



Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université Mouloud Mammeri, Tizi-Ouzou
Faculté Génie Electrique et Informatique
Laboratoire Vision Artificielle et Automatique des Systèmes



Proceedings

of

the 4th international Conference on Artificial Vision

CVA' 2021

Tizi-Ouzou, Novembre 21-23, 2021



Lucius Apuleius Madaurensis or Afulay in amazigh language
He was born in M'Daourouch, Algeria
He wrote "The Golden Ass", the first novel in the history of humanity



Honorary Presidents of the Conference

Pr Ahmed BOUDA, Rector of the University Mouloud Mammeri of Tizi-Ouzou (UMMTO)

Pr Redouane KARA, Dean of the faculty of Electrical Engineering and Computer Science.

President of the Conference

Pr Moussa DIAF, FGEI, UMMTO

Organizing Committee

M-Outahar BENSIDHOUM, President

Dalila HOCINE

Abdellatif BOUZID-DAHO

Boussad IDJERI

Ouiza BOUKENDOUR

Samir YESLI

Salim IRATNI

Merzouk YOUNSI

Yasmine SILHADI

Azouaou AOUJ

Program Committee

Pr Patrick SIARRY (UPEC, France)
Pr Tahar KECHADI (Univ. Dublin, Ireland)
Pr Mounir BENAÏSSA (ENIT, Tunisie)
Pr Yacine YADDADEN (Univ. Québec, Canada)
Pr Mourad FAKHFAKH (Univ. Sfax, Tunisie)
Pr Amine NAIT-ALI (UPEC, France)
Pr A.TALEB-AHMED (Univ. Valenciennes, France)
Pr Zahia GUESSOUM (Univ. Paris 6, France)
Pr Mhania GUERTI (ENP, Algeria)
Pr Amer RAGAB ZEREK (Univ. Zawia, Libya)
Pr Nacera BENAMRANE (Univ. Oran, Algeria)
Pr Abderrezak GUESSOUM (Univ. Blida, Algeria)
Pr Slimane LARABI (USTHB, Algeria)
Pr M-Said DJOUADI (EMP, Algeria)
Pr Malika MIMI (Univ. Mostaganem, Algeria)
Pr Abdelouahab MOUSSAOUI (Univ. Sétif, Algeria)
Pr Nasr-Eddine BERRACHED (Univ. Oran, Algeria)
Pr Youcef SMARA (USTHB, Algeria)
Pr Hicham TEBBIKH (Univ. Guelma, Algeria)
Pr Fella HACHOUF (Univ. Constantine 1, Algeria)
Pr Salim SBAA (Univ. Biskra, Algeria)
Pr Soltane AMEUR (UMMTO, Algeria)
Pr Nacereddine BENAMROUCHE (UMMTO, Algeria)
Pr Moussa DIAF (UMMTO, Algeria)
Pr Mourad LAGHROUCHE (UMMTO, Algeria)
Pr Zohra MAZOUZI (UMMTO, Algeria)
Pr Mourad LAHDIR (UMMTO, Algeria)
Pr Kamal HAMMOUCHE (UMMTO, Algeria)
Pr Arezki ZIANI (UMMTO, Algeria)
Pr Malik SI-MOHAMMED (UMMTO, Algeria)
Pr Mohamed BOUDOUR (USTHB, Algeria)
Pr Arezki BENFDILA (UMMTO, Algeria)
Pr Rachid AHMED-OUAMAR (UMMTO, Algeria)
Pr Mourad LAZRI (UMMTO, Algeria)

FULL PAPERS

Walid HARIRI, Lina BOUREFIS Accurate Transfer Learning and Deep Ensemble Models for COVID-19 Detection using X-ray Images	1
Amira CHEKIR Deep Neural Network Using Local Shape Features Learning for White Matter Fibers Classification	8
Imed Eddine HAOU LI, Walid HARIRI, Hassina SERIDI-BOUCHELAGHEM Non-trainable and Semi-trainable modes for COVID-19 Detection using Vision Transformers and X-ray Images	13
Amira CHEKIR Self-organizing Algorithm based on Multi-agent system for White Matter fibers Clustering with Outliers Elimination	20
Salim IRATNI, Merzouk YOUNSI, Moussa DIAF Human Detection Method based on Combined Face and Upper-Body Part Detectors	25
Laidi AMEL, Mohammed AMMAR Recognition Process of the Right and the Left Ventricles in short-axis MRI images Based on ANFIS and KFCM Algorithm	32
Ayoub Abderrazak MAAROUF, Fella HACHOUF, Soumia KHARFOUCHI K-NN Classification Using Two Dimensional Random Coefficients Autoregressive Model (2D-RCA)	38
Adel BOUDIAF, Khaled HARRAR, Mostefa Mohamed TOUBA, Rachid ZAGHDOUDI, Slimane ZIANI Automatic System Inspection of Surface Defects on X-Ray Images using Gray-level Co-occurrence Matrix and Decision Tree Model	43
Soumia FAOU CI, Djamel GACEB, Mohammad HADDAD Deep Transfer learning for Arabic handwritten text recognition: A comparative study	50
Noureddine BENS AFI, Youcef ATTA F, Mourad LAHDIR Meteosat Image Compression Based 2d Gradient Tensor	57
Mohand OUAHIOUNE, Leila AKROUR, Boujemaa AYOUNI, Mourad LAHDIR, Soltane AMEUR Wavelet learning-based super-resolution technique for hyperspectral image	65
Leila AKROUR, Mohand OUAHIOUNE, Mourad LAHDIR, Soltane AMEUR, Régis FOURNIER, Amine NAIT ALI Hyperspectral images compression based multimodal schemes	71
Boussad IDJERI, M.T. BENSIDHOUM, Mourad LAGHROUCHE Design of a New Multi-sensor System Probe For Remote Wind Parameters Monitoring	77
Fares KARA OUI, Hamida BOUGHERIRA, Nadjia BENBLIDIA, Abderrezak GUESSOUM, Nadia KHETTIB, Djamila NACEUR, Zoubir BENSELAMA, Zoheir MELIANI, Abdoulkarim MAMOUDOUMOUMOUNI, Bouba TANGARA Laser shaping by Digital Micromirrors for multiple targets tracking	82
Katia HANNOUN, Ouerdia MEGHERBI, Hamid HAMICHE, Mourad LAHDIR, Mourad LAGHROUCHE A Novel Image Watermarking Scheme In DWT Domain Based on Discrete-Time	86

Chaotic Systems Of Fractional and Integer Orders	
Farid HAMMOU, Kamal HAMMOUCHE An Improved PSO Based Approach For Two-Dimensional IIR Filter Design	93
Ahcene HAMOUDI, Nadia DJEGHALI, Maamar BETTAYEB Video encryption-decryption based on synchronisation of chaotic systems for secure communication	99
Mohammed Amin KHELIFA, Badis LEKOUAGHET, Abdelkrim BOUKABOU An efficient WSN Localization Scheme using the Improved Chaos-GBO algorithm	106
Salim IRATNI, Ouiza BOUKENDOUR, Souhila BOUMATI, Moussa DIAF Contribution to Automatic Tomatoes Sorting According to Their Maturity using an Evolutionary Classification Method	112

ABSTRACTS

Abdellatif BOUZID-DAHO, Patrick SIARRY Feature Extraction and Clustering of Abnormal Blood Cells	119
Samir YESLI, Moussa DIAF Anisotropic Diffusion Filtering Optimized by Simplex Method Applied to MRI images with a Priori on the Rayleigh Probability Density Function of the Noise	120
Abderrahim BENMOHAMED, Adil BOUGUERRA Intelligent system for detecting faults in the industrial area	121
Sidali BOURENANE, Sid Ahmed HENNI, Benabdellah YAGOUBI The safety of people in a intelligent building by classifying speech with neuron networks and hidden markov chains	122
Ali HAMROUN, Karim LABADI, Fethi OUALOUCHE, Mourad LAZRI, Soltane AMEUR Stochastic and discrete-event model for simulation and performance evaluation of car-sharing networks	123
Fatiha ZAGHRAT, Brahim GASBAOUI, Boussemaha BOUCHIBA, Ismail Khalil BOUSSERHANE, Imane CHAOUFI Based robust Sliding Mode Control of two wheels EV	124
Mourad KHETATBA, Rachid BOUDOUR Enhancing Firefly Algorithm by Modifying Evolution Rules and Introducing Position based Crossover	125

Accurate Transfer Learning and Deep Ensemble Models for COVID-19 Detection using X-ray Images

Walid Hariri * ¹, Lina Bourefis ¹

¹ Labged Laboratory, Computer Science Department, Badji Mokhtar Annaba University, Algeria

hariri@labged.net; bourefis.lina97@gmail.com

Abstract—The pandemic of Coronavirus disease (COVID-19) has become one of the main causes of mortality over the world. In this paper, we employ a transfer learning-based method using five pre-trained deep convolutional neural networks (CNN) architectures fine-tuned with an X-ray image dataset to detect the COVID-19. Hence, we use VGG-16, ResNet50, InceptionV3, ResNet101 and Inception-ResNetV2 models. The purpose consists of classifying the input images into three classes (COVID-19 / Healthy / Other viral pneumonia). The results of each model are presented in detail using 10-fold cross-validation and comparative analysis has been given among these models by taking into account different elements in order to find the more suitable model. To further enhance the performance of single models, we propose to combine the obtained predictions of these models using the majority vote strategy. The proposed method has been validated on a publicly available chest X-ray image database that contains more than one thousand images per class. Evaluation measures of the classification performance have been reported and discussed in detail. Promising results have been achieved compared to state-of-the-art methods where the proposed ensemble model achieved higher performance than using any single model. This study gives more insights to researchers for choosing the best models to accurately detect the COVID-19 virus.

Index Terms—COVID-19, CNN, Transfer learning, Ensemble model, X-ray images

I. INTRODUCTION

Since the spread of COVID-19, the real-time polymerase chain reaction (RT-PCR) was the more popular technique applied to detect this virus. Despite the good performance achieved by this technique, it still has many problems like time-consuming, false negative results, and its expensive price [1]. Since the mortality cases with COVID-19 is constantly increasing, the aforementioned drawbacks of RT-PCR test could further complicate the situation. Recently, deep learning techniques to detect and diagnose the COVID-19 have become an active research area using X-ray images or (Computerized Tomography) CT scans [2]. Their high performance achieved using many deep learning models motivated the researchers to adopt this novel technique to prevent against the COVID-19 pandemic. Besides the high performance, deep learning-based techniques are very fast compared to RT-PCR test [3]. Therefore, we propose in this paper various deep learning based-strategies to deal with the COVID-19 detection and diagnosis using publicly available dataset of X-ray images. The remainder of the paper is structured as follows: in Section

II, we present the related works. Section III explains the contribution of the paper. The proposed method is presented in detail in Section IV. Experimental results and comparative study are reported in Section V. Conclusions end the paper.

II. RELATED WORKS

Few weeks after the propagation of COVID-19 pneumonia, many works to detect the virus from radiography imaging are carried out using deep learning-based techniques. Among these techniques we can find "*traditional deep learning methods*" that aim to train deep models from scratch using a specified labeled dataset. Since the appearance of the COVID-19 pandemic, some datasets have been introduced to allow researchers to test their models. For example, Zheng et al. [4] trained a supervised deep learning model. The segmentation of the lung region is applied using Unet model from CT-scans. Other methods based on "*deep features extraction*" where the deep pre-trained models have been widely used as feature extractors, in which the last convolutional layers or the fully connected layers are used to feed a machine learning classifier. For example, Ismael et al. [5] applied five pre-trained models including VGG-16, ResNet18, ResNet50, ResNet101, and VGG19 to train an SVM classifier. Different kernel functions are then used in the SVM classification stage such as Linear, Quadratic, Cubic, and Gaussian kernels. An other method used AlexNet-based features to feed an SVM classifier is introduced in [6]. In this work, the deep features are extracted from the fully-connected and convolution layers. Another method proposed by Rahimzadeh et al. [7] aims to combine the deep extracted features from Xception [8] and ResNet50V2 [9] networks. A global feature vector is then generated to train a classifier. From another point of view, to be able to make a real time detection of COVID-19, training a deep model from scratch has many problems, especially the insufficiency of representative data and also it is time-consuming and requires high performance machines. In this case "*transfer learning (TL)*" were the most useful technique to figure out train time and data troubles. TL is one of the deep learning approaches that consists of reusing a pre-trained model for one job to accomplish another one in the same domain of missions. By way of example, Vaid et al. [10] applied a transfer learning method using VGG-16

pre-trained model. They used a labeled frontal X-ray images dataset of patients from different countries around the world. The particularity of the used dataset lies in the additional information of each patient such as location, old and gender. Das et al. [11], however, used the extreme version of Inception (Xception) model, in order to develop an automated deep transfer learning to detect COVID-19 pneumonia in X-ray images. Transfer learning has also been used to classify the CT scans of lungs into COVID-19 or NORMAL cases as presented in [12]. Four pre-trained models are then used including ResNet18, ResNet50, ResNet101, and SqueezeNet. A different transfer learning-based method using the DetRaC model is presented in [13]. The combination of TL and the DetRaC model makes the proposed method able to deal with any irregularities in the image dataset by investigating its class boundaries using a class decomposition mechanism. Authors in [14] used eight pre-trained models namely, AlexNet, VGG-16, GoogleNet, MobileNet-V2, Squeezenet, ResNet34, ResNet50 and InceptionV3. They evaluated the pre-trained models with X-ray illustration taken from covid-chestxray-dataset [15].

To deal with the lack of grand amount of labeled datasets, "generative models" have been widely used to generate new images using the existing ones. Many strategies have been carried out such as flipping the image horizontally or vertically, zooming in or out. For example, Loey et al. [16] proposed a model of two axis, the first one about the data augmentation using common techniques across Conditional generative adversarial network (CGAN), the second axe is about deep TL model, which is formed of five model, named as following: AlexNet, VGG-16, VGG-19, GoogleNet and ResNet50. All of these models are fine-tuned with COVID-19 CT-image dataset.

III. CONTRIBUTION OF THE PAPER

A transfer learning-based technique is applied in this paper to detect COVID-19 virus using labeled datasets of X-ray images. To avoid training a deep CNN from scratch on a limited labeled dataset, we propose in this paper to carry out a transfer learning technique using five pre-trained models and acquired data only to fine-tune them. This is very useful when the data is abound for an auxiliary domain, but very limited labeled data is available for the domain of experiment. Figure 1 presents our proposal overview.

We opted for the following pre-trained models: VGG-16, ResNet50, InceptionV3, ResNet101 and Inception-ResNetV2. This choice is based on the diversity of these models, the difference of their architecture as well as their structure. A comparative study is then conducted between these models in terms of training accuracy, loss accuracy, validation accuracy and validation loss during the training stage. A confusion matrix is then generated after the classification of test samples. Other performance measures are computed to show the efficiency of each model (e.g. recall, precision, F-score). The difference between the applied models can be useful in our second step where their outputs will be combined using an ensemble learning technique (also called ensemble model)

using the majority vote strategy. This combination enhances the classification performance of the non-learned samples compared to the obtained rates using each pre-trained model separately.

IV. THE PROPOSED METHOD

A. Pre-trained models

In this Section, we present the architecture of the five models that we used in our TL system. These models are pre-trained on ImageNet database [17].

- **VGG-16:** [18] is trained on the very large ImageNet dataset which has over 14 million images and 1000 classes. It contains 16 layers including 13 convolutional layers, 3 dense layers and 5 Max Pooling layers. Each convolutional layer is 3x3 layer with a stride size of 1 and the same padding. The pooling layers of VGG-16 are all 2x2 pooling layers with a stride size of 2. Figure 2 presents modified VGG-16 architecture as an example of TL.
- **ResNet50:** ResNet50 is a variant of ResNet pre-trained model on ImageNet dataset which has 48 Convolution layers along with 1 MaxPool and 1 Average Pool layer. It has 3.8×10^9 Floating points operations [19].
- **ResNet101:** Another variant of ResNet deep neural network series, trained on more than a million images from the ImageNet database [20]. It consists of 101 deep layers with identity connection between them.
- **Inception-ResNetV2:** A CNN that builds on the Inception family of architectures [21]. Its architecture combines the Inception architecture with residual connections. This CNN contains 164 deep layers trained on ImageNet dataset and is able to classify images into 1000 classes.
- **InceptionV3:** presented by google is the third version of Inception DL convolution architectures, with 42 deep layers contain Convolution layer, AvgPool, MaxPool, Concat, Dropout, Fully Connected layer and Softmax activation function [22]. The input layer size of this model is different from the other models ($299 \times 299 \times 3$ instead of $224 \times 224 \times 3$).

B. Transfer learning

We aim to transfer the acquired information from the CNN models pre-trained on ImageNet dataset to our specific task. The issue that needs to be addressed is the highly dependance of these models on the initial dataset, whereas our Chest X-ray images are different. Consequently, the generalization of the network will be poor since the extracted features from the large amount of data are not inadequate to represent our target images (to feed a classifier or softmax function). The solution consists of fine-tuning the pre-trained models on our Chest X-ray images dataset that is a very small dataset compared to the ImageNet. In other words, pre-trained CNN structures are updated to suit our classification task. This strategy is generally much faster than the traditional training of the CNN model from scratch with arbitrary weights. For example, using VGG-16, the total number of parameters after training is 14,789,955.

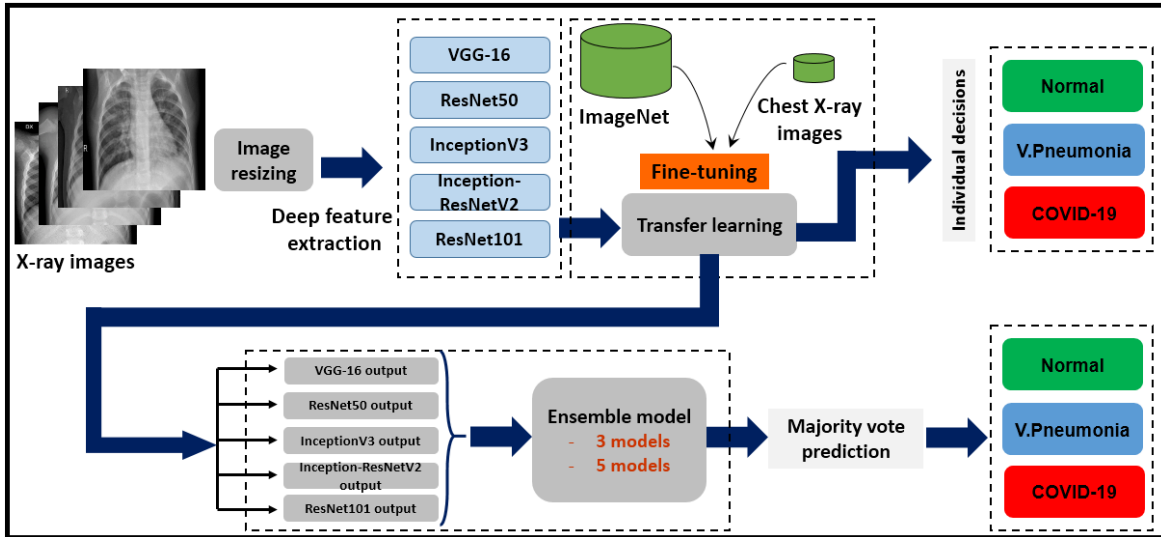


Fig. 1: Overview of the different steps of our proposal.

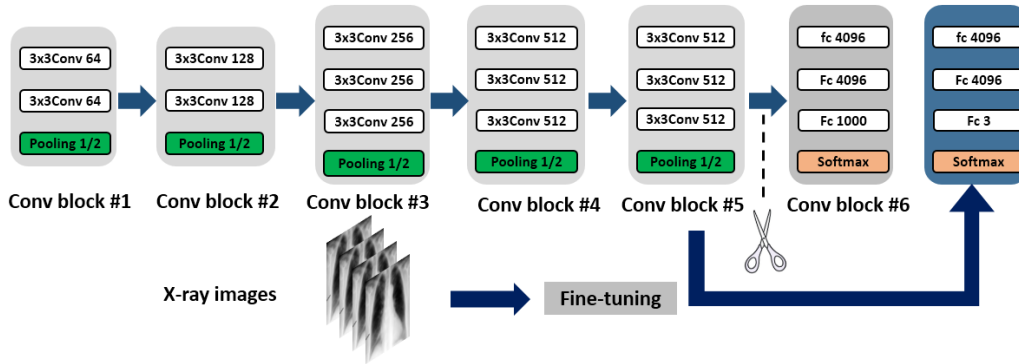


Fig. 2: The modified VGG-16 network architecture.

The number of trainable parameters is 75,267 which is very small compared to 14,714,688 of non-trainable parameters.

C. Deep ensemble learning

To enhance the classification performance, we exploit the different architectures of the five pre-trained models and we fuse their outputs to make a global decision. Thus, we propose to apply a majority voting strategy as an ensemble learning stage. We then use the combination of 5 and 3 models, respectively using the output of the last epoch (30) of the 10th training fold. The obtained results are presented and discussed in detail in the following section.

V. EXPERIMENTAL RESULTS

All the experiments were performed on Windows 7 operating system 64 bits the TensorFlow/Keras framework of python language. The implementation of our proposal is provided by Google Colaboratory Notebooks. The obtained results by each pre-trained model, and using the deep ensemble model strategy are presented in the following.

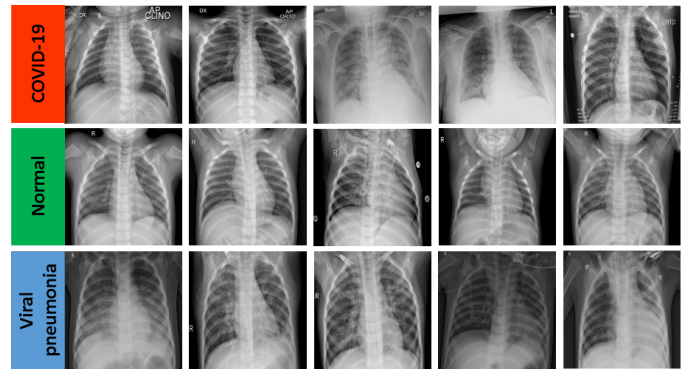


Fig. 3: X-ray images from COVID19 Radiography Database [23].

A. Database description

COVID19 Radiography Database [23]: contains 3616 COVID-19 positive cases along with 10,192 Healthy, 6012 Lung Opacity (Non-COVID lung infection), and 1345 Viral Pneumonia images. In this work, we carry out 3 class clas-

sification (COVID-19, Healthy and Viral pneumonia). Some scans from this database are shown in Figure 3.

B. Method performance

The 10 fold-cross validation setting is applied using the 4th version of COVID19 Radiography Database. It is randomly split into training and test datasets. The initial learning rate is of 0.0003 and cross entropy loss. The models are trained for 30 epochs where the batch size is 32.

In the following, we show the training performances (accuracy and loss) as well as the validation performances using the three best models including Inception-ResNetV2, VGG-16 and InceptionV3, respectively. The performance during the fold 1, 6 and 10 are displayed in detail in Figure 6. These curves show that using VGG-16 architecture, the highest training accuracy is observed 98.85% in epoch 7 where the highest validation accuracy is 99.65% in epoch 24. The loss is 0.0324 and 0.024 at epoch 30 of training and validation respectively. All the reported results are put out from fold 10.

On the contrary, Inception-ResNetV2 achieved higher training accuracy compared to VGG-16 by 99.46% in epoch 29. The highest validation accuracy is 100%. Training and validation loss are respectively 0.026 and 0.005 (See Figure 7). Although the initial loss value of Inception-ResNetV2 is very high compared to that of VGG-16 (1.00 vs 0.2), the previous values show that Inception-ResNetV2 is more efficient during the training and validation stages compared to VGG-16 over all the epochs of the 10th fold. InceptionV3, however, is the third best model among the five models (See Figure 8).

More details about the performance measures of the best three models as well as the two remaining ones (ResNet50 and ResNet101) are shown in Table 1. According to the displayed values of precision, recall, and F-score, ResNet101 is slightly better than ResNet50. This performance can be clearly seen in the confusion matrix of these models. Figure 4 shows that ResNet101 has less false classifications. From the same Figure, we can also notice that false classified samples of VGG-16 and Inception-ResNetV2 are very limited compared to the two previous models. This explains the very high rates registered as recall, precision and F-score.

The use of ensemble learning, however, enhances all the previous performances using the combination of the three and five models' outputs as shown in the Table 1. Using this strategy, we achieve 1.00 of precision, recall and F-score when dealing with the COVID-19 classes. Compared to recent state-of-the-art methods, our proposed methods achieved higher classification rate of the test set by 0.96 using the ensemble model of the combination (VGG-16+ResNet50+ResNet101).

C. Discussions

The experimental study proves that the TL is one of the best deep learning techniques to efficiently detect the COVID-19 using X-ray images.

The obtained results show that the ensemble models enhance the classification performance of the fine-tuned CNNs. This paper proposes five pre-trained models that give promising

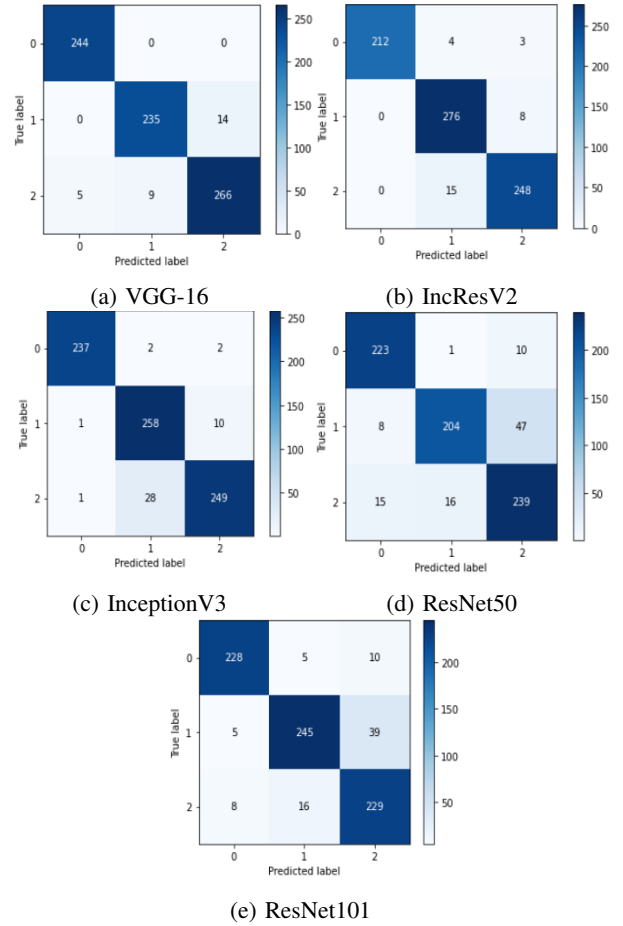


Fig. 4: Confusion matrix of the pre-trained models. **Label 0** refers to COVID-19 class, **label 1** refers to Healthy class, and **label 2** refers to the VP.

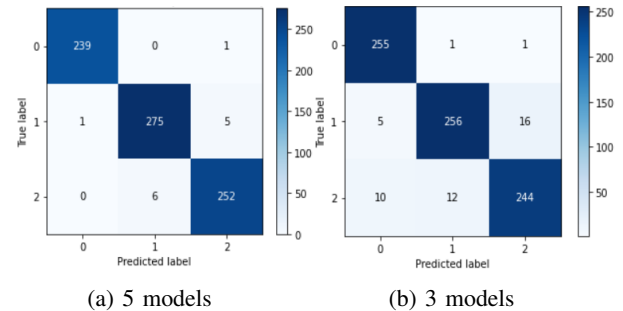


Fig. 5: Confusion matrix of the majority vote prediction. (a) refers to all the five models. (b) refers to the following three models: VGG-16, ResNet50 and ResNet101.

results which are still competitive to those of the state-of-the-art method. The TL strategy using the combination of the best models outperforms the other methods and gives 0.98 accuracy of test images. The high performances achieved are explained by the fact that TL is suitable since the first CNN layers learn low-level features. These features are mostly invariable from a classification task to another (i.e. edges). The fine-tuning

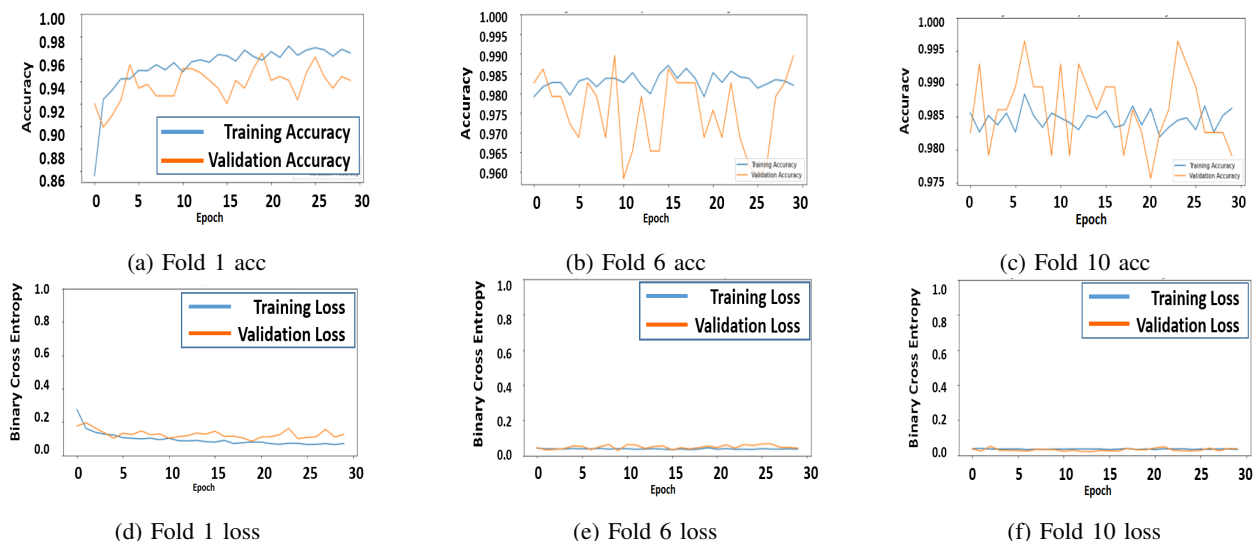


Fig. 6: Training and validation accuracy / loss using VGG-16.

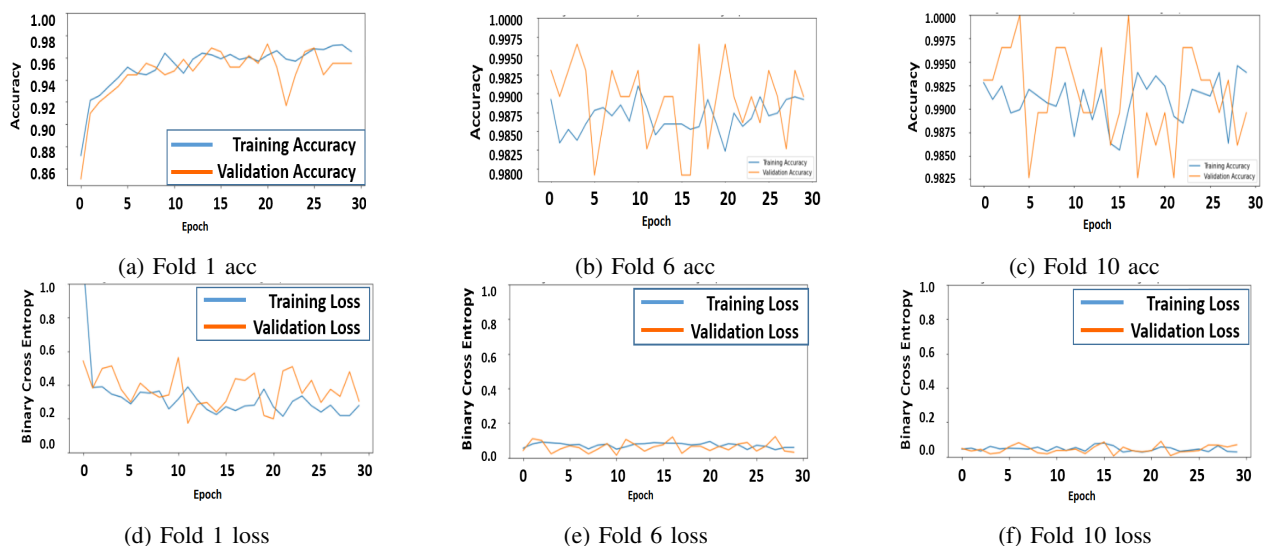


Fig. 7: Training and validation accuracy / loss using Inception-ResNetV2.

provides specific features of the target domain such as COVID-19 detection.

VI. CONCLUSION

Since the epidemic is still fast-spreading, the proposed method seems to be a good solution to early diagnose the virus. We fine-tuned five pre-trained CNN models to our COVID-19 dataset of X-ray images. High classification performances have been achieved especially with VGG-16 and Inception-ResNetV2. Slightly lower performances have been achieved using the other models. This transfer learning technique reduces considerably the training cost compared to learning from scratch that becomes an amassed technique. To exploit the different features extracted by each model, we proposed to combine their outputs to find a global decision through a majority voting strategy. This strategy further enhances the

performance of the proposed transfer learning-based method. In future work, we look at the application of Graph neural network along with CNNs to improve the performance and reduce the computation time.

VII. ACKNOWLEDGMENTS

The authors would like to thank the Agency for Research Results Valuation and Technological Development, 'Agence Nationale de Valorisation des Résultats de la Recherche et du Développement Technologique' Algérie (DGRSDT).

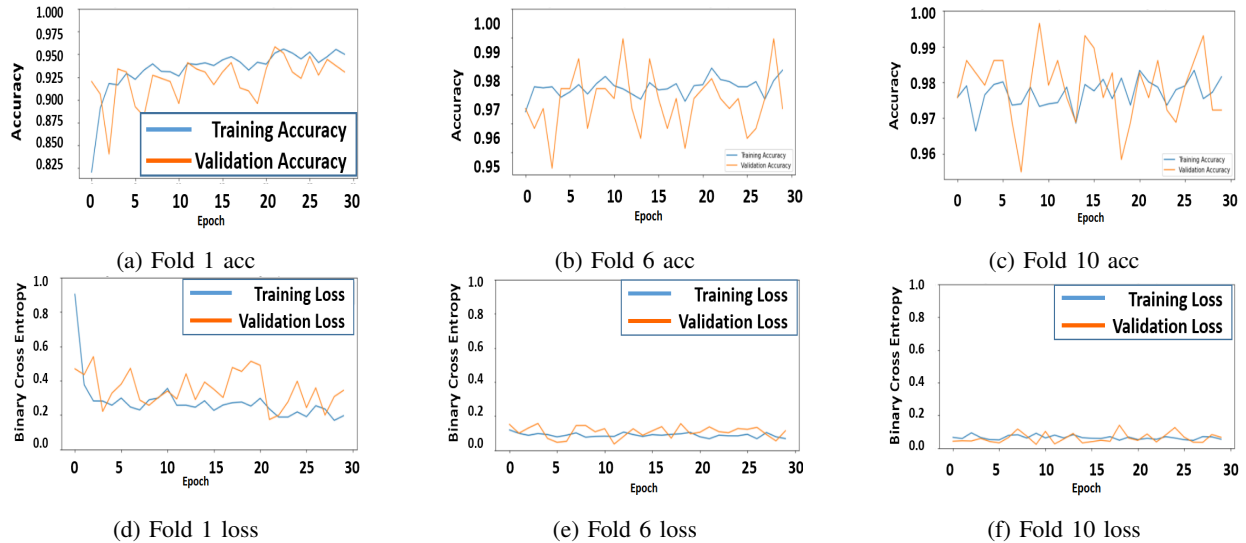


Fig. 8: Training and validation accuracy / loss using InceptionV3.

TABLE I: Performance of 5 fine-tuned CNNs and ensemble learning with X-ray images of COVID-19, Healthy and VP cases.

Model / measure	Precision			Recall			F-score		Accuracy	
	COVID	Healthy	VP	COVID	Healthy	VP	COVID	Healthy		
Class	COVID	Healthy	VP	COVID	Healthy	VP	COVID	Healthy	Test set	
VGG-16	0.98	0.96	0.95	1.00	0.94	0.95	0.99	0.95	0.95	0.96
ResNet50	0.91	0.92	0.81	0.95	0.79	0.89	0.93	0.85	0.84	0.88
ResNet101	0.95	0.92	0.82	0.94	0.85	0.91	0.94	0.88	0.86	0.90
Inception-ResNetV2	1.00	0.94	0.96	0.97	0.97	0.94	0.98	0.95	0.95	0.96
InceptionV3	0.99	0.90	0.95	0.98	0.96	0.90	0.99	0.93	0.92	0.95
All five models	0.94	0.95	0.93	0.99	0.92	0.93	0.97	0.94	0.93	0.94
VGG-16+ResNet50+ResNet101	1.00	0.98	0.98	1.00	0.98	0.98	1.00	0.98	0.98	0.98
Pham et al. [24]	/	/	/	/	/	/	/	/	/	0.96
Cavallo et al. [25]	/	/	/	/	/	/	/	/	/	0.91
Rajpal et al. [26]	0.99	0.91	0.94	0.96	0.96	0.92	0.97	0.93	0.93	0.94

REFERENCES

- [1] A. Altan and S. Karasu, "Recognition of covid-19 disease from x-ray images by hybrid model consisting of 2d curvelet transform, chaotic salp swarm algorithm and deep learning technique," *Chaos, Solitons & Fractals*, vol. 140, p. 110071, 2020.
- [2] S. H. Yoo, H. Geng, T. L. Chiu, S. K. Yu, D. C. Cho, J. Heo, M. S. Choi, I. H. Choi, C. Cung Van, N. V. Nhung *et al.*, "Deep learning-based decision-tree classifier for covid-19 diagnosis from chest x-ray imaging," *Frontiers in medicine*, vol. 7, p. 427, 2020.
- [3] H. Panwar, P. Gupta, M. K. Siddiqui, R. Morales-Menendez, P. Bhardwaj, and V. Singh, "A deep learning and grad-cam based color visualization approach for fast detection of covid-19 cases using chest x-ray and ct-scan images," *Chaos, Solitons & Fractals*, vol. 140, p. 110190, 2020.
- [4] C. Zheng, X. Deng, Q. Fu, Q. Zhou, J. Feng, H. Ma, W. Liu, and X. Wang, "Deep learning-based detection for covid-19 from chest ct using weak label," *MedRxiv*, 2020.
- [5] A. M. Ismael and A. Şengür, "Deep learning approaches for covid-19 detection based on chest x-ray images," *Expert Systems with Applications*, vol. 164, p. 114054, 2021.
- [6] M. Turkoglu, "Covidetectionnet: Covid-19 diagnosis system based on x-ray images using features selected from pre-learned deep features ensemble," *Applied Intelligence*, vol. 51, no. 3, pp. 1213–1226, 2021.
- [7] M. Rahimzadeh and A. Attar, "A new modified deep convolutional neural network for detecting covid-19 from x-ray images," *arXiv preprint arXiv:2004.08052*, 2020.
- [8] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.
- [10] S. Vaid, R. Kalantar, and M. Bhandari, "Deep learning covid-19 detection bias: accuracy through artificial intelligence," *International Orthopaedics*, vol. 44, pp. 1539–1542, 2020.
- [11] N. N. Das, N. Kumar, M. Kaur, V. Kumar, and D. Singh, "Automated deep transfer learning-based approach for detection of covid-19 infection in chest x-rays," *Irbm*, 2020.
- [12] S. Ahuja, B. K. Panigrahi, N. Dey, V. Rajjikanth, and T. K. Gandhi, "Deep transfer learning-based automated detection of covid-19 from lung ct scan slices," *Applied Intelligence*, vol. 51, no. 1, pp. 571–585, 2021.
- [13] A. Abbas, M. M. Abdelsamea, and M. M. Gaber, "Classification of covid-19 in chest x-ray images using detrac deep convolutional neural network," *Applied Intelligence*, vol. 51, no. 2, pp. 854–864, 2021.
- [14] S. R. Nayak, D. R. Nayak, U. Sinha, V. Arora, and R. B. Pachori, "Application of deep learning techniques for detection of covid-19 cases using chest x-ray images: A comprehensive study," *Biomedical Signal Processing and Control*, vol. 64, p. 102365, 2020.
- [15] J. P. Cohen, P. Morrison, L. Dao, K. Roth, T. Q. Duong, and M. Ghassemi, "Covid-19 image data collection: Prospective predictions are the future," *arXiv preprint arXiv:2006.11988*, 2020.
- [16] M. Loey, G. Manogaran, and N. E. M. Khalifa, "A deep transfer learning model with classical data augmentation and cgan to detect covid-19 from chest ct radiography digital images," *Neural Computing and Applications*, pp. 1–13, 2020.
- [17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
- [18] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

- [20] J. Dai, Y. Li, K. He, and J. Sun, "R-fcn: Object detection via region-based fully convolutional networks," *arXiv preprint arXiv:1605.06409*, 2016.
- [21] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.
- [22] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [23] T. Rahman, "COVID-19 Radiography Database," <https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/>, 2020, [Online; accessed 05-April-2021].
- [24] T. D. Pham, "Classification of covid-19 chest x-rays with deep learning: new models or fine tuning?" *Health Information Science and Systems*, vol. 9, no. 1, pp. 1–11, 2021.
- [25] A. U. Cavallo, J. Troisi, M. Forcina, P. Mari, V. Forte, M. Sperandio, S. Pagano, P. Cavallo, R. Floris, and F. Geraci, "Texture analysis in the evaluation of covid-19 pneumonia in chest x-ray images: a proof of concept study," 2020.
- [26] S. Rajpal, M. Agarwal, A. Rajpal, N. Lakhyani, and N. Kumar, "Cov-elm classifier: an extreme learning machine based identification of covid-19 using chest x-ray images," *arXiv preprint arXiv:2007.08637*, 2020.

Deep Neural Network Using Local Shape Features Learning for White Matter Fibers Classification

CHEKIR AMIRA

*Laboratoire de Robotique, Parallélisme et systèmes Embarqués (LRPE)
Université des sciences et de la technologie Houari-Boumediène (USTHB)
Algiers, Algeria
chekir.amira@gmail.com*

Abstract—Diffusion Magnetic Resonance Imaging (dMRI) provides a promising way for estimating the neural fiber pathways in the human brain non-invasively via White Matter (WM) tractography process. The fibers classification generated by tractography into meaningful bundles is of general interest in neuroscience. It produces a quantitative description of the brain connection, applied in the study of brain disorders. However, fiber classification is a hard task due to the huge size of tractography dataset and his complexity. We propose a deep-learning approach for WM fiber classification based on estimating local 3D fiber shape properties. In our approach, we consider WM fibers as 3D point clouds and we directly estimate local shape properties such as normals (both unoriented and oriented) and curvature from raw point clouds. The performances of our method are demonstrated by experimental results using real WM fiber dataset.

Index Terms—White Matter fiber, Diffusion Magnetic Resonance Imaging, deep learning, 3D point clouds, Shape classification.

I. INTRODUCTION

Diffusion Magnetic Resonance Imaging (dMRI) [2], [1], is a non-invasive and in-vivo modality extensively used for the study of brain tissue microstructure and WM connection structure. The dMRI data enable measurement of the neural fibers trajectory using computational methods called tractography [3], [4]. Tractography methods trace these trajectories by following probable fiber orientations.

The analysis of dMRI tractography dataset is of very important interest, generating a quantitative description of WM connections called "connectome". Two popular categories of methods were proposed in the literature. The first category, parcellation-based segmentation methods, that divided fibers according to the cortical anatomical regions, that they connect [5].

The second one is the classification of fibers by measuring their property or properties. The pairwise similarity between fibers was calculated and stored in a matrix [6], [7]. However, the first category of methods requires extensive knowledge about complex 3D WM fibers anatomy and intensive user interaction. The second category, although it is fully automatic, it stays a very complicated task due to the huge size of the WM tractography dataset and its composition by various fibers with complex geometries; with different shape, size, and density.

On the others hand, several methods have been proposed for analyzing and processing 3D shapes by building on the success of deep learning methods. These methods are based on the notion that the solutions to many problems in geometric data analysis can benefit from large data repositories. Many learning-based approaches are aimed at estimating global properties for tasks such as shape classification and often are based either on 2D projections (views) of 3D shapes or on volumetric representations [8]. However, several methods have also been proposed for shape segmentation [9] and even shape correspondence [10], [11], among many others. In recent years, few approaches have been proposed to directly operate on point clouds. The approaches of Boulch et al. [12] and Qi et al. [13] propose an architecture for estimating unoriented normals on point clouds by creating a Hough transform-based representation of local shape patches and using a convolutional neural network for learning normal directions from a large ground-truth corpus. While not projection based, this method still relies on a 2D-based representation for learning and moreover loses orientation information, which can be crucial in practice. More recently, the PointNet architecture [13] has been designed to operate on 3D point clouds directly. This approach allows to estimate properties of 3D shapes without relying on 2D projections or volumetric approximations. The original architecture is purely global, taking the entire point cloud into account, and has since then been extended to a hierarchical approach in a very recent PointNet++ [13], which is designed to better capture local structures in point clouds. The authors in proposed PCPNet method which is based on the original PointNet architecture, but rather than using it for estimating global shape properties for shape classification or semantic labeling, as has also been the focus of PointNet++ [13], it was used explicitly for estimating local differential properties including oriented normals and curvature. In this work, we propose a new deep learning architecture using an adaptation of PCPNet method [14] to directly learn local properties (normals and curvatures) using groundtruth white matter bundle atlas and we extend it by adding several fully connected layers to classify the WM fibers from new subjects. Our approach allows computing local feature vectors that describe the local neighborhood around a fiber point. These features are better suited to estimate local fiber properties. The originality

of this work is that the WM fibers are considered as 3D point cloud series. The paper is organized as follows: In Section II, we present the concept of the method. Section III introduces the datasets, used to train our model and validate the proposed method, and contains extensive experiments demonstrating the efficiency of our method. Section V concludes the paper.

II. METHODOLOGY

A. Deep Learning on White Matter Fibers

1) *Geometric Features Estimation:* Given a brain point cloud $P = p_1, \dots, p_N$, the PCPNet [14] method is applied to a set of points of interest for each fiber $P_i^r \in P$ composed a bundle. We then estimate local shape properties at these points of interest for all fibers of the bundle. The network learns a set of k non-linear functions in the local fiber neighborhoods, which can intuitively be understood as a set of density estimators for different regions of the fiber. These give a k -dimensional feature vector per fiber that can then be used to regress various local features. The PCPNet method estimated the local fiber properties, such as normal n_i and principal curvature values k_i^1 and k_i^2 at the points of interest. Our network takes a fixed number of points by fiber as input. The fibers were down sampled to have the same number of point, we choose 100 points by fiber.

2) *Deep Learning Architecture:* Our architecture follows the PCPNet framework, with two changes: We add the fully-connected layers to organize the 3D points of fibers in bundle and we used the max-pooling layers to concatenate multiples learned features and to summarize the statistics of each fiber. Our architecture is multi-scale. We input all bundles of interest and we triple the number of point functions to apply each function to a bundle. The sum, is computed over each bundle separately. This results in nine-fold increase in patch features H , which are then used to regress the output properties. Our architecture is composed by the following layers:

Second Spatial Transformer layers: The Second Spatial Transformer (STN) layers were used for the transformation as a fully connected layer with weights that are computed from the feature vectors g of all points of interest in the fiber. The STN layer operates on the feature vector $g_j = (g_1(p_j), \dots, g_{64}(p_j))$, giving a 64×64 transformation matrix. This introduces global information into the point functions and increasing the performance of the network.

Point functions and symmetric operation layers: One important property of the network is that it should be invariant to the input point ordering. It can be achieved by applying a set of functions h_1, \dots, h_k with shared parameters to each point of fiber separately and then combine the resulting values for each point using a symmetric operation:

$$H_l(P_i^r) = \sum_{p_j \in P_i^r} h_l(p_j) \quad (1)$$

$H_l(P_i^r)$ is then a feature of the fiber and h_l are called point functions; they are scalar-valued functions, defined in the in the local coordinate frame of the fiber. The functions H_l can

intuitively be understood as density estimators for a region given by h_l . Their response is stronger the more points are in the non-zero region of h_l . We used the max-pooling to give our point functions a simple conceptual interpretation as density estimators. The point functions h_l are computed as:

$$h_l(P_j) = (MLP_2 \circ STN_2)(g_1(p_j), \dots, g_{64}(p_j)) \quad (2)$$

where MLP_2 is a three-layer fully-connected network and STN_2 is the second spatial transformer. The functions g can be understood as a less complex set of point functions, since they are at a shallower depth in the network. They are computed analogous to h .

Fully connected layers: In a trained model, the bundle feature vector $H_j = (H_l(P_i^r), \dots, H_k(P_i^r))$ provides a rich description of the bundle. We used a fully connected layer to perform classification on the learned shape features from 3D point cloud of the entire brain.

III. EXPERIMENT

In this section, we studied the WM fibers classification results accuracy and effectiveness, through various experiments performed on the training and testing datasets. To evaluate the impact of our method improvement, we compared our method with others methods in the literature.

A. Real Datasets

Training database: We used for training reference bundles called bundle atlas composed by 72 major high quality bundles semi-automatically generated from 105 Human Connectome Project (HCP) subjects [15]. This bundle atlas is openly available: <https://doi.org/10.5281/zenodo.1088277>. Those reference bundles are used as labels for training and learning shape features of bundles. The following bundles are included: Arcuate fascicle (AF), Anterior thalamic radiation (ATR), Commissure anterior (CA), Corpus callosum (Rostrum (CC1), Genu (CC2), Rostral body (CC3), Anterior midbody (CC4), Posterior midbody (CC5), Isthmus (CC6), Splenium (CC7)), Cingulum (CG), Corticospinal tract (CST), Middle longitudinal fascicle (MLF), Fronto-pontine tract (FPT), Fornix (FX), Inferior cerebellar peduncle (ICP), Inferior occipito-frontal fascicle (IFO), Inferior longitudinal fascicle (ILF), Middle cerebellar peduncle (MCP), Optic radiation (OR), Parieto-occipital pontine (POPT), Superior cerebellar peduncle (SCP), Superior longitudinal fascicle I (SLFI), Superior longitudinal fascicle II (SLFII), Superior longitudinal fascicle III (SLFIII), Superior thalamic radiation (STR), Uncinate fascicle (UF), Thalamo-prefrontal (T-PREF), Thalamo-premotor (T-PREM), Thalamo-precentral (T-PREC), Thalamo-postcentral (T-POSTC), Thalamo-parietal (T-PAR), Thalamo-occipital (T-OCC), Striato-fronto-orbital (ST-FO), Striato-prefrontal (ST-PREF), Striato-premotor (ST-PREM), Striato-precentral (ST-PREC), Striato-postcentral (ST-POSTC), Striato-parietal (ST-PAR), Striato-occipital (ST-OCC).

Testing database: We selected 10 subjects from the HCP dataset to construct testing database. The corresponding dMRI images have 1,25 mm isotropic resolution and 270 gradient directions with 3 b-values (1000, 2000, 3000 s/mm^2) and 18 $b = 0$ images [15]. The multi-shell constrained spherical deconvolution was used to extract the fODF and anatomically constrained probabilistic tractography (iFOD2) for the whole brain was performed using MRtrix [16] to generate 10 tractography datasets. each one contains 10 million fibers with a minimum length of 40 mm . These fibers were classified on bundles using our framework.

B. Training and classification

Firstly, our network learned bundle point cloud shape features from training database and established the bundle model. Secondly, this model bundle was used to classify WM fibers from testing tractography database on anatomical bundles. Our network was trained jointly to output normals and curvatures at the same time. We trained our multi-scale network with 1024 point functions h . We selected point of interest randomly for each fiber. We trained for up to 2000 epochs on our dataset, or until convergence. All our training was performed in PyTorch using stochastic gradient descent with batch size 64, a learning rate of 10^{-4} and 0.9 momentum.

C. Qualitative evaluation

To appreciate the relative quality of the WM bundles obtained with our method, we selected 5 bundles in both hemisphere (left and right), with different shapes. These bundles were obtained with our classification network from different subjects in the testing tractography database. The following bundles are included: The right cingulum (CG-R), the left Corticospinal (CST-L), the right Fornix (FX-R), the left Inferior Longitudinal Fascicle (ILF-L) and Superior Thalamic Radiation (STR-L). The figure 2 shows qualitative results for classification of some bundles with different shape colored according to their unoriented normals. We can see clearly from this figure 2 that the point of interest summarizes the skeleton of the shape. We also see that the CST-L bundle is well constructed by our method, so it is among the rare methods, that succeed the good CST bundle extraction, due to his shape resemblance with their neighbors bundles. We observe a good classification of the other bundles.

We compared our network normal prediction results to the ground-truth normals computed from bundles of the training database (figure 1). We observe a reasonable normal reconstruction. Our predictions are more smooth and continuous than the ground-truth which includes flipped normal directions in some region.

D. Quantitative evaluation

In this section, we compare our classification network with two traditional methods FiberNet [17] and RecoBundles [18]. The FiberNet [17] is based on a similar concept that our method and used Convolutional Neural Network (CNNs) to learn shape features of the fiber bundles, then used it to

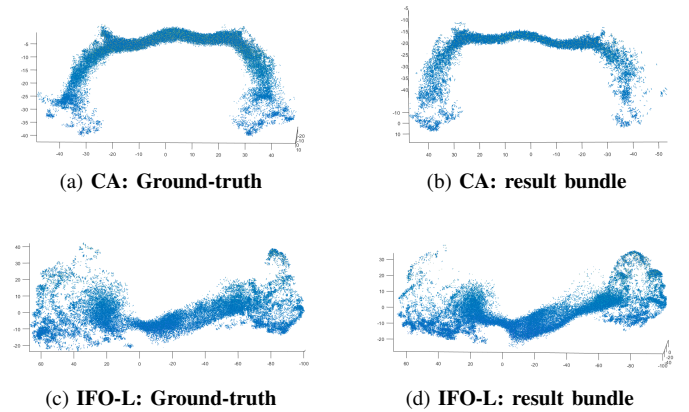


Fig. 1: Our network normal reconstruction results. In the first column we show the reconstructed normals for the points of interest in two classification result bundles and in the second column we show ground-truth normals computed on training database

classify WM fibers from other subjects. The architecture of FiberNet was implemented using TFLearn library [1], based on two convolutional layers with 32 and 64 feature maps with are rectified linear units (ReLU) as non-linearity inducing functions and the hyperbolic tangent function (tanh). To prevent overfitting, the authors used 80% dropout. The RecoBundles is fibers-based method proposed in [18] to extract corresponding bundles given by an atlas from tractography dataset. We used the default RecoBundles parameters; $reduction_{thr} = 20$, $cluster_{thr} = 15$ and $pruning_{thr} = 8$.

The table 1 shows results obtained with leave-one-out cross-validation accuracy for ten bundles chosen randomly from the testing database. Our method achieved the best and high accuracy in most bundles, with a mean accuracy of 85.2%. This result proves the capacity of our method to effectively assign the fibers to their coherent bundles. Nevertheless, the FiberNet achieved better accuracy compared to RecoBundles, with a mean accuracy of 75% and 71.3% respectively.

IV. CONCLUSION

In this study, we have presented a deep-learning approach for WM fiber classification based on estimating local 3D fiber shape properties. We considered WM fibers as 3D point clouds and we directly estimate local shape properties from raw point clouds. Various experiments were performed on the real HCP, allowing extraction of 72 WM bundles. Our method was a solution approximating the optimal and the most reasonably stable results across all experiments. This work demonstrated the advantages of using the deep learning methods in the WM classification, we plan to test other more recent methods in this contexte, such as PC-Net [19] and including other WM fiber features such as orientation, which may increase our method results accuracy.

¹<http://tflearn.org/>

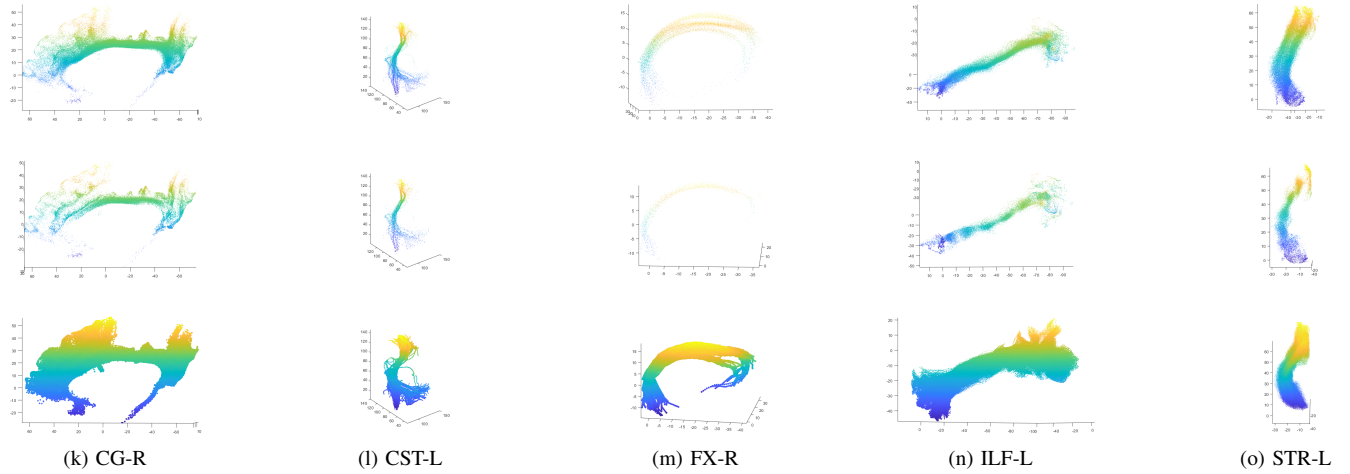


Fig. 2: Qualitative results for classification of some bundles with different shape colored according to their unoriented normals, chosen randomly from testing database. First row is input point cloud bundles of reference (ground-truth). second row is bundle points of interest. Bottom row is output classification results.

TABLE I: Accuracy results obtained for ten bundles chosen randomly from the testing database, with our method RecoBundles and FiberNet.

Method \ Bundles	Mean	CST	ILF	IFO	CG	FX	SLF	MCP	AF	ATR	OR
Our method	0.852	0.860	0.910	0.905	0.878	0.745	0.853	0.840	0.872	0.780	0.886
FiberNet	0.750	0.811	0.840	0.835	0.758	0.655	0.566	0.751	0.861	0.695	0.730
RecoBundles	0.713	0.752	0.688	0.701	0.738	0.665	0.732	0.780	0.719	0.670	0.689

REFERENCES

- [1] Basser, P., Mattiello, J., LeBihan, D., "Imagerie de diffusion in vivo par resonance magnetique nucleaire". *Notdefined* 301, 1109-1112, 1985.
- [2] Basser, P., Mattiello, J., LeBihan, D., "Mr diffusion tensor spectroscopy and imaging". *Biophysical journal* 11, 259, 1994.
- [3] Behrens, T., Berg, H., S.Jbabdi, Rushworth, M., Woolrich, M., "Probabilistic diffusion tractography with multiple fiber orientations: what can we gain?" *Neuroimage* 34, 144-155, 2007.
- [4] Basser, P., Pajevic, S., Pierpaoli, C., Duda, J., Aldroubi, A., "Bootstrap white matter tractography (boottrac)". *Neuroimage* 24, 524-532, 2003.
- [5] Catani, M., de Schotten, M.T., "A diffusion tensor imaging tractography atlas for virtual in vivo dissections". *Cortex* 44, 1105-1132, 2008.
- [6] A. Brun, H. Knutsson, Westin, C., "Clustering fiber traces using normalized cuts", in: *Proceedings in MICCAI, France*. pp.368-375, 2004.
- [7] O'Donnell, L., Kubicki, M., Shenton, M., Dreusicke, M., Grimson, W., Westin, C., "A method for clustering white matter fiber tracts". *Am. J. Neuroradiol* 27(5), 1032-1036, 2006.
- [8] Qi C. R., Su H., Niessner M., Dai A., Yan M., Guibas L.J. "Volumetric and multi-view cnns for object classification on 3d data". In *Proc. CVPR*, pp. 5648–5656, 2016.
- [9] Maron H., Galun M., Aigerman N., Trope M., Dymn., Yumer E., KIM V. G., Lipman Y. "Convolutional neural networks on surfaces via seamless toric covers". *ACM Trans. Graph.* 36, 4, 71:1–71:10, 2017.
- [10] Boscaini D., Masci J., Rodola E., Bronstein M.: "Learning shape correspondence with anisotropic convolutional neural networks". In *Advances in Neural Information Processing Systems*, pp. 3189–3197, 2017.
- [11] Wei L., Huang Q., Ceylan D., Vouga E., Li H., "Dense human body correspondences using convolutional networks". In *Proc.CVPR*, pp. 1544–1553, 2016.
- [12] Boulch A., Marlet R. "Deep learning for robust normal estimation in unstructured point clouds". *Computer Graphics Forum* 35, 5, 281–290. doi:10.1111/cgf.12983, 2016.
- [13] Qi C. R., Su H., MO K., Guibas L. J., "Pointnet: Deep learning on point sets for 3d classification and segmentation". In *Proc.CVPR*, 2017.
- [14] Guerrero P, Kleiman Y, Ovsjanikov M, Mitra NJ, "PCPNet: Learning Local Shape Properties from Raw Point Clouds", *Computer Graphics Forum*, 37, 2, 75-85, 2018
- [15] Sotiropoulos S.N, Moeller S, Jbabdi S, Xu J, Andersson J. L, Auerbach E. J, Yacoub E, Feinberg D, Setsompop K, Wald L.L, Behrens T. E. J, Ugurbil K, and Lenglet Cs "Effects of Image Reconstruction on Fibre Orientation Mapping from Multichannel Diffusion MRI: Reducing the Noise Floor Using SENSE". *Magnetic Resonance in Medicine* 00:000–000, 2013.
- [16] Tournier J.-D., Calamante F, Connelly A, "Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions". *Proceedings of the International Society for Magnetic Resonance in Medicine*, 2010.
- [17] Vikash G, Thomopoulos I, "Fibernet: An ensemble deep learning framework for clustering white matter fibers", pp. 548-555, 2017.
- [18] Garyfallidis E, Cote MA, Descoteaux, M., "Recognition of white matter bundles using local and global streamline based registration and clustering". *NeuroImage* 170(7), 2017.
- [19] Chen Z, Guan T, Luo Y, Wang Y, Luo K and Xu L, "PC-Net: A Deep Network for 3D Point Clouds Analysis," 2020 25th International Conference on Pattern Recognition (ICPR), pp. 465-472, 2021.

Non-trainable and Semi-trainable modes for COVID-19 Detection using Vision Transformers and X-ray Images

Imed Eddine Haouli¹, Walid Hariri¹, Hassina Seridi-Bouchelaghem¹

¹ *Labged Laboratory, Computer Science Department, Badji Mokhtar Annaba University, Algeria*

haouli-imededdine@hotmail.com; hariri@labged.net; seridi@labged.net

Abstract—Coronavirus Disease (COVID-19) pandemic has become an increasing threat to human life. Although Convolutional Neural Networks (CNN) have achieved reliable results to automatically detect the virus, they are still suffering from some drawbacks. CNNs are not robust against scale and rotation variations which may cause false classifications. Vision transformer (ViT) has recently demonstrated its potentials in image processing and achieved competitive results where CNN's problems are solved. Specifically, ViT appears to demonstrate excellent performance in image processing and classification when trained on sufficient data, outperforming a comparable state-of-the-art CNN with less computational requirements. Therefore, we propose in this paper to apply for the first time a ViT-based technique for COVID-19 detection. Two transfer learning modes have been conducted and tested, the simple fine-tuning (also called non-trainable mode) and the semi-trainable mode along with different two ViT architectures including ViT-B16 and ViT-B32. Finally, 2-class and 3-class classification methods are tested. Very high classification rates have been achieved compared to classical CNN-based methods.

Index Terms—COVID-19, ViT, Transformer block, CNN, Self-attention, Transfer learning.

I. INTRODUCTION

Coronavirus is a fast-spreading contagious disease that infects the respiratory system especially the lungs. During the last two years, it has caused a healthcare crisis worldwide. Despite the good detection results of the traditional reverse transcription-polymerase chain reaction (RT-PCR) to detect this virus, it is still has many limitations since it is not fast, and it can occur false-negative results [1]. Therefore, detecting the virus using radiology imaging and deep learning techniques has become the best solution that can replace RT-PCR.

Convolutional Neural Networks (CNN) are becoming more and more used to detect the virus due to their high performance that can solve the RT-PCR drawbacks [2]. One of the more challenging problems for COVID-19 detection using CNNs is the lack of training data that considerably restrains the power of deep-learning-based techniques as in other image classification tasks. Specifically, the majority publicly available databases contain a limited number of samples from COVID-19 suspected cases and do not contain sufficient data of these cases. Therefore, in order to avoid blending suspected cases with real COVID-19 cases with the considerable increase of the former cases, it has become a highly needed task to

diagnose the suspected cases to classify them according to their pneumonia, and distinguish them from real COVID-19 cases.

Recently, many researchers proposed to automatically detect the virus using machine and deep learning approaches with different strategies. The following surveys can be referred to [3]–[5]. Many drawbacks have been noticed when dealing with deep learning-based techniques such as their sensitivity to scale and rotation variations. Based on self-attention architectures, vision transformer becomes the leading model in natural language processing (NLP). For NLP, by employing attention models, i.e. transformers, training speed can be significantly improved hence enhancing the performance of neural machine translation applications. For image processing and especially medical imaging, vision transformers are emerging and starting to show potentials by applying to computer vision tasks, such as image recognition [6].

The rest of this paper is organized as follows: Section II presents the related works. The contribution of the paper is presented in Section III. An overview of our proposal as well as the main steps of our proposed method are shown in Section IV. Section V presents in detail the experimental results. Discussions are presented in Section VI. Conclusions end the paper.

II. RELATED WORKS

In the last years, deep neural networks have become the mainstream approach for diseases diagnosis and detection. Since the spread of COVID-19 in 2020, CNNs have been widely used to distinguish between positive and negative cases using different modalities including images and voice.

Transfer learning (TL) is becoming more and more popular to accurately detect the virus. It consists of transforming the knowledge of an already trained machine learning model to a different but related problem. That is to say that TL aims to employ the knowledge of a model has learned from a task with a lot of available labeled training data in a new task that doesn't have much data. For example, Jaiswal et al. [7] proposed a pre-trained model DenseNet201 to classify chest CT images on two classes, as COVID-19 infected or Not. The proposed model is divided on two parts, the first is used to extract discriminative features, by using weights

learned from ImageNet dataset, and the second its used as a classifier based on fully connected layers, ended by sigmoid function composed to two neurones. In the work made by Mamalakis et al. [8], the authors presented DenResCov-19 model, deep transfer learning network to detect the following diseases : COVID-19, tuberculosis and pneumonia on X ray images. To improve performance, they combined two models ResNet50 and DensNet by concatenate four specific block for each model. Another work presented by Apostolopoulos et al. [9], five pre-trained models VGG19, MobileNet v2, Xception, Inception and InceptionResNetV2 have been fine-tuned to automatic diagnosis COVID-19 on two public datasets, with two scenarios, the first with two classes normal cases and affected by COVID-19, in the second they added third class pneumonia.

Although CNNs are highly efficient in image classification and pattern recognition tasks, understanding their results is going to be fruitless endeavor. Vision Transformer (ViT), however, offers the possibility to divide the images into patches, providing more interpretability to these sub-images even during the training process. Also, ViT is more robust against scale and translation variations compared to CNN-based architectures. Few methods have employed ViT to detect the COVID-19 disease, for example, Mondal et al. [10] proposed to use multi-stage transfer learning using ViT on CT scans and X-ray images called xViTCOS. They showed that the ViT gives more interpretability to the learned features compared to CNN. Also, xViTCOS give higher accuracy rate 6 erreurs12 avertissements compared to InceptionV3 with 0,96 and 0,946 respectively.

Other methods have only used CT scans with ViT to detection COVID-19. For example, Gao et al. [11] applied two different deep learning strategies based on ViT and DenseNet pre-trained model. As a result; ViT has enhanced the DenseNet performance when dealing with F1 score from 0,72 to 0,76.

III. CONTRIBUTION

This paper aims to propose new system in order to automatically detect the COVID-19 virus using X-ray images. Despite the challenging results obtained using deep CNN models through different settings, they are still sensitive to some variations such as scale and translation. Hence, we propose in this paper a transformer vision-based method to deal with COVID-19 detection with two-class (COVID, normal), and three-class (COVID, normal, other pneumonia) classification problems. Hence, we show the efficiency of two different ViT models already pre-trained on a large amount of data (ImageNet) including ViT-B16 and ViT-B32, and compare it with the pre-trained ResNet152V2 [12]. To further investigate the efficiency of each model, we have carried out two scenarios, the first scenario consists of freezing all the transformer blocks (non-trainable mode), and adjusting the output layer to 2 or 3 classes. The second scenario, however, aims to train more block layers and keep the previous layers frozen. Therefore, we train gradually four blocks and show the efficiency of the whole model with 2-class and 3-class classification tasks.

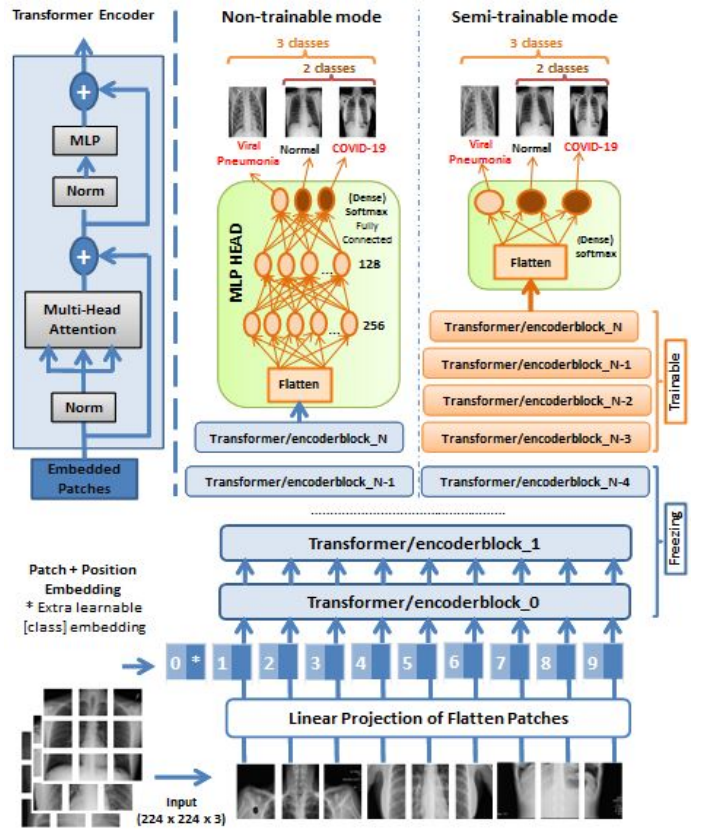


Fig. 1: Overview of our proposal.

Since our database is not balanced (viral pneumonia is a minority class). Therefore, to treat this problem, we have applied class weight algorithm to avoid biased classification. Figure 1 presents the overview of our vision transformer-based COVID-19 detection method.

IV. METHOD OVERVIEW

Our proposed method contains three main steps:

A. Pre-processing

In our approach, we have applied image resizing of 224×224 pixels, and data normalisation that refers to rescaling pixel values into the range of $[0,1]$.

Moreover, to overcome the overfitting problem and to increase the accuracy by generating new images that we call also data generation, we applied four methods including translation, shift, zoom and horizontal flipping.

B. Unbalanced dataset adjustment

Among the problems that we encounter during the automatic image classification process, especially medical images, we can find the problem of unbalanced classes. It is the case when the occurrence of one of the classes is very high compared to the other classes, results in a biased algorithm towards the majority class. To overcome this problem, we have applied in this paper *class weights technique*, it consists of giving a weight to each class depending on its

number of images. Hence, this strategy will penalize the misclassification made by the minority class by giving them a higher class weight while giving the majority class a lower weight [13]. Values of weights are then defined by the Eq.(1) :

$$W_j = \frac{N_{samples}}{(N_{classes} \times N_{samples_j})} \quad (1)$$

Where,

- W_j is the weight for each class (j represent the class).
- $N_{samples}$ represent the total number of samples.
- $N_{classes}$ represent the number of unique classes in dataset.
- $N_{samples_j}$ represent number of samples for each class.

The Eq.(2) defines the Categorical Cross-Entropy loss :

$$CCE = - \sum_{i=1}^C t_i \log(f(s)_i) \quad (2)$$

where,

- C represents the number of classes.
- t_i refers to the target label for training example i
- $f(s)_i$ represents the Softmax function defined by Eq.(3) :

$$f(s)_i = \frac{e^{s_i}}{\sum_j^C e^{s_j}} \quad (3)$$

Softmax is an activation function, not a loss, it results into a vector in the range of $[0, 1]$, the sum of all values is equal to 1, where each value represents a class probability.

After adding the class weight method through the Categorical Cross-Entropy loss, Eq.(2) is updated and it is given by Eq.(4):

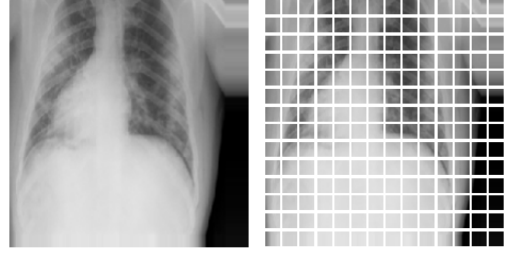
$$CCE = - \sum_{i=1}^C W_i \cdot t_i \log(f(s)_i) \quad (4)$$

Where W_i represents the weight of each class i .

C. Vision transformer-based feature extraction

Vision Transformer has mainly proposed for natural language processing in 2017 by Research team of Google [14]. Based on the remarkable success of this new representation in NLP domain, it has become a very interesting alternative to CNN models in the field of image processing and classification thanks to self-attention architectures [6]. One of the main improvements of ViT compared to CNN is its reduced use of storage equipment and computation complexity (more information about the number of parameters of ViT models are shown in Table I). Accordingly, the performance on ImageNet has been improved using ViT from 87,54% to 88,55%. Although the ViT models contain an important number of parameters as shown in Table I, they are still fast compared to their CNN counterparts. For these reasons, we have used in this paper the ViT as the main feature extractor based on their transformer blocks along with the self-attention mechanism for COVID-19 detection, and we compare its performance with ResNet152V2 as a baseline CNN. More details about the transformer block architecture can be found in the top left of the overview figure (Figure 1).

The input X-ray image, $\mathbf{x} \in \mathbb{R}^{d1 \times d2 \times C}$ is decomposed of N image patches, $\mathbf{x}_p^{(i)} \in \mathbb{R}^{P \times P \times C}$ as shown in Figure 2, where $i \in \{1, \dots, N\}$, and each path is of shape $P \times P$ in 2-D, C represents number of channels (e.g. $C = 3$ for RGB images).



(a) X-ray image

(b) Image size: (224 × 224)
Patch size: (16 × 16)
Patches per image: 196

Fig. 2: Example of patches extraction using ViT model.

D. Deep residual feature extraction

For the sake of completeness, and to compare the proposed ViT-based model, we have chosen the pre-trained ResNet152V2 [12] that is one of the last ResNet models. This model contains hundreds of CNN layers, it has shown very high performance in image classification tasks. The advantage of this model comparing to other CNNs is that it doesn't have important generalization difficulties. Moreover, with very important number of layers, it is still accurate.

E. Training modes

We present in this section the two training modes along with ViT and ResNet152V2 models.

a) Non-trainable mode: Consists of freezing all the transformer blocks where the last part of the network architecture is modified. Accordingly, we replace the last fully connected layer by a multi-layer perceptron composed of three layers: the first layer contains 256 nodes, the second contains 128 nodes and the last one contains 2 or 3 nodes representing (COVID-19, Normal) and (COVID-19, Normal and Viral pneumonia) respectively. This proposal aims to apply the transfer learning with a simple fine-tuning stage at the last part of our ViT-based model as well as the CNN of reference to compare with (i.e. ResNet152V2). The Softmax function is then used to produce the output of model in the range of $[0,1]$.

b) Semi-trainable mode: In this mode, we freeze all the first blocks and we gradually train the last four blocks to carry out the transfer learning with a four-levels semi-trainable mode. Next, we adjust the last fully connected layer (1000 nodes) of each model by replacing it with a fully connected layer of 2 nodes or 3 nodes according to our classification task (2 or 3 classes) followed by a Softmax function to make a final decision. It is worth mentioning that in this mode, we don't add more fully connected layers contrary to the first mode. This strategy is explained by the fact that the four-levels semi-trainable mode is sufficient to accurately fine-tune

TABLE I: Number of trainable parameters (in million) according to the number of the used layers on semi-trainable mode of each Model.

Number of trainable Blocks		1	2	3	4	Total
Models	ViT-B16	7	14	21	28	85,8
	ViT-B32					87,4

the model to our target dataset compared to the first mode in which we freeze all the blocks.

V. EXPERIMENTS

We have conducted our experiments on Colab PRO environment using GPU Tesla P100-PCIE-16GB and the TensorFlow/Keras framework of Python language. In this section, we present our experiments that we conducted on publicly available COVID-19 Radiography Database. To do so, we adopted two different classification strategies, *three classes* in which we take into account the positive and negative COVID-19 cases as well as other viral pneumonia. *Two classes* in which we use only positive and negative COVID-19 cases. We present the results of each scenario (non-trainable and semi-trainable).

A. Dataset description

Publicly available COVID-19 Radiography Database contains 21165 X-ray images divided on four classes COVID-19, Normal, Viral Pneumonia and Lung Opacity [15]. This database has been created by collaboration with team of researches from university of Dhaka, Bangladesh, university of Qatar, Doha and medical doctors. This dataset has been updated several times. In our experiments, we have used the last version V4 published on 2021 and we only consider the first three classes. Table II below describes our data arrangement of this dataset.

TABLE II: Data arrangement.

Classes	number of images	Train (80%)	Test (20%)
COVID-19	3616	2893	723
Normal	3616	2893	723
Viral Pneumonia	1345	1076	269
Total	8577	6862	1715

B. Evaluation criteria

We have divided the dataset into 80% for train and 20% for the test, where we have adopted the 10-fold cross-validation strategy in the training process. It consists of dividing our training data into 10 k folds, then the model is trained on nine folds successively, ended by evaluating the model on the tenth fold (unseen data). In our experiments, we have carried out 30 epochs in the training process.

C. Results on three classes

According to the experiments made on the COVID-19 Radiography Database and the obtained results shown in Table III (in this table, **Yes** means that we apply the class weight, whereas **No** means that we don't apply it). We can notice

that the impact of using the class weights can be useful to avoid biased classification. However, the improvement is not very high (about 1%). This is due to the values of the weights, finding the optimal values will be a big challenge. Also we noticed that the ViT models are more efficient than ResNet152V2 model in both scenarios (i.e. with and without class weight).

Moreover, in semi-trainable mode, all models have shown that the accuracy increases where the number trainable blocks / layers increases using ViT and ResNet152V2, respectively. Figure 3 presents the better result of each model on semi-trainable mode. We can notice from the confusion matrices that the ViT models outperform the ResNet152V2 model especially when dealing with COVID-19 cases. Therefore, ViT models are more robust and accurate to detect COVID-19 compared to residual features-based CNNs. More specifically, in the 3-class classification task, ViT-B16 and ViT-B32 have recognized 701 and 701 COVID-19 cases respectively, where the ResNet152V2 has correctly classified only 693 COVID-19 cases. Similar result has been obtained when dealing with normal cases where ResNet152V2 is slightly better with the minority class of pneumonia. Overall, the best model is ViT-B32 with an accuracy of 96,97% and precision of 99%.

Furthermore, in non-trainable mode, the results have proven that ViT-B16 and ViT-B32 outperformed ResNet152V2 with 95,16%, 94,99% and 94,58%, respectively. To give more details about the robustness of the ViT models and ResNet152V2, we show in the second and the third columns of Figure 3 the accuracy and loss of the training and the validation process in the fold 1 and the fold 10, respectively. From this figures, it is clear than ViT-based models are steadier from the fold 1 compared to ResNet152V2. Therefore, ViT models converge faster than CNN model and give higher classification rates. To further show the efficiency of the proposed models, especially using ViT models, we have compared our obtained results with other state-of-the-art methods using the same dataset. For example, Das et al. [16] has proposed a CNN with reduced number of parameters to classify COVID-19, normal and pneumonia cases. They achieved an accuracy 93,67% on older version of COVID-19 Radiography Database, where the recent version is more challenging. Another method has been proposed by Ouchicha et al. [17] using residual deep features that are based on different parallel stages of kernel sizes. They have trained all the proposed model and achieved 96,69% of accuracy of 3-class classification task. Wang et al. [18] proposed a new CNN called COVID-Net, and achieved 93,3% on a similar amount of data.

D. Binary classification

In this experiment, we tested the same dataset using only COVID-19 and normal cases. From Table IV, in the semi-trainable mode, we can see that ViT-B16 gives the highest accuracy using 2 trainable blocks with 96,96%, whereas, ViT-B32 is more efficient with 3 trainable blocks and gives 97,72%. Also, ViT-B32 outperforms ViT-B16 over the 4 gradually trainable blocks. With comparison to ResNet152V2, it is clear

TABLE III: Comparison between **ResNet152V2**, state-of-the-art methods, and two **ViT** models on Chest X-Ray images of **Normal**, **COVID-19** and **Viral Pneumonia** cases when dealing with the two scenarios of transfer learning (**non-trainable** and **semi-trainable**).

Mode Number of trainable Layers / blocks	Semi-trainable								Non-trainable	
	1		2		3		4		yes	no
Class weight / Models	yes	no	yes	no	yes	no	yes	no	yes	no
ViT-B16	93,99	93,35	93,82	94,64	93,64	95,34	96,38	96,85	95,16	93,94
ViT-B32	96,20	95,56	95,45	95,51	95,74	94,40	96,97	96,68	94,99	95,45
ResNet152V2	92,59	91,95	94,34	95,22	95,69	94,81	96,79	95,80	94,58	95,04
Das et al. [16]	/	/	/	/	/	/	/	/	/	93,67
Ouchicha et al. [17]	Fully trainable mode : 96,69								/	/
Wang et al. [18]	Fully trainable mode : 93,30								/	/

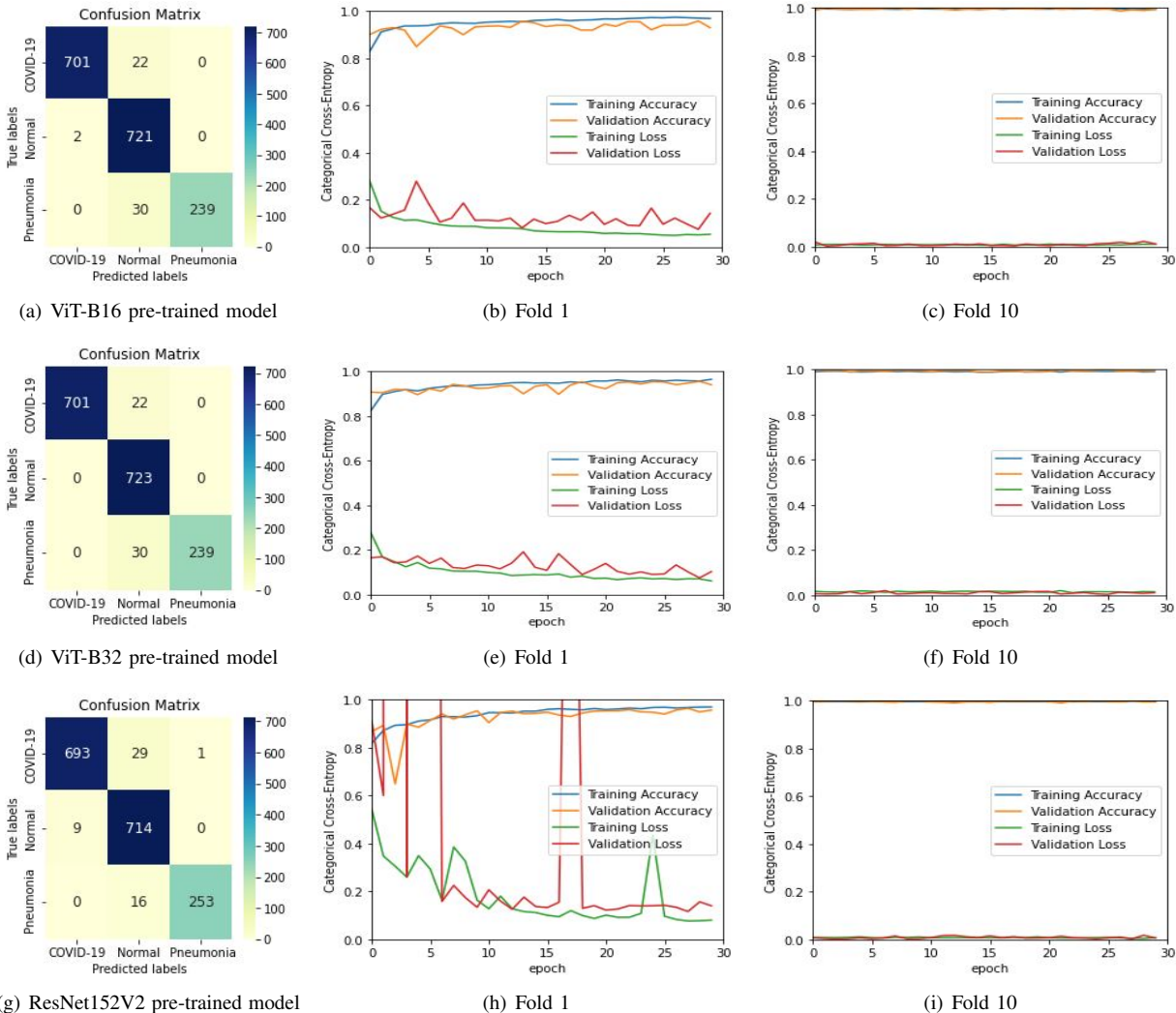


Fig. 3: Performance comparison between ViT models and ResNet152V2 CNN with 3-class classification task. From left to right, the first column presents the confusion matrix of each model, the second and the third column show the accuracy and loss of train and validation during the fold 1 and fold 10 respectively. Each row refers to the model mentioned on the left figure.

that the ViT models are more accurate over the 4 gradually trainable blocks, where ResNet152V2 requires more trainable blocks to enhance the performance (i.e. 4 blocks). In the non-trainable mode, the ViT-B32 outperforms the two other models

that are slightly similar. In comparison with other state-of-the-art methods, our proposed method outperformed Cavallo et al. [19]’s proposal that used statistical-based method and obtained 91,8%. Wu et al. [20] has proposed and trained a

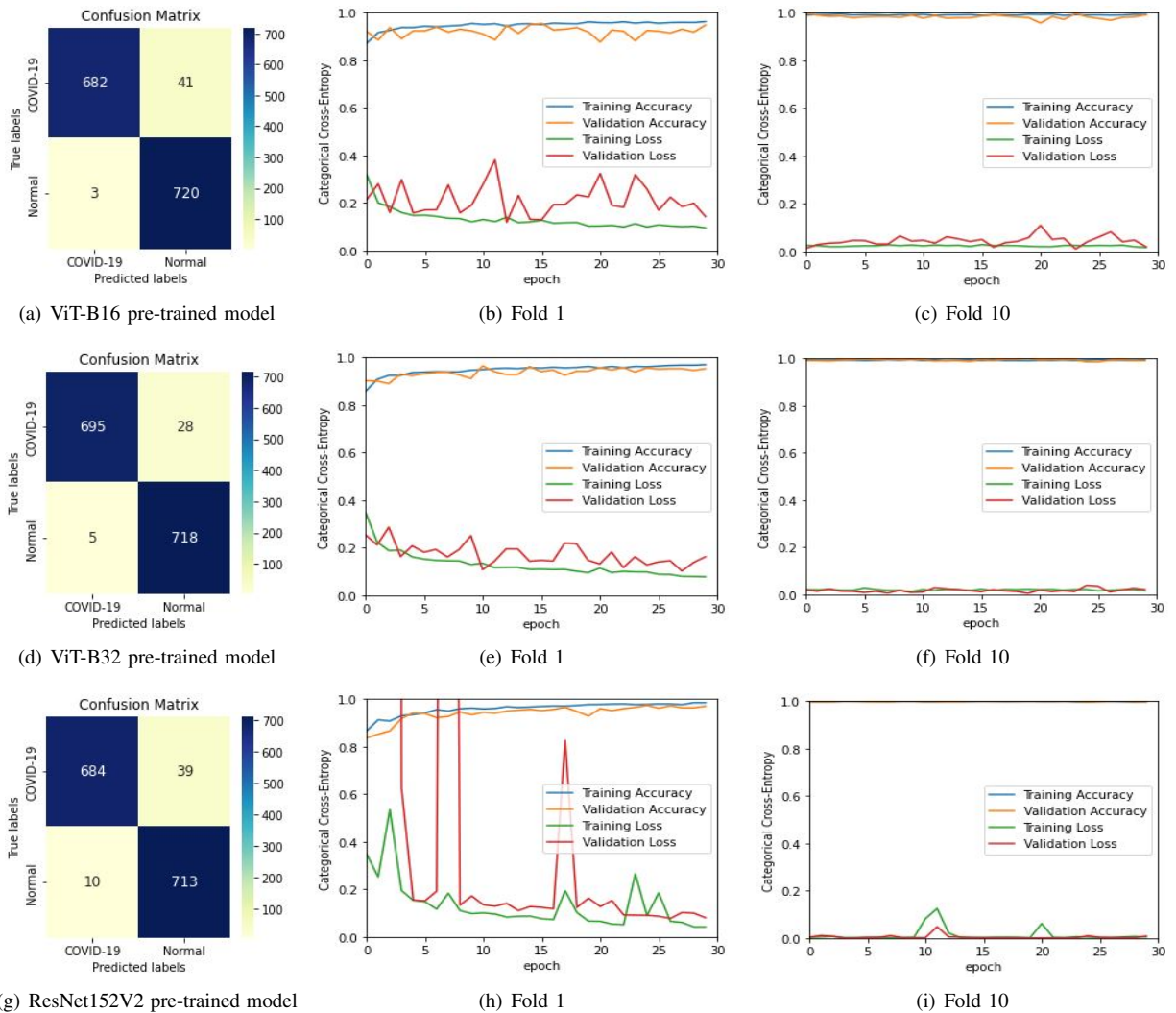


Fig. 4: Performance comparison between ViT models and ResNet152V2 CNN with 2-class classification task. From left to right, the first column presents the confusion matrix of each model, the second and the third column show the accuracy and loss of train and validation during the fold 1 and fold 10 respectively. Each row refers to the model mentioned on the left figure.

TABLE IV: Comparison between **ResNet152V2**, state-of-the-art methods, and two **ViT** models on Chest X-Ray images of **Normal** and **COVID-19** cases when dealing with the two scenarios of transfer learning (**non-trainable** and **semi-trainable**).

Mode	Semi-trainable				Non-trainable
	1	2	3	4	
Trainable layers/blocks					
Models					
ViT-B16	96,68	96,96	96,40	94,88	95,16
ViT-B32	97,16	97,03	97,72	97,44	96,75
ResNet152V2	92,25	94,54	95,64	96,61	96,20
Cavallo et al. [19]	/	/	/	/	91,8
Wu et al. [20]	Fully trainable mode : 93,08				/

To show more details about the correctly / wrongly classified images, the left column of the Figure 4 shows the confusion matrices of the three models in 2-class classification task. From this Figure, we can see that COVID-19 cases are better classified by ViT-B32 with 695 images. Normal images, however, are better classified by ViT-B16 with 720 images. The second and the third columns of Figure 4 further show the same discipline as the 3-class classification task where ViT models are steadier compared to ResNet152V2. From this figure, it is clear that in fold 1, the training and validation losses are already low over the 30 epochs, whereas, it is wobbling with ResNet152V2. To make them steady, additional epochs during the training process are required in order to ensure the convergence of the model.

new CNN based on U-net and obtained 93,08% with positive and negative classes.

VI. DISCUSSION

From the experimental results with two classification strategies (2 and 3 classes), the accuracy of ViT models is higher than one of the best selected CNNs to deal with scale and translation variations (i.e. ResNet152V2). This efficiency is also noticeable over the four gradually trainable blocks / layers of ViT models and CNN, respectively. This discipline is explained by the robustness of the vision transformer blocks that are inspired by natural language processing, where the self-attention mechanism is taken into consideration. Dividing the images into a fixed number of patches and embedding them to feed a transformer encoder further makes the ViT models more robust against rotation and scale variations compared to its CNN counterpart. More specifically, ViT-B32 slightly outperforms ViT-B16 when dealing with the two classification scenarios, and almost over all the 4 gradually trainable blocks. Also, although the number of parameters of the ViT models is relatively high, it is still very fast during the training stage compared to CNNs. Moreover, the class weight algorithm came through with solution to the problem of biased classification toward the majority class. Nevertheless, it is still possible to find optimal weights to further improve the performance when dealing unbalanced data.

VII. CONCLUSION

This paper presents a new method for COVID-19 detection using vision transformers and X-ray images. Two classification scenarios have been carried out with two different ViT-based models according to the number of patches (16 and 32). After the pre-processing stage, the images are divided into a fixed number of ordered patches. Next, these patches are then embedded to feed transformer encoders based on the self-attention mechanism. To further evaluate the efficiency of the proposed ViT-based models, we have compared their results with ResNet152V2 CNN which is the last ResNet family models. This comparison is carried out using two different scenarios (non-trainable, and semi-trainable). The first scenario is a simple fine-tuning with additional fully connected layers in the output, whereas, the second scenario aims to train gradually four blocks (from ViT) and layers (From ResNet152V2) and freeze all the previous ones. Results also show that the ViT models are more robust than ResNet152V2 and accurately classify the COVID-19 images as well as normal and pneumonia cases. In future work, we further analyze the time complexity of the ViT models and test more trainable blocks. The combination of ViT models with other CNN models is also conceivable where the class weight algorithm can be further optimized to accurately classify unbalanced data without the need to oversampling or undersampling solutions.

ACKNOWLEDGMENT

The authors would like to thank the Agency for Research Results Valuation and Technological Development, Algeria (DGRSDT).

REFERENCES

- [1] M. Teymouri, S. Mollazadeh, H. Mortazavi, Z. N. Ghale-Noie, V. Keyvani, F. Aghababaei, M. R. Hamblin, G. Abbaszadeh-Goudarzi, H. Pourghadamyari, S. M. R. Hashemian *et al.*, "Recent advances and challenges of rt-pcr tests for the diagnosis of covid-19," *Pathology-Research and Practice*, p. 153443, 2021.
- [2] F. M. Shah, S. K. S. Joy, F. Ahmed, T. Hossain, M. Humaira, A. S. Ami, S. Paul, M. A. R. K. Jim, and S. Ahmed, "A comprehensive survey of covid-19 detection using medical images," *SN Computer Science*, vol. 2, no. 6, pp. 1–22, 2021.
- [3] W. Hariri and A. Narin, "Deep neural networks for covid-19 detection and diagnosis using images and acoustic-based techniques: a recent review," *Soft computing*, pp. 1–18, 2021.
- [4] S. Bhattacharya, P. K. R. Maddikunta, Q.-V. Pham, T. R. Gadekallu, C. L. Chowdhary, M. Alazab, M. J. Piran *et al.*, "Deep learning and medical image processing for coronavirus (covid-19) pandemic: A survey," *Sustainable cities and society*, vol. 65, p. 102589, 2021.
- [5] T. Alafif, A. M. Tehame, S. Bajaba, A. Barnawi, and S. Zia, "Machine and deep learning towards covid-19 diagnosis and treatment: survey, challenges, and future directions," *International Journal of Environmental Research and Public Health*, vol. 18, no. 3, p. 1117, 2021.
- [6] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," *arXiv preprint arXiv:2010.11929*, 2020.
- [7] A. Jaiswal, N. Gianchandani, D. Singh, V. Kumar, and M. Kaur, "Classification of the covid-19 infected patients using densenet201 based deep transfer learning," *Journal of Biomolecular Structure and Dynamics*, pp. 1–8, 2020.
- [8] M. Mamelakis, A. J. Swift, B. Vorselaars, S. Ray, S. Weeks, W. Ding, R. H. Clayton, L. S. Mackenzie, and A. Banerjee, "Denrescov-19: A deep transfer learning network for robust automatic classification of covid-19, pneumonia, and tuberculosis from x-rays," *arXiv preprint arXiv:2104.04006*, 2021.
- [9] I. D. Apostolopoulos and T. A. Mpesiana, "Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks," *Physical and Engineering Sciences in Medicine*, vol. 43, no. 2, pp. 635–640, 2020.
- [10] A. K. Mondal, A. Bhattacharjee, P. Singla, and P. AP, "xvitcos: Explainable vision transformer based covid-19 screening using radiography," 2021.
- [11] X. Gao, Y. Qian, and A. Gao, "Covid-vit: Classification of covid-19 from ct chest images based on vision transformer models," *arXiv preprint arXiv:2107.01682*, 2021.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*. Springer, 2016, pp. 630–645.
- [13] Y. Ho and S. Wookey, "The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling," *IEEE Access*, vol. 8, pp. 4806–4813, 2019.
- [14] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, E. Kaiser, and I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.
- [15] T. Rahman, "COVID-19 Radiography Database," "<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database/>", 2020, [Online; accessed 15-August-2021].
- [16] A. K. Das, S. Kalam, C. Kumar, and D. Sinha, "Tlcv-an automated covid-19 screening model using transfer learning from chest x-ray images," *Chaos, Solitons & Fractals*, vol. 144, p. 110713, 2021.
- [17] C. Ouchicha, O. Ammor, and M. Meknassi, "Cvdnet: A novel deep learning architecture for detection of coronavirus (covid-19) from chest x-ray images," *Chaos, Solitons & Fractals*, vol. 140, p. 110245, 2020.
- [18] L. Wang, Z. Q. Lin, and A. Wong, "Covid-net: A tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images," *Scientific Reports*, vol. 10, no. 1, pp. 1–12, 2020.
- [19] A. U. Cavallo, J. Troisi, M. Forcina, P. Mari, V. Forte, M. Sperandio, S. Pagano, P. Cavallo, R. Floris, and F. Geraci, "Texture analysis in the evaluation of covid-19 pneumonia in chest x-ray images: a proof of concept study," 2020.
- [20] T. Wu, C. Tang, M. Xu, N. Hong, and Z. Lei, "Ulnet for the detection of coronavirus (covid-19) from chest x-ray images," *Computers in Biology and Medicine*, p. 104834, 2021.

Self-organizing Algorithm based on Multi-agent system for White Matter fibers Clustering with Outliers Elimination

CHEKIR AMIRA

*Laboratoire de Robotique, Parallélisme et systèmes Embarqués (LRPE)
Université des sciences et de la technologie Houari-Boumediène (USTHB)
Algiers, Algeria
chekir.amira@gmail.com*

Abstract—The study of white matter (WM) connectivity is of general interest in neuroscience, which is achieved by the analysis and clustering of the streamlines made the tractography dataset. However, the WM streamlines clustering is a challenge, due to the presence of outliers, the complexity, and the huge size of the WM tractography dataset. In this paper, we propose a new framework of distributed multiagent, improving, and adapting a bio-inspired model called Multiple Species Flocking (MSF) for WM streamlines clustering and automatic outlier elimination. Specifically, each streamline is associated with a mobile agent. The agents move onto 3D virtual space to form a group following defined rules. Only agents attributed to similar streamlines form a cluster, whereas the agents attributed to dissimilar streamlines are sidelined and considered as outliers. The experimental results on real datasets affirm our approach accuracy and scalability to large datasets.

Index Terms—White Matter Clustering, Outliers detection, Multiple Species Flocking Model, Swarm intelligence, multiagent.

I. INTRODUCTION

Diffusion-Weighted Magnetic Resonance Imaging (dMRI) offers a unique approach to study in vivo the structure of brain tissues, allowing the non-invasive reconstruction of brain bundle trajectories using tractography. The tractography dataset presents high complexity and is very huge, containing over a million fibers for the whole brain.

How to analyze and quantify the structures defined by tractography, and in particular, how to extract the streamline bundles is, therefore, an unsolved problem. Since the development of tractography, several methods have been proposed to analyze WM streamlines and cluster them automatically. We can divide the existing automated clustering method into two categories: connectivity-based and streamline based.

The connectivity-based methods use anatomical knowledge, by specifying some regions of interest (ROIs) based on domain knowledge and select streamlines that pass through these predefined ROIs [1]. This kind of method is limited, especially in terms of time-consuming and the necessity of anatomical knowledge.

Therefore, streamline based connectivity methods are most promising. They use the affinity matrix representing pairwise

distances between streamlines. These methods are fully automatic but are limited by the number of streamlines they can analyze. Indeed, their complexity increases quadratically as the number of streamlines increases [2]–[4].

On the other hand, the tractography dataset contains a large number of false streamlines due to data quality, tractography errors, and other factors. These false streamlines are considered as anatomical outliers affecting connectivity, thus, the WM clustering methods must be robust to the outliers to find valid bundles.

A few recent WM clustering methods incorporate outliers rejection; the authors in [5] used generalized Expectation-Maximization (EM). P. Guevara et al in [4] used a track density map, whereas they only consider the cluster size. C. Ros et al in [6] incorporated Local Outlier Factors (LOFs) in their clustering technique, based on a relative degree of streamlines isolation to eliminate outliers. L. Doderio et al in [7] detected outlier streamlines using the Gaussian curve and statistical test. Recently, M. Côté et al in [8] proposed a probabilistic outliers removal based on a hierarchical utilization of QB [2]. However, the results of these methods depend strongly on a manually fixed outlier threshold defined or requires knowing the number of clusters.

A recent method [9] developed the deep reinforcement learning Track-to-Learn framework. The framework allows for data-driven improvements over classical methods by making a learning agent improve its tracking abilities iteratively via trial-and-error.

In the last few years, many clustering approaches based on the bio-inspired paradigm have been proposed in the literature to deal with the limitations that exist in traditional clustering methods. The Flocking algorithm is the most efficient. The authors in [10] studied the basic flocking model proposed in [11] and extend it to a Multiple Species Flocking model (MSF) model applied to text clustering. Another methods [12] combined the MSF model with DenStream algorithm [13] for a data stream clustering. More recently, the authors in [14] proposed a multi-agent clustering algorithm applying the MSF rules to detect anomalies in distributed data streams such

as intrusions, faults, and system failures. In this method the agent moves into 2D virtual space following the MSF rules, specifically, all the rules - cohesion, alignment and separation - are applied to form a flock with similar agents, otherwise when two agents are dissimilar only the separation rule is applied.

In this work, we extend and adapt the method proposed in [14], by modifying the manner of applying the MSF rules and extend it to 3D space, for WM inter-subject streamlines clustering with automatic outliers elimination. In the best of our knowledge, the flocking approaches have never been applied to the WM brain study, except [15], which used the flocking model for brain tractography and another work in [16], introduced 3D-SSF algorithm, a dynamic multi-subject clustering framework, considering the WM streamlines from various subjects as a sequential data stream. This streamline stream is divided into chunks. Each chunk is operated independently, based on a distributed multiagent implementation of the MSF model.

Precisely, Each streamline from tractography dataset is assigned to a mobile agent. Agents are deployed into a 3D space and work simultaneously following the new rules defined in this work, simulating MSF model. When an agent encounters other agents in its neighborhood, it evaluates the similarity measure and then it decides whether to form a flock or no. If they are similar, we apply the cohesion rule in addition to similarity velocity, to maintain them in the same cluster. Otherwise, if they are dissimilar, we apply the separation and dissimilarity rule to more isolate them. Furthermore, the agent interacts only with the other agents present in its visibility range, it does not compare itself with all the other agents. Thus a reduced number of distance computations are performed. Such a number depends on how many agents it encounters during its movement in the virtual space and we do not need to calculate all pairwise, allowing to our method to achieve a linear speedup concerning the input size. The appartenance of the agent to the flock is not definitive, if during the space exploration a more similar agent is encountered, the current flock can be dropped and the agent can join another more similar flock, contrary to the existing methods in the literature, that are all statics, once a streamline is affected to the cluster, its stay in this one. This improvement provides significant benefits over previous work; it gives at our method the capacity to form consistent bundles quickly and increases the clustering accuracy, while, pushing away automatically outlier streamlines.

The paper is organized as follows: Section II describes the details of our method, followed by data acquisition and Pre-processing. Section III describes the evaluation of clustering results and outlier robustness. Lastly, Section IV presents discussion and conclusion of our method.

II. METHOD AND MATERIAL

A. Self-organizing Algorithm Based on MSF model

The proposed self-organizing algorithm is an automatic, bio-inspired WM clustering using distributed agents with distinct simple functionalities to mimic the flocking behavior. Each

streamline from a tractography dataset is assigned to an agent. The agent is represented by his position on 3D space (x, y, z) , his characteristic, and his velocity vector $\vec{V}_{velocity}$. Agents deployed randomly into 3D virtual space, They move and work together according to the new rules establish in our work. We modified the MSF model proposed in [10] and adapted it to our context. We introduced a new similarity function and we used it otherwise, instead of moving agents following the velocity vector calculating by the sum of the four rules described in [10], each agent applied two rules only depending on the case. Then, in the case of two dissimilar agents, the separation rule with the dissimilarity velocity was applied to isolate dissimilar streamlines and distinguish automatically the outliers. Whereas, regarding two similar agents, the distance between them was verified. If they are separated by an optimal distance (superior to the minimum distance R_2 that must be maintained among agents to avoid collision), the cohesion rule and the similarity velocity were applied to bring them together in the same flock. However, if they are too close (distance between them below R_2), only the alignment rule was applied, to keep them in this good stat. When an agent meets another in its visibility range, the similarity between them is compared. Agents that share similar streamline vector features will group and form a flock, while dissimilar agents will be moved away from the flock. Agents use the similarity function, called Minimum average Direct-Flip (MDF) proposed in [2]. It is assumed that two agents, A_1 and A_2 , are similar if their MDF distance was inferior to a fixed threshold (ϵ).

The algorithm exhibits an adaptive behavior, an agent can leave his group and join another group containing agents with higher similarity during his movement. Thus, during this predefined number of iterations, the streamlines join and leave the cluster forming different flocks. At the end of a prefixed number of iterations, each created flock is recuperated representing final bundles, and outliers are automatically eliminated. The following pseudo code [1] presents a global concept of the proposed algorithm.

B. Data Acquisition and Pre-processing

1) *Real Dataset:* We used for different experiments the real dMRI dataset from the Human Connectome Project (HCP) [17], [18], we choose randomly 10 healthy subjects (100307, 124422, 161713, 199655, 201111, 239944, 245333, 366446, 528446 and 856766). The dataset was acquired on a Siemens Skyra 3T scanner (90 gradients; b-value= $1000s/mm^2$; anatomical scan $(1,25 \times 1,25 \times 1,25mm^3)$). We reconstructed the tractography datasets for each subject, using a probabilistic tracking algorithm, applied to the CSD model [19]. The CSD model is based on Spherical Deconvolution (SD), which provides an estimation of the Fiber Orientation Distribution (FOD). The response function was measured using voxels with an FA value greater than 0.7 and a maximum harmonic order of 8 was used. The reconstruction and tractography were generated by DIPY software [1] [20].

¹<https://dipy.org/>

Algorithm 1 Pseudo-code of the self-organization algorithm

```

1: for  $i = 1$  to  $MaxIterations$  do
2:   for each agent  $A_1$  do
3:     if  $d_{direct}(A_1, A_2) < R_1$  ( $A_1$  contains in its
       visibility range an agent  $A_1$ ) then
4:       if  $A_1$  is dissimilar with  $A_2$  then
5:         applying the separation and dissimilarity
           rules
6:         Move the agent following  $V_{separation} +$ 
            $V_{dissimilarity}$  [10]
7:       end if
8:       else if  $A_1$  is similar to  $A_2$  then
9:         if  $d_{direct}(A_1, A_2) > R_2$  then
10:        applying the cohesion and similarity rules
11:        Move the agent following  $V_{cohesion} +$ 
            $V_{similarity}$  [10]
12:       else
13:         applying the alignment rule
14:         Move the agent following  $V_{alignment}$  [10]
15:       end if
16:     end if
17:   end for
18: end for

```

The obtained tractography dataset for each subject contains approximately 8×10^5 streamlines.

III. EXPERIMENTAL RESULTS

A. Evaluation of Clustering Results

The main problem with the WM clustering is the validation of any method on real datasets, due to the lack of ground truth. To evaluate the performance of method results on the real dataset, a WM atlas must be constructed. The spatial agreement between obtaining clusters and the atlas bundles called bundles of interest (BOI) was determined.

1) *White matter atlas creation:* We constructed our bundle atlas automatically using white matter query language (WMQL) tool [21], which is considered as our gold standard. WMQL segments tractography datasets of the ten subjects according to the parcellation in FreeSurfer and the WM bundles definitions. The HCP data of all the subjects are in the MNI 152 standard space, the atlas creation can be done without any inter-subject registration. We have obtained for each subject 37 neuroanatomical BOIs; such as projection, association, and commissural bundles. Afterward, we generated a WM atlas containing all these BOIs using a method described in [22].

2) *Quantitative evaluation:* We compared our method with alternative techniques in literature such as a popular WM clustering approach called QuickBundle (QB), proposed by E. Garyfallidis et al in [2] because it is a powerful method able to cluster tractography dataset of any size. We compared our method also with the basic flocking clustering method proposed by X. Cui et al in [10] to evaluate the impact of our method improvement. The clustering on all real ten tractography datasets was successfully performed using the

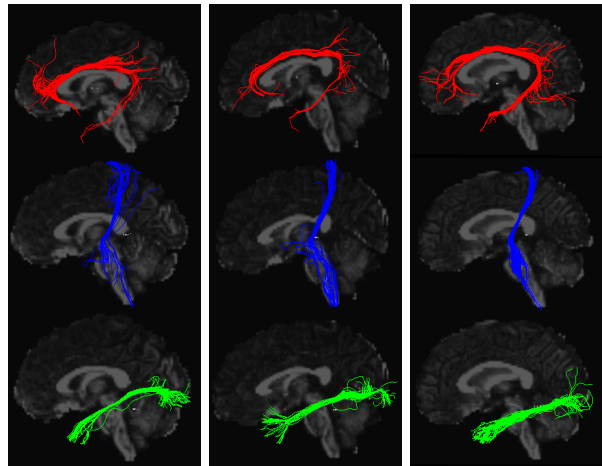


Fig. 1. Example of WM bundles results from our clustering method on several subjects with different colors: The left cingulum (CG-L) in red, the left corticospinal (CST-L) in blue, the left inferior longitudinal fasciculus (ILF-L) in green. Each column illustrates the bundles of a specified subject.

three clustering techniques (our method, QB, and basic flocking clustering (BFC)). We have obtained for each subject 37 neuroanatomical bundles namely; left/right corticospinal tract (CST), left/right cingulum (CG), left/right arcuate fasciculus (AF), left/right inferior longitudinal fasciculus (ILF), left/right inferior fronto-occipital fasciculus (IFO), left/right uncinate fasciculus (UNC), 1, 2, and 3 left/right superior longitudinal fasciculus (SLF1, 2, 3), and finally, corpus callosum (CC) in its seven parts (CC1-7). Figure 1 shows examples for bundles obtained with different subjects randomly chosen using our clustering method.

The QB method was applied using the same similarity function, used in our method, the MDF distance with a clustering threshold of $15mm$. The MDF distance requires streamlines to have the same number of points, streamline has been automatically downsampled to have only 21 points. In basic flocking clustering, each streamline was assigned to an agent and the other parameters were kept the same as the original method in [10]. For our method, we defined a set of optimal parameters. More specifically, the similarity threshold ϵ , radius of the boids visibility range R_1 , the minimum distance that must be maintained among boids R_2 , and we observed that 100 iterations were sufficient for our approach to generating good clustering results from WM tractography datasets. To quantitatively evaluate the proposed framework, the spatial agreement between results bundles and atlas bundles (BOIs) was determined after superimposing their probabilistic maps. For each subject, the 37 bundles result from every tractography dataset were matched to the atlas BOIs. Then we estimated the spatial matching ratio (SMR), spatial volume agreement (SVA) and kappa statistics between the bundles result from our method, QB, and BFC methods, to determine the best matching clustering results for each bundle. The average spatial agreement of result bundles and atlas BOIs cross the ten subjects with the three different methods (Our

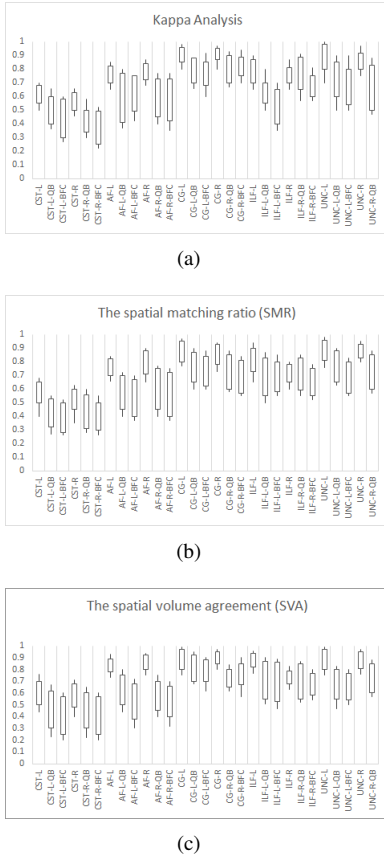


Fig. 2. The average spatial matching comparison between bundles of the ten subjects obtained with different methods (for example CST-L obtained with our method, CST-L-QB obtained with QB and CST-L-BFC obtained with BFC). (a) Kappa analysis. (b) spatial matching ratio (SMR) and (c) spatial volume agreement (SVA).

method, BFC and QB) is shown in the figure 2. The figure 2(b) shows the spatial matching ratio (SMR), the figure 2(c) shows the spatial volume agreement (SVA) and figure 2(a) shows the Kappa analysis.

The three clustering techniques were able to group streamlines into bundles that had a high spatial agreement with the BOIs of atlas (see Figure 2). Nevertheless, major differences between the clustering methods were observed. For every result bundle, the variability of the spatial agreement for our clustering method is considerably lower compared to clustering with QB and BFC methods.

By inspecting the average spatial agreement of our method and QB, a higher spatial agreement is remarked for our method in most bundles. The average SMR, SVA, and kappa for all bundles of the ten subjects are respectively 78%, 79% and 77% for our method versus 63%, 64% and 64% for QB clustering.

On the other hand, the results with basic BFC are more diverse and the spatial agreement was lower for all bundles, compared with our method. The average SMR, SVA, and kappa for basic flocking clustering are 59%, 64% and 60%.

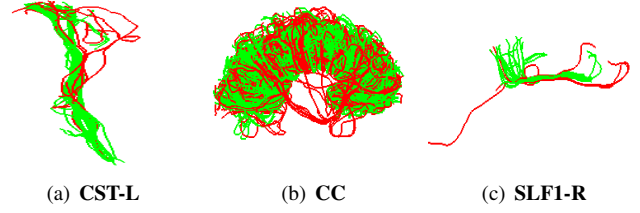


Fig. 3. Demonstration of the outlier elimination strategy robustness on real datasets. Example of bundles obtained illustrated with different colors, in green the bundles results and outlier streamlines in red. CST-L: Cortico-Spinal Tract Left ,CC: Corpus Callosum, SLF1-R: Superior Longitudinal Fascicle I Right

B. Outlier Robustness

To demonstrate the benefits of our outlier elimination strategy, our method was performed on the synthetic dataset, with introducing four additional noisy streamlines considered as outliers. We quantitatively evaluated the performance of results using some clustering validation indexes, which require a ground truth, such as silhouette coefficient, completeness and homogeneity [23].

The figure 3 shows the capacity of our method to reproduce effectively ground truth of some bundles on real datasets, thanks to our outliers elimination strategy. Streamlines component bundle results in green color and streamlines considered as outliers in red color.

IV. CONCLUSION

In this study, we have proposed a new bio-inspired algorithm, by modify and adapting the MSF rules to our context. The proposed algorithm was applied for WM streamlines clustering into anatomically meaningful bundles with outliers elimination. Precisely, each streamline from the tractography dataset is assigned to a mobile agent. Agents work autonomously and interact only with the immediate neighbor entities following the MSF rules defined in this work. These specifications make our method scalable for large tractography dataset and allow automatic streamline outliers elimination . Furthermore, both of the similarity measure and the dynamism concept gave the agent the capacity to leave his flock for another one containing agents with higher similarity, helping to get more precise clustering results. All these contributions in a very reasonable running time. Our method is a complete solution address WM clustering issues in a satisfying way. Various experiments were performed on real datasets. The first one, a quantitative evaluation of our clustering method accuracy, compared with alternative techniques in the literature. This experiment used the spatial agreement between results bundles and atlas bundles (BOIs), that is constructed in this study. Possible directions for future work, include given the distributed mobile agents, working following the defined rules, the capacity to learn the WM environment and integrate this anatomical knowledge to reinforce the decision criteria. This may improves the WM clustering results.

REFERENCES

- [1] M. Catani and M. T. de Schotten, "A diffusion tensor imaging tractography atlas for virtual in vivo dissections," *Cortex*, vol. 44, pp. 1105–1132, 2008.
- [2] E. Garyfallidis, M. Brett, M.-M. Correia, G.-B. Williams, and I. Nimmo-Smith, "Quick-bundles, a method for tractography simplification," *Frontiers in Neuroscience*, vol. 6, p. 175, 2012.
- [3] A. Demir, "Online agglomerative hierarchical clustering of neural fiber tracts," in *Proceedings in IEEE EMBC*, (Osaka, Japan), pp. 85–88, 2013.
- [4] P. Guevara, D. Duclap, C. Poupon, L. Marrakchi-Kacem, and P. Fillard, "Automatic fiber bundle segmentation in massive tractography datasets using a multi-subject bundle atlas," *Neuroimage*, vol. 61, p. 1083–1099, 2012.
- [5] U. Ziyang, M. Sabuncu, W. Grimson, and C.-F. Westin, "Consistency clustering: A robust algorithm for group-wise registration, segmentation and automatic atlas construction in diffusion mri," *International journal of computer vision*, vol. 85(3), pp. 279–290, 2009.
- [6] C. Ros, D. Gllmar, M. Stenzel, H.-J. Mentzel, and J.-R. Reichenbach, "Atlas-guided cluster analysis of large tractography datasets," *PLoS ONE*, vol. 10, p. 1371, 2013.
- [7] L. Doderio, S. Vascon, V. Murino, A. Bifone, A. Gozzi, and D. Sona, "Automated multi-subject fiber clustering of mouse brain using dominant sets," *Front Neuroinform*, vol. 8, p. 87, 2015.
- [8] M.-A. Cote, E. Garyfallidis, H. Larochelle, and M. Descoteaux, "Cleaning up the mess: tractography outlier removal using hierarchical quick-bundle clustering," in *Proceedings of: International Society of Magnetic Resonance in Medicine (ISMRM)*, (Toronto, Canada), 2015.
- [9] A. Theberge, C. Desrosiers, M. Descoteaux, and P.-M. Jodoin, "Track-to-learn: A general framework for tractography with deep reinforcement learning," *Medical Image Analysis*, vol. 72, p. 102093, 2021.
- [10] X. Cui and T. E. Potok, "A distributed agent implementation of multiple species flocking model for document partitioning clustering," in *Proceedings in Cooperative Information Agents*, (Edinburgh, UK), pp. 124–137, 2006.
- [11] C. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *Proceedings in M.C. Stone (Ed.)*, (SIGGRAPH), pp. 25–34, 1987.
- [12] Y. Li, M. Li, J. Shao, and G. Wang, "Study of data stream clustering based on bio-inspired model," *Advanced Science and Technology Letters*, pp. 412–418, 2014.
- [13] F. Cao, M. Ester, W. Qian, and A. Zhou, "Density-based clustering over evolving data stream with noise," in *Proceedings in SIAM international conference on data mining*, (Bethesda), pp. 326–337, 2006.
- [14] A. Forestiero, "Bio-inspired algorithm for outliers detection," *Multimed Tools*, vol. Appl 76, p. 25659–25677, 2017.
- [15] R. Aranda, M. Rivera, and A. Ramirez-Manzanares, "A flocking based method for brain tractography," *Medical Image Analysis*, vol. 18(3), pp. 515–530, 2014.
- [16] A. Chekir, S. Hassas, M. Descoteaux, M. Côté, E. Garyfallidis, and F. Oulebsir-Boumghar, "3d-ssf: A bio-inspired approach for dynamic multi-subject clustering of white matter tracts," *Comput Biol Med*, pp. 10–21, 2017.
- [17] D. V. Essen, M. Glasser, D.L. Dierker, J. Harwell, and T. Coalson, "Parcellations and hemispheric asymmetries of human cerebral cortex analyzed on surface-based atlases," *Cereb. Cortex*, vol. 22, pp. 2241–2262, 2011.
- [18] S. Sotiropoulos, S. Moeller, S. Jbabdi, J. Xu, J. Andersson, E. Auerbach, E. Yacoub, D. Feinberg, K. Setsompop, L. Wald, T. Behrens, K. Ugurbil, and C. Lenglet, "Effects of image reconstruction on fibre orientation mapping from multichannel diffusion mri: Reducing the noise floor using sense," *Magnetic Resonance in Medicine*, vol. 70(6), p. 1682, 2013.
- [19] I. Aganj, C. Lenglet, G. Sapiro, E. Yacoub, K. Ugurbil, and N. Harel, "Reconstruction of the orientation distribution function in single- and multiple-shell q-ball imaging within constant solid angle," *Magnetic Resonance in Medicine*, vol. 64(2), pp. 554–566, 2010.
- [20] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. van der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, "Dipy, a library for the analysis of diffusion mri data," *Frontiers in Neuroinformatics*, vol. 8, p. 720., 2014.
- [21] D. Wassermann, M. Nikos, R. Yogesh, S. Martha, K. Ron, K. Marek, and W. Carl-Fredrik, "The white matter query language: a novel approach for describing human white matter anatomy," *Brain Structure and Function*, vol. 1(17), pp. 1863–2661, 2016.
- [22] K. Hua, J. Zhang, S. Wakana, H. Jiang, X. Li, D. Reich, P. Calabresi, J. Pekar, P. van Zijl, and S. Mori, "Tract probability maps in stereotaxic spaces: Analyses of white matter anatomy and tract-specific quantification," *NeuroImage*, vol. 39, pp. 336–347., 2008.
- [23] B. Moberts, A. Vilanova, and J.-J. van Wijk, "Evaluation of fiber clustering methods for diffusion tensor imaging," in *Proceedings IEEE Visualization*, (Minneapolis MN, USA), pp. 65–72, 2005.

Human Detection Method based on Combined Face and Upper-Body Part Detectors

Salim Iratni, Merzouk Younsi, Moussa Diaf

*Laboratoire Vision Artificielle et Automatique des Systèmes (LVAAS)
Mouloud Mammeri University (UMMTO); Tizi-Ouzou – Algeria*

salim.iratni@ummto.dz
merzouk.younsi@ummto.dz
moussa.diaf@ummto.dz

Abstract–Human detection is an important issue in computer vision, with numerous applications in several fields, starting from video surveillance, human-robot interaction to driving assistance systems. Hence, in the recent years, different approaches for human detection have been proposed in the literature. However, many of these approaches are holistic-based, which means that they consider the human body as a whole, so that, they are not able to handle some specific challenges, such as occlusions and non-rigidity of the human body. Thus, in this paper, we propose a novel part-based human detection method using combined face and upper-body part detectors. This method consists mainly of two parts. In the first one, the face and upper-body part regions are independently detected by using two detectors based on Haar-like features and AdaBoost cascade classifier. Then, in the second, the outputs of the two detectors are combined using a simple rule-based approach to decide whether a human is present or not in the image. Experiments are performed on our created dataset, and the obtained results as well as the comparative study with other state-of-the-art methods have demonstrated the superiority and excellent performances of the proposed method.

Keywords– Human detection, Upper-body part, Haar-like features, AdaBoost cascade classifier.

I. INTRODUCTION

Human detection is defined as the problem of localization of all objects having a shape of a human in still images or image sequences, and it is actually among the most popular research areas in the fields of computer vision and machine learning due to its wide range of applications in many domains, such as video surveillance [1], human-robot interaction [2], smart homes [3], driver assistance system [4], autonomous vehicles [5], health care monitoring [6], image and video content retrieval [7]. However, detecting humans from real-world images or image sequences is still a challenging task due to a number of significant factors, such as illumination changes, non-rigidity of the human body, background clutters, variability in clothing, occlusions, shadows, view-point changes, and variations related to the skin color. Examples of these factors are shown in Fig. 1.

In order to overcome some of the aforementioned challenges and attain high performances, several approaches for human detection have been proposed in the literature. These approaches can be divided into two main categories [8], namely holistic-based and part-based approaches. In the holistic-based approaches, the human body is taken as a whole, and global features from its image region are extracted.



Fig. 1 Examples of human detection challenges. (a) Illumination change. (b) View-point change. (c) Shadow. (d) Occlusion.

In [9], for example, Beiping and Wen used the Histogram of Oriented Gradient (HOG) [10], and they combined it with a Support Vector Machines (SVM) classifier for detecting humans. In [11], Hoang et al. extracted parallelogram Haar-like features from the input image, and then fed them into a cascade structure classifier for the human detection. Ko et al. [12], for their part, combined three types of Center-Symmetric Local Binary Patterns (CS-LBP) descriptors with three-level cascade of random forest classifier. Nigam and Khare [13] integrated central moments with multi-resolution gray scale invariant LBP for human shape representation, and then used SVM to discriminate human from non-human objects. Umar et al. [14] used Speed-Up Robust Feature (SURF) and SVM classifier to detect humans from aerial images. Prakash et al. [15] combined two feature types, the biorthogonal wavelet transform and Zernike moments, and for the classification, they compared two classifiers, AdaBoost and SVM. In [16], Bastian and Jiji combined Aggregate Channel Features (ACF) and Histograms of Sparse Codes (HSC) for human detection. Since the holistic-based approaches do not consider the spatial relationships between the human body parts, they are not able to handle specific challenging situations such as occlusions and large variations due to non-rigidity of the human body. In order to overcome these limitations, part-based approaches have been proposed. In these approaches, the region of the human body is first divided into different parts such as arms, torso, legs, etc., then, separate detectors are applied to detect each part independently. Learned models or geometric constraints between the detected body parts are finally used to achieve the human detection task. Li et al. [17], for example, divided the human into three parts, namely head-shoulder, torso and legs, and then used an And-Or graph grammar

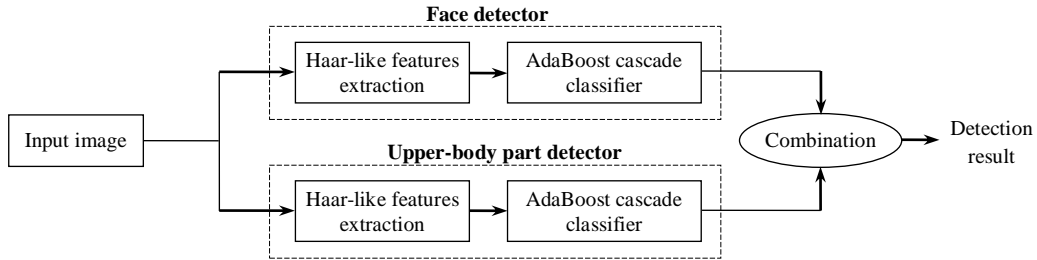


Fig. 2 Flowchart of the proposed human detection method.

model in combination with two features, HOG and Active Basis Mode (ABM), to capture large variations in the human pose and appearance. Yao et al. [18] represented the human body as a symbolic sequence. The different human body parts such as head, shoulders and waist are considered as letters in the alphabet, whereas the different poses are considered as words. A part validation for the human detection is performed using a string matching. To address the view-variance and partial occlusions, Ko et al. [19] extracted a random set of rectangular parts from three views, namely frontal-rear, left profile, and right profile. These parts are generated, as Gaussian distribution from the upper, middle, and lower region of the human body. The human detection is performed using trained part classifiers based on random forests. In [20], Tian et al. proposed a method, named DeepParts, that generates proposals using the LDCF (Locally Decorrelated Channel Features) method [21] and then trains a set of Convolutional Neural Network (CNN) to detect the different parts of the human body. Ouyang et al. [22] also used a deep CNN model that they combined with part filters to reduce imperfections of the part detectors. The part-based approaches can deal effectively with occlusions and non-rigidity of the human body than the holistic-based approaches, since they do not require the whole information about the human body to perform the detection. Motivated by these advantages, in this paper, we propose a novel part-based human detection method using combined face and upper-body part detectors. This method consists mainly of two parts. In the first one, the face and upper-body part regions are independently detected by using two detectors based on Haar-like features and AdaBoost cascade classifier. Then, in the second, the outputs of the two detectors are combined using a simple rule-based approach to decide whether a human is present or not in the image. Experiments are performed on our created dataset, and the obtained results as well as the comparative study with other state-of-the-art methods have demonstrated the superiority and excellent performances of the proposed method.

The rest of the paper is organized as follows. In Section II, we present the details of the proposed method. In Section III, we describe the experiments performed to evaluate the performance of the proposed method. In Section IV, we conclude the paper and mention some perspectives and improvements of the presented work.

II. PROPOSED METHOD FOR HUMAN DETECTION

The method we propose consists of three main steps, namely face detection, upper-body part detection, and combination. Fig. 2 shows the flowchart of this method. The

first and second steps also consist of two sub-steps, Haar-like features extraction, and classification using AdaBoost (Adaptive Boosting) algorithm. The details of these steps are described in the following subsections.

A. Haar-like features extraction

Feature extraction is the first and most important step of any human detection system, and which aims to convert the input image into a set of feature vectors that can be exploited by a machine learning algorithm. In order to achieve this objective, in this paper, we used the Haar-like features already introduced by Viola and Jones [23], and which are similar to the Haar wavelets used in signal processing. The Haar-like features at a given location in a detection window are calculated by subtracting the sum of the pixel intensities of two, or more, rectangular sub-regions adjacent to that location. These features can be calculated at any position or scale in the original image. Fig. 3 shows some examples of Haar-like features, where the feature values are obtained by subtracting the pixel intensities under the white rectangles from the pixel intensities under the dark rectangles.

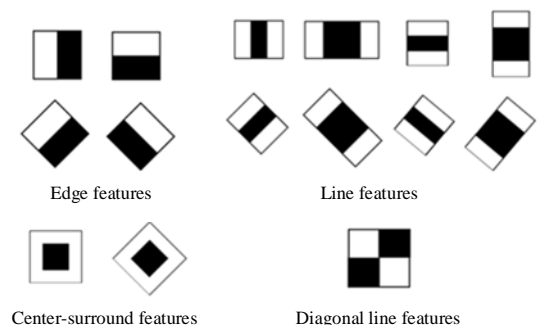


Fig. 3 Examples of Haar-like features.

The Haar-like features are calculated as explained above. However, this operation is known to be too time-consuming. So, to significantly increase the speed of detection, the authors Viola and Jones proposed the use of Integral Image concept. This integral image, denoted by “ IP ”, is an intermediate representation to compute rapidly Haar-like features for real-time use. Its size is the same as the original image “ I ”. The value of each pixel of coordinates (x, y) in the integral image is the sum of the values of the pixels above and to the left of that pixel in the original image, as shown in (1). An example of integral image computation is given in Fig. 4(a).

$$II(x, y) = \sum_{x'=1}^x \sum_{y'=1}^y I(x', y') \quad (1)$$

The sum of the pixels of a given rectangle in the image I is calculated with only just four values, and this is done by subtracting the sum of the top-left and bottom-right pixels from the sum of the top-right and bottom-left pixels. An illustration is shown in Fig. 4(b).

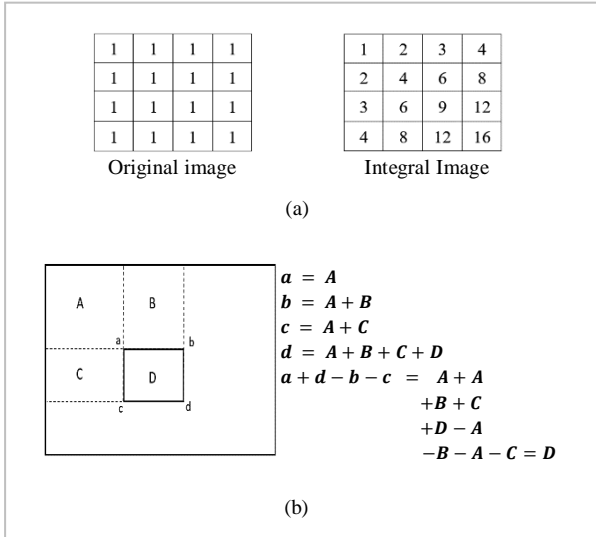


Fig. 4 Integral Image. (a) Integral image computation. (b). Using integral image to compute the sum over a given rectangle (D).

B. AdaBoost cascade classifier

For the detection of face and upper-body part, we used the AdaBoost cascade classifier, which consists of two steps. In the first one, it uses the AdaBoost algorithm to select the best and informative features, and then, in the second, it feeds these features into a cascade of classifiers to make the detection more efficient.

For the face and upper-body part detection, a large number of feature types for different scales and positions is required, which makes the detection process computationally expensive. To overcome this problem, the AdaBoost algorithm is an effective way that is proposed to drastically reduce the number of features from several thousand to only several hundred by selecting only those features that are relevant for the detection. AdaBoost algorithm evaluates each feature against the training data. The features that separate the classes more strongly are taken as relevant features.

After the selection of best features, the classification can finally be done. This is achieved by using a cascade classifier, which combines a large number of weak classifiers to form a strong classifier. Not that, a classifier is said to be weak if it performs slightly better than a random classifier.

The cascade classifier is an effective and an easy to implement method that was initially proposed by Viola and Jones in their work [23]. The use of a cascade of classifiers will give better accuracy to the detection algorithm, and this after having determined the best and informative features by

the AdaBoost algorithm for each sub-region of the image. If a sub-region of the input image is more likely to contain the object of interest (the face and upper-body part in our case), then, the detection process continues the verification with the subsequent classifiers. If all the classifiers in the cascade respond positively, thus, the algorithm concludes that the object of interest is present within the sub-region. Otherwise, the detection process does not proceed to the next classifier in the cascade, and the algorithm concludes that there is no presence of the object of interest within the sub-region of the input image. Not that, the more classifiers are included in the cascade, the better the detection accuracy of the algorithm, but at the expense of the computation time, which will be more important, so there is a trade-off between the accuracy and speed.

The scheme of the cascade classifier is shown in Fig. 5.

C. Combined face and upper-body part detectors

The human detection method proposed in this paper is based on face and upper-body part detectors by using Haar-like features as inputs into an AdaBoost cascade classifier. The major challenge encountered while detecting the upper-body part are the false positives. The method proposed in this paper is based on the fact that the head is an inseparable part of the human upper body. Therefore, in order to reduce the number of false positives, we propose to check the presence of a face inside the bounding box of the detected upper-body part.

The procedure is summarized as follows. If a face is detected inside the bounding box of an upper-body part, then, we consider that a human is detected. In that case, the bounding box of the human is estimated by using the bounding box of the detected upper-body part. In general, the proportion of the upper part of the human body in height is almost equal to a quarter of the total height of a human in upright posture. Thus, the bounding box of the human can be estimated by extending vertically downwards by 3 times the bounding box of the detected upper-body part (see Fig. 6 for an illustration).

The overall start-to-end algorithm of the procedure described above is shown in Fig. 7, where BB_{Face} , BB_{Upper_body} are the bounding boxes of the detected face and upper-body part, respectively, and BB_{Human} is the estimated bounding box of the detected human.

III. EXPERIMENTS

In this section, we present the experiments conducted in order to evaluate the performance of the proposed human detection method. First, in Section III-A, we describe the dataset used for the experiments. Then, in Section III-B, we present the obtained results and comparison with other state-of-the-art approaches.

All the experiments are conducted under the MATLAB environment on a standard PC equipped with 3.20 GHZ Intel Core i7 CPU, and 8 GB of RAM.

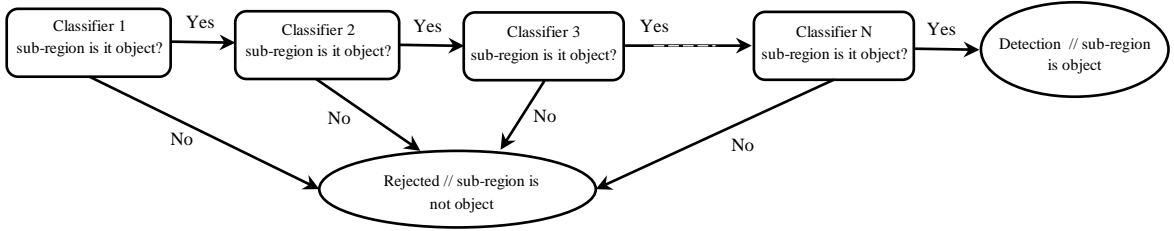


Fig. 5 Cascade classifier.

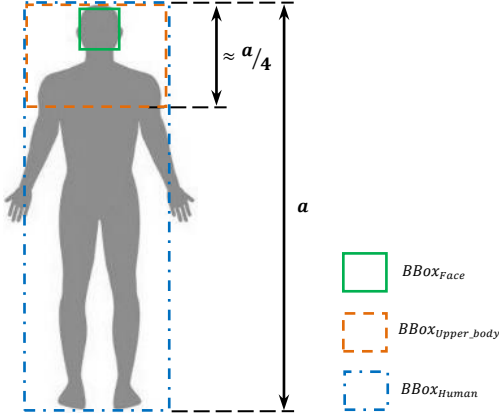


Fig. 6 Human bounding box estimation.

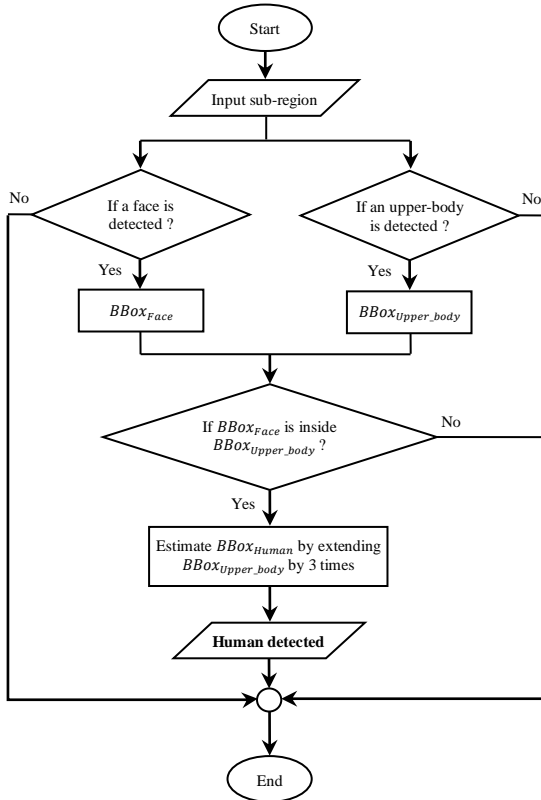


Fig. 7 The overall start-to-end algorithm of the proposed procedure for combining face and upper-body part detectors.

A. Dataset description

In order to evaluate the performance of the proposed method, we collected a dataset of 5030 from 14 people including man and woman adults. The images were acquired using a Canon G15 camera, with a resolution of 1024×768 pixels, recorded to JPEG format. The acquisition was performed in various indoor and outdoor environments (e.g., university campus areas, working rooms, office rooms, corridors, etc.) and under different challenging situations such as partially occluded human, illumination changes, complex background, variability in clothing, blurring, shadow, different skin color, facial occlusions (e.g., eyeglasses, hats, beards, masks, etc.). The dataset is divided into two sets of data. The first one contains 1248 human images as a positive class, and the second one contains 3782 images as a negative class. The samples of the positive class are cropped manually from the original images and they only contain region of interest (human), while the samples of the negative class are cropped randomly and they contain regions that belong to the background of the images. All the cropped samples are finally resized into 256×96 pixels. Fig. 8 shows some sample images from the two classes (human and non-human).

B. Experimental results

To test and validate the proposed method, we used the N -folds-cross validation approach, in which the whole created dataset is randomly divided into N (where $N = 10$) subsets. Then, the i -th subset is taken as a test set, while the rest $N-1$ subsets are taken as a training set. To demonstrate the effectiveness of the proposed method, we performed a comparison with two state-of-the-art approaches using the same dataset. These approaches are the Histogram of Oriented Gradient (HOG) [10] and Aggregated Channel Features (ACF) [24]. This comparison is conducted on the basis of the following evaluation metrics:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 - Score = \frac{2 \times (Recall \times Precision)}{(Recall + Precision)} \quad (4)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (5)$$



Fig. 8 Sample images from our created dataset. (a) Positive (Human) samples. (b) Negative (Non-human) samples.

where, TP (True Positives) is the number of humans correctly detected, TN (True Negatives) is the number of non-humans correctly detected, FP (False Positives) is the number of non-humans incorrectly detected as humans, and FN (False Negatives) is the number of humans incorrectly detected as non-humans.

The values of TP, TN, FP and FN are obtained directly from the confusion matrix (contingency matrix), which represents the predicted versus the actual class for each sample.

In addition to the aforementioned evaluation metrics, we also conducted the comparison in terms of area under the ROC (Receiver Operating Characteristic), which is a curve plotting TP rate against FP rate at different thresholds of the classifier scores.

The confusion matrices of the detection results obtained using the HOG, ACF and the proposed method are given in Tables I, II, and III, respectively, and the overall comparison in terms of precision, recall, F1-Score, and accuracy is given in Table IV.

TABLE I
Confusion matrix (in %) using HOG [10]

Actual	Predicted	
	Human	Non-Human
Human	59.31	40.69
Non-Human	00.78	99.22

TABLE II
Confusion matrix (in %) using ACF [24]

Actual	Predicted	
	Human	Non-Human
Human	79.31	20.69
Non-Human	02.34	97.66

TABLE III
Confusion matrix (in %) using the proposed method

Actual	Predicted	
	Human	Non-Human
Human	84.14	15.86
Non-Human	00.00	100

From the tables, we can observe that our proposed method is more accurate and more robust than the HOG and ACF methods, providing an overall accuracy of 91.57% and an F1-

Score of 91.38%. On the other hand, we can observe that the HOG method achieves the worst results (accuracy = 78.02%, and F1-Score = 74.13%), which demonstrate that this method is not robust enough to detect humans in the presence of challenging situations such as facial occlusions, blurring, background clutters, different skin color and variability in clothing. In Fig. 9, we also compare the performances of the different methods in terms of ROC curve. From this figure, we can observe that the proposed method produces an AUCROC of 0.92, which is superior to that of the HOG and ACF methods, which provide an AUCROC of 0.88 and 0.79, respectively. These results confirm those reported in Table IV.

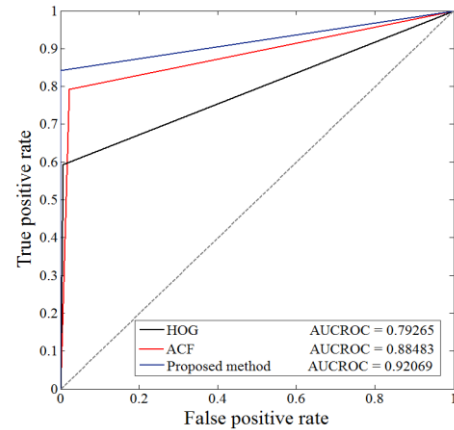


Fig. 9 Comparison in terms of ROC curve between the different methods.

We also tested the proposed method on some challenging images, and the results we obtained in comparison to the two other state-of-the-art methods are shown in Fig. 10. From Fig. 10(a), we can observe that the HOG, ACF, the face detector only, and the upper-body part detector only have all detected some false positives, but the proposed method based on combined face and upper-body part detectors is able to remove the false positives due to the background clutter. From Fig. 10(b), we can observe that the HOG and ACF method have not detected the human present in the image due to its similar clothes color to the background. However, the proposed method has successfully detected the human by combining the results of the face and upper-body part detectors.

TABLE IV
Comparison of detection performance for the proposed method
and other state-of-the-art methods

Method	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
HOG [10]	98.85	59.31	74.13	78.02
AFC [24]	97.45	79.31	87.45	87.91
Proposed method	100	84.13	91.38	91.57

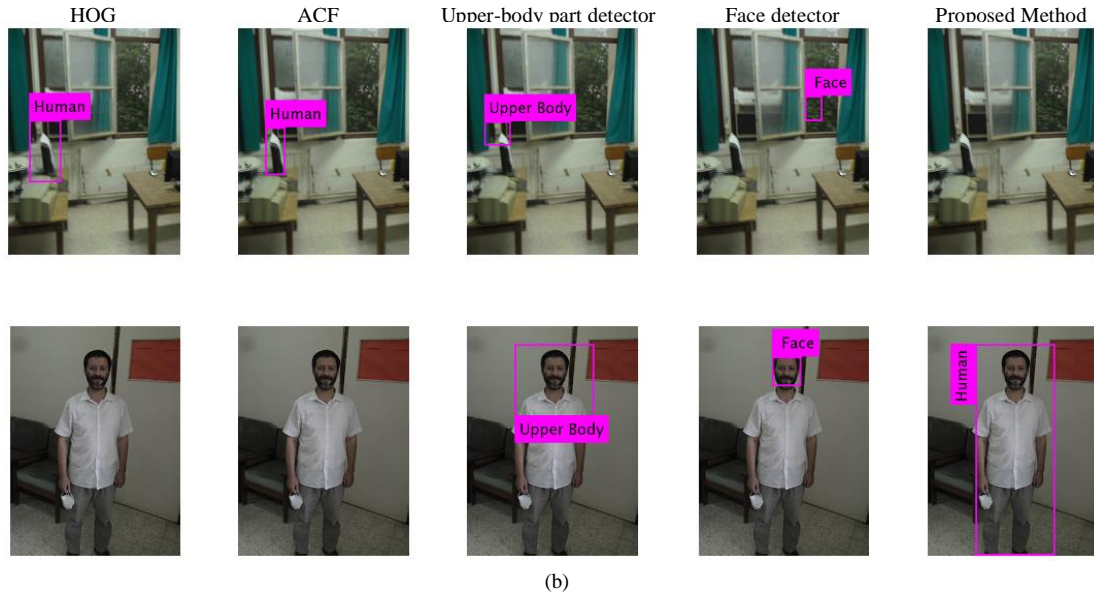


Fig. 10 Comparison of the detection results on challenging images.

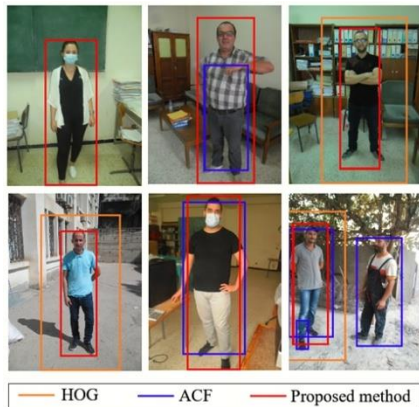


Fig. 11 Some detection results on real images using the HOG, ACF and the proposed method.

Some other detection results obtained on real images using the HOG, ACF and the proposed method are shown in Fig. 11.

IV. CONCLUSION

In this paper, a new method for human detection based on combined face and upper-body part detectors is proposed. The experimental results obtained on our created dataset have shown that the proposed method achieves the best detection performances in comparison with other state-of-the-art methods. In the future, we plan to include more features, such as LBP, in the proposed method in order to improve its

robustness against view-point changes. Furthermore, we plan to evaluate the detection performances of the proposed method on public datasets with a comparison against more recent state-of-the-art approaches.

REFERENCES

- [1] V. Gajjar, Y. Khandhediya, et A. Gurnani, « Human Detection and Tracking for Video Surveillance: A Cognitive Science Approach », in *2017 IEEE International Conference on Computer Vision Workshops (ICCVW)*, Venice, oct. 2017, p. 2805-2809.
- [2] S. Hoshino et K. Niimura, « Robot Vision System for Human Detection and Action Recognition », *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 24, n° 3, p. 346-356, mai 2020.
- [3] D. Chowdhry, R. Paranjape, et P. Laforge, « Smart home automation system for intrusion detection », in *2015 IEEE 14th Canadian Workshop on Information Theory (CWIT)*, juill. 2015, p. 75-78.
- [4] A. Mulyanto, R. I. Borman, P. Prasetyawan, W. Jatmiko, et P. Mursanto, « Real-Time Human Detection and Tracking Using Two Sequential Frames for Advanced Driver Assistance System », in *2019 3rd International Conference on Informatics and Computational Sciences (ICICoS)*, oct. 2019, p. 1-5.
- [5] T. N. Nizar, R. Hartono, et D. Meidina, « Human Detection and Avoidance Control Systems of an Autonomous Vehicle », *IOP Conference Series: Materials Science and Engineering*, vol. 879, p. 012103, août 2020.
- [6] A. Jalal, S. Kamal, et D. Kim, « A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems », *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, p. 54, janv. 2017.

- [7] S. Ding *et al.*, « SurvSurf: human retrieval on large surveillance video data », *Multimed Tools Appl*, vol. 76, n° 5, p. 6521-6549, mars 2017.
- [8] D. Sangeetha et P. Deepa, « A low-cost and high-performance architecture for robust human detection using histogram of edge oriented gradients », *Microprocessors and Microsystems*, vol. 53, p. 106-119, août 2017.
- [9] H. Beiping et Z. Wen, « Fast Human Detection Using Motion Detection and Histogram of Oriented Gradients », *JCP*, vol. 6, n° 8, p. 1597-1604, août 2011.
- [10] N. Dalal et B. Triggs, « Histograms of oriented gradients for human detection », in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, juin 2005, vol. 1, p. 886-893 vol. 1.
- [11] V.-D. Hoang, A. Vavilin, et K.-H. Jo, « Fast human detection based on parallelogram haar-like features », in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, oct. 2012, p. 4220-4225.
- [12] B. C. Ko, D.-Y. Kim, J.-H. Jung, et J.-Y. Nam, « Three-level cascade of random forests for rapid human detection », *OE*, vol. 52, n° 2, p. 027204, févr. 2013.
- [13] S. Nigam et A. Khare, « Multiresolution approach for multiple human detection using moments and local binary patterns », *Multimedia Tools and Applications*, vol. 74, n° 17, p. 7037-7062, sept. 2015.
- [14] B. U. Umar, J. Agajo, A. Aliyu, J. G. Kolo, O. S. Owolabi, et O. M. Olaniyi, « Human detection using speeded-up robust features and support vector machine from aerial images », in *2017 IEEE 3rd International Conference on Electro-Technology for National Development (NIGERCON)*, nov. 2017, p. 577-586.
- [15] O. Prakash, J. Gwak, M. Khare, A. Khare, et M. Jeon, « Human detection in complex real scenes based on combination of biorthogonal wavelet transform and Zernike moments », *Optik*, vol. 157, p. 1267-1281, mars 2018.
- [16] B. T. Bastian et J. C.V., « Integrated feature set using aggregate channel features and histogram of sparse codes for human detection », *Multimedia Tools and Applications*, vol. 79, n° 3, p. 2931-2944, janv. 2020.
- [17] B. Li, Y. Li, B. Tian, F. Zhu, G. Xiong, et K. Wang, « Part-based pedestrian detection using grammar model and ABM-HoG features », in *Proceedings of 2013 IEEE International Conference on Vehicular Electronics and Safety*, juill. 2013, p. 78-83.
- [18] C. Yao, X. Bai, W. Liu, et L. J. Latecki, « Human Detection Using Learned Part Alphabet and Pose Dictionary », in *Computer Vision – ECCV 2014*, Cham, 2014, p. 251-266.
- [19] B. C. Ko, J. E. Son, et J.-Y. Nam, « View-invariant, partially occluded human detection in still images using part bases and random forest », *OE*, vol. 54, n° 5, p. 053113, mai 2015.
- [20] Y. Tian, P. Luo, X. Wang, et X. Tang, « Deep Learning Strong Parts for Pedestrian Detection », 2015, p. 1904-1912.
- [21] W. Nam, P. Dollar, et J. H. Han, « Local Decorrelation For Improved Pedestrian Detection », in *Advances in Neural Information Processing Systems*, 2014, vol. 27, p. 424-432.
- [22] W. Ouyang, X. Zeng, et X. Wang, « Partial Occlusion Handling in Pedestrian Detection With a Deep Model », *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, n° 11, p. 2123-2137, nov. 2016.
- [23] P. Viola et M. J. Jones, « Robust Real-Time Face Detection », *International Journal of Computer Vision*, vol. 57, n° 2, p. 137-154, mai 2004.
- [24] P. Dollár, R. Appel, S. Belongie, et P. Perona, « Fast Feature Pyramids for Object Detection », *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, n° 8, p. 1532-1545, août 2014.

Recognition Process of the Right and the Left Ventricles in short-axis MRI images Based on ANFIS and KFCM Algorithm

Laidi AMEL¹, Mohammed AMMAR²

¹LIMOSE Laboratory, ²Systems and Telecommunication Laboratory

¹a.laidi@univ-boumerdes.dz

²m.ammar@univ-boumerdes.dz

University M'Hamed Bougara, Boumerdes, Algeria

Abstract— This work presents an interpretable segmentation process for cardiac MRI short-axis images. we started by detecting the right and left ventricles using an Adaptive Neuro-Fuzzy Inference System (ANFIS), following a set of interpretable rules. We then ran a classification process on the segmented images using Kernel Fuzzy C Means algorithm (KFCM) and characterization steps. For this purpose, we carefully selected a number of descriptors: circularity, elongation factor and Hu moments. The performance of our method was compared to a perceptron multilayers neural network. We achieved an overall correct rate, sensitivity, and specificity of 96%, 93.33%, and 100% respectively.

Keywords— Left and right ventricles, shape descriptors, interpretable process, ANFIS

I. INTRODUCTION

The right ventricle is one of the four chambers of the heart, it lays in the lower right portion of the heart below the right atrium. It is the chamber responsible for pumping oxygen-depleted blood to the lungs. It has a rather complex shape, compared to the left ventricle, which defies easy geometric approximations and kept it unexplored by scientific research for a long time. However, many recent studies have demonstrated the prognostic value of the RV function in cardiovascular diseases including heart failure, RV myocardial infarction, congenital heart disease and pulmonary hypertension.

Many works have been proposed in order to localize or detect the left and right ventricles in cardiac images. These localization methods can be grouped into two main categories [1]:

1. Time-based recognition: This approach exploits the movement of the heart, which is in constant motion during the cardiac cycle [2-8].

2. Shape-based recognition [9-11].

Several segmentation methods have been published to detect LV in MRI or Echocardiographic images. An example of these approaches published in [12] was based on the conserved myocardial volume concept .

In [13] the authors proposed a strategy that combines the border detection process and the generalized Hough transform to initialize an active-contour algorithm. Meanwhile, approaches based on CHT or Generalized Hough Transform use an accumulator space, which makes them cost more memory space and computational time.

In a previous work, we have proposed an automatic level set based method [14] for endocardial border segmentation from MRI images. The process was automatically initialized with the circularity index.

Similarly, many works interested in RV segmentation have been published. In [15] Two semi-automated level-set-based approaches were used for right ventricular (RV) endocardial border detection from MR images.

In [16], the authors have proposed a statistical shape analysis model to detect and locate regional wall motion abnormalities (RWMA). They have used the independent component analysis (ICA) to train the proposed model from normal and abnormal myocardial contraction.

In [17], a new active-shape-model (ASM) based method was proposed to segment the right ventricle (RV) from cardiac MRI images. The proposed method achieved a significantly low segmentation error compared to existing literature.

Recently, several methods based on Deep Convolutional Neural Networks (CNN) were developed in the literature for detecting endocardial and epicardial borders in LV [18,19]. These approaches use a stack of filters that can automatically learn hierarchies of features with no need of input pre-processing. The algorithm presented in [20] starts with the application of U-Net architecture for the detection of LV. The LV segmentation results are then used to identify some regional cardiac indices. In other works, CNN has been combined with other algorithms, such as deformable models, to achieve better segmentation and classification performances [21].

The goal of this paper is to develop an interpretable segmentation process for right and left ventricles in MRI cardiac short-axis images based on the KFCM segmentation

method and the ANFIS classifier. We propose to characterize the left and right ventricles by using some selected parameters extracted from the shape descriptor. Indeed, the left ventricle usually has a circular shape, and the right ventricle is more elongated. For this reason, we have chosen the following attributes as input of the ANFIS classifier: elongation, circularity factor.

II. MATERIALS AND METHODS

A. Fuzzy-C Means (FCM)

the Fuzzy C Means (FCM) developed by Bezdek ([22],[23]) attempts to minimize the objective function with respect to each fuzzy membership degree u_{ik} and each prototype v_i .

$$J = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m \|x_k - v_i\|^2, 1 \leq m \leq \infty = \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m d_{ik}^2 \quad (1)$$

Similarly, the membership degrees or prototypes are first initialized, then each iteration of the process, affiliations and prototypes are updated, and the data vectors move from one class to another to minimize the objective function until stabilization.

Membership degrees must satisfy the following conditions:

$$0 \leq u_{ik} \leq 1 \quad \forall i \in [1, c], \forall k \in [1, n]$$

$$\sum_{i=1}^c u_{ik} = 1 \quad \forall k \in [1, n]$$

$$0 < \sum_{k=1}^n u_{ik} < n \quad \forall i \in [1, c]$$

Membership's degrees u_{ik} and prototype v_i are given as following:

(2)

$$\mu_{ik}^m = \frac{(d(x_k, v_i))^{2/1-m}}{\sum_{j=1}^c (d(x_k, v_j))^{2/1-m}} \quad (2)$$

$$v_i = \frac{\sum_{k=1}^n u_{ik}^m x_k}{\sum_{k=1}^n u_{ik}^m} \quad (3)$$

The FCM algorithm can be described as follows :

Initializing parameters:

c : number of classes.

ε : Error of convergence .

m : Fuzzy degree.

2. Initializing the membership's degrees U between $[0, 1]$.

3. Upload prototype V with relation (2.6).

4. Upload Membership's degrees with the relation (2.5).

5. Repeat steps 3 to 4 until:

$$\|U_{old} - U_{new}\| < \varepsilon.$$

Kernel Fuzzy c-means KFCM

In case of the classification algorithm based on Kernel Fuzzy c-means (KFCM) the objective function is minimized iteratively by the equation defined by ([24],[25],[26]) :

$$J_{m=\sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m \|\Phi(I_k) - \Phi(CC_i)\|^2 \quad (4)$$

Φ : is an implicit nonlinear function. In this application, the selected function is the Gaussian function, $K(x, x) = 1$, so after simplification we obtain the following formula:

$$J_{m=2 \sum_{i=1}^c \sum_{k=1}^n \mu_{ik}^m (1 - K(I_k, CC_i)) \quad (5)$$

Center or prototype and membership degree are given by:

$$\mu_{ik}^m = \frac{(1 - K(I_k, CC_i))^{-1/m-1}}{\sum_{j=1}^c (1 - K(I_k, CC_j))^{-1/m-1}} \quad (6)$$

$$CC_i = \frac{\sum_{k=1}^n \mu_{ik}^m K(I_k, CC_i) I_k}{\sum_{k=1}^n \mu_{ik}^m K(I_k, CC_i)} \quad (7)$$

We resume the KFCM algorithm as follows :

Inputs: number of classes c , maximum number iteration t_{max} , and $m > 1$, $\varepsilon > 0$, δ is the parameter of the Gaussian function.

Output: CC_i and μ_{ik} , $i=1, 2, \dots, c, k=1, 2, \dots, N$.

Step 1: Initialization of centers CC_0 , membership degree

μ_{ik} .

Step 2:

for $t=1, 2, \dots, t_{max}$:

(a) : upload CC_0 and μ_{ik} .

(b) Calculate :

$$E^t = \max_{i,k} |\mu_{ik}^t - \mu_{ik}^{t-1}|. \text{ If } E^t \leq \varepsilon \quad (8)$$

Finish,

else repeat (a).

Step 3: output ; CC_i and μ_{ik}

B. Characterization of the right and left ventricle

The characterization step is very important. It allows the detection and localization of cardiac structures as left and right ventricles. In the field of pattern recognition, many

descriptors are available. The choice depends on the descriptors of the object characterized.

The left ventricle has generally a circular shape and the right ventricle is longer, for this reason, we chose the following attributes [33]:

Circularity factor:

$$F_{cH} = P(R)/(2\sqrt{\pi \cdot A(R)}) \quad (9)$$

Where, $P(R)$: is the perimeter.

$A(R)$: Area

Elongation factor = RME length / width of RME

RME : rectangle minimum supervision

Type Factor :

$$F_t = \frac{A^2}{4\pi\sqrt{I_{xx} + I_{yy}}} \quad (10)$$

Hu moments computation

In [27], Hu proposed six moments which are invariant to position, scale, orientation, and parallel projections [28-29].

The calculation of the Hu moments is based on spatial and central moments. The spatial moments m_{pq} of an image I_{xy} are calculated with the following equation (11):

$$m_{pq} = \sum_{x=1}^M \sum_{y=1}^N x^p y^q I_{xy} \quad (11)$$

This allows us to compute the center of mass of the image and of a region in case of a binary mask. Centralized moment \bar{m}_{pq} are geometric moments of the image computed relatively to the center of mass (\bar{x}, \bar{y}) :

$$(\bar{x}, \bar{y}) = \left(\frac{m_{10}}{m_{00}}, \frac{m_{01}}{m_{00}} \right) \quad (12)$$

$$\bar{m}_{pq} = \sum_{x=1}^M \sum_{y=1}^N (x - \bar{x})^p (y - \bar{y})^q I_{xy} \quad (13)$$

Centralized moments are invariant under translation. To enable invariance to scale, normalized moments λ_{pq} are used:

$$\Lambda_{pq} = \frac{\bar{m}_{pq}}{\bar{m}_{00}^\gamma}, \quad (14)$$

$$\text{with } \gamma = \frac{p+q}{2} + 1, \forall p + q \geq 2 \quad (15)$$

Hu proposed a set of orthogonal moment invariants, which can be used for scale, position, and rotation invariant pattern identification. We define the three orders Hu moments with the following formulas:

$$I_1 = \lambda_{20} + \lambda_{02} \quad (16)$$

$$I_2 = (\lambda_{20} - \lambda_{02})^2 + \lambda_{11}^2 \quad (17)$$

$$I_3 = (\lambda_{30} - 3\lambda_{12})^2 + (3\lambda_{21} - \lambda_{03})^2 \quad (18)$$

C. Adaptive Neuro-Fuzzy Inference System (ANFIS)

ANFIS is an adaptive network that uses a neural network combined with fuzzy logic. In fact, ANFIS has a fuzzy inference system optimized with the feed-forward back-propagation algorithm.

The Consequent parameters of the rule are calculated forward, and the premise parameters are calculated backward. Many fuzzy inference architectures have been proposed by different researchers [30].

In our work, we use zero-order Takagi- Sugeno fuzzy inference system, where the premise part of the fuzzy rule is a fuzzy proposition, and the conclusion part is a constant.

III. RESULTS AND DISCUSSIONS

A. Image acquisition

For our experiments, we used two free databases available for research purposes.

The Sunnybrook Cardiac Data set (SCD) contains 45 cine-MRI images:

12 ischemic heart failures, 12 non-ischemic heart failures, 12 LV hypertrophies, and 9 normal cases. The whole complete data set is now available on the CAP server with a public domain license [31].

The second free database is available at [32].

B. Methodology:

- 1) *Segmentation and localisation of the left and the right ventricles* : the KFCM algorithm is applied to segment the short axis view. We present in Fig. 1 a segmentation example by using the KFCM algorithm with 3 clusters.

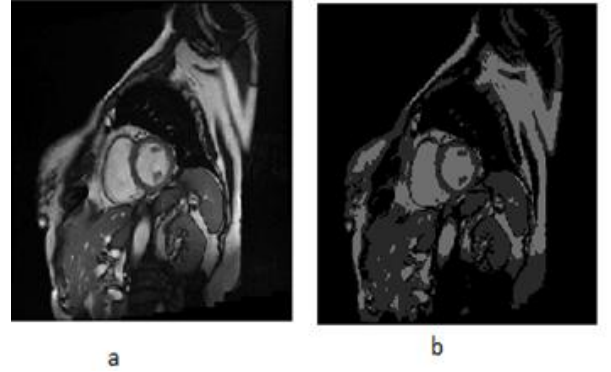


Fig. 1 a. original image. b. KFCM segmentation

According to Fig. 1.b, the algorithm KFCM segments the image into three classes. The first represents the low intensity pixels containing the lung and the background of the image, the second gathers the pixels of average intensities like the myocardium, while the third gathers the pixels of high intensities like the left and right ventricle.

The segmentation results are evaluated on the data set [33] where a manual segmentation performed by two experts is

available. We obtain an average similarity of 94.82% (expert1) and 95.52% (expert2).

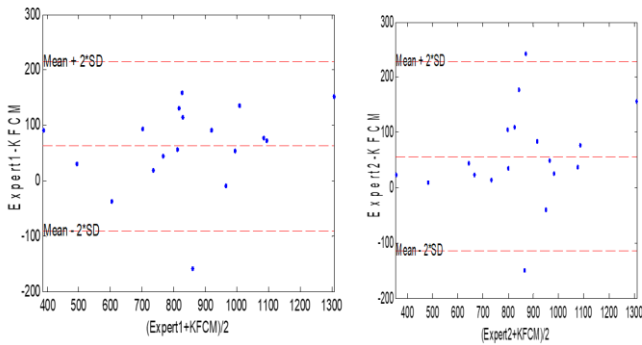


Fig.2 Bland-Altman representation for expert 1 and 2

To obtain a clearer interpretation of the results, we have drawn the Bland-Altman graph between expert1 and the KFCM method (Fig. 2.a), and between expert2 and the KFCM method. (Fig. 2.b). The Bland-Altman plot compares the means of the measurements to their differences. It is constructed as follows:

- The first step is to determine the difference between the tested method and the reference method. The two limits of agreement are calculated by the average of this difference \pm two standard deviations.
- The second step is to make a graph where we write the differences between the two measurements according to the average value of the two measurements.

2) *Characterization of the right and left ventricle:* We have used ANOVA for feature selection. We note that having an interference between the “values” representing a studied feature means both the RV and the LV can be identified with it, which makes it impossible to use said feature to identify one ventricle from the other. Similarly, having a “specific value” of a feature for one structure without the other, gives a possibility to associate said feature with that ventricle.

A graphical presentation based on boxplot is given in Fig. 3 and Fig. 4 to show the importance and uniqueness of the selected parameters.

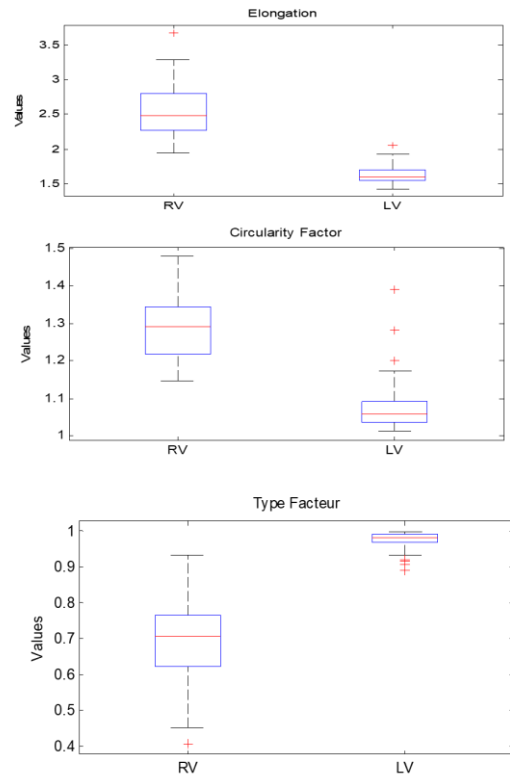


Fig. 3 Elongation, circularity and type factor box plots presentations for RV and LV.

We note that the right ventricle has the highest elongation factor. For the 45 images selected, the elongation factor is greater than 2 for the right ventricle, and it is less than 2 for the left ventricle.

In addition, the left ventricle has a circularity factor generally between 1 and 1.2. The right ventricle has a circularity factor greater than 1.3. We note that the left ventricle has a type factor close to 1. For images the type factor is between 0.95 and 1. For the right ventricle this factor is less than 0.9.

We can conclude from the box plot representations of the different descriptors that the most relevant characterization parameter for the right ventricle is the elongation factor, where for all 45 selected images, the right ventricle is characterized by an elongation factor greater than 2.

While the type factor and circularity are the main parameters characterizing the left ventricle, when analysing all three descriptors it is safe to say that the factor type is the best parameter characterizing the circularity of the left ventricle.

3) *Recognition of the right and left ventricle using ANFIS:* After the characterization step, we select the elongation and the circularity factor as input in the ANFIS classifier. For each parameter we choose two functions: less and high.

The proposed model has 2 inputs; each input has two trapezoidal membership functions.

According to the characterization of the right ventricle, the elongation factor is generally greater than 2, and it is less than two for the left ventricle.

The left ventricle has a type factor close to 1 (between 0.95 and 1). while, for the right ventricle, it is less than 0.9.

The neuro-fuzzy classifier generates automatically a set composed of 4 fuzzy rules to justify the classification:

1. If (Elongation is less) and (Circularity is less) then (output is unknown)
2. If (Elongation is less) and (Circularity is high) then (output is LV)
3. If (Elongation is high) and (Circularity is less) then (output is RV)
4. If (Elongation is high) and (Circularity is high) then (output is unknown)

We achieved an overall correct rate, sensitivity, and specificity of 98%, 96.67%, and 100% respectively for the perceptron multilayers, and a correct rate, sensitivity, and specificity of 96%, 93.33%, and 100% respectively for ANFIS. A final segmentation can be shown on an actual short axis slice of an MRI image in Fig. 5. As it is visually clear, the left ventricle has a circle shape while the right ventricle has more of an elongated shape.

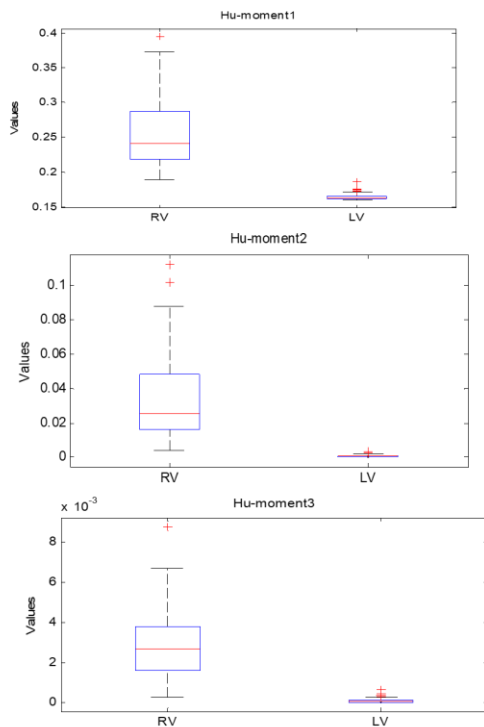


Fig. 4 Hu moment1, Hu moment2 and Hu moment3 box plots presentations for RV and LV.

We can conclude that the two classifiers provide the best performances. A little decrease in the accuracy of the ANFIS

model is noted. On the other hand, the results obtained by the ANFIS classifier are explicit and interpretable.

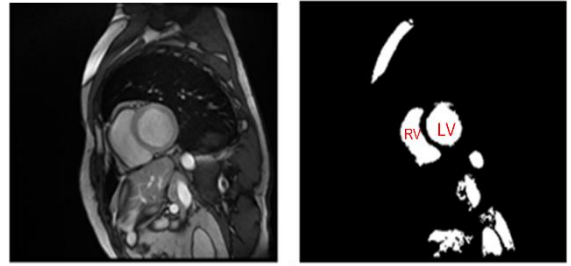


Fig. 5A demonstration of RV and LV segmentation on a short axis MRI image.

IV. CONCLUSIONS

This paper presents a new approach to MRI cardiac structures segmentation and localization using KFCM and ANFIS.

The classification process was to previously segmented and characterized images using KFCM. The descriptors selected as input for the ANFIS classifier were circularity and elongation factors.

The end results of 96% correct rate, 93.33% sensitivity, and 100% specificity are quite promising and comparable to novel neural network models, while keeping an explicit explainable aspect that is sacrificed when working with neural networks (which are referred to as black boxes).

In future work, we aim to apply this model to detect pathological shapes of the left and the right ventricles.

REFERENCES

- [1] Petitjean C, Dacher JN. A review of segmentation methods in short axis cardiac MR images. *Med Image Anal.* 2011 Apr;15(2):169-84. doi:10.1016/j.media.2010.12.004. Epub 2010 Dec 24.
- [2] .Pednekar, A., Kurkure, U., Muthupillai, R., Flamm, S., 2006. Automated left ventricular segmentation in cardiac MRI. *IEEE Trans Biomed Eng* 53 (7), 1425-1428.
- [3] Huang, J., Huang, X., Metaxas, D., Axel, L., 2007. Dynamic Texture based Heart Localization and Segmentation in 4-D Cardiac Images. In: *Proc. Of the IEEE Intl Symposium on Biomedical Imaging: From Nano to Macro (ISBI)*. pp. 852–855.
- [4] Cocosco, C., Netsch, T., Senegas, J., Bystrov, D., Niessen, W., Viergever, M., 2004. Automatic cardiac region-of-interest computation in cine 3D structural MRI. In: *Proc. of the Conference on Computer Assisted Radiology and Surgery (CARS)*. Chicago, pp. 1126–1131.
- [5] Gering, D., 2003. Automatic segmentation of cardiac MRI. In: *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. No. 1 in LNCS. pp. 524-532.
- [6] Otsu, N., 1979. A threshold selection method for gray-level histograms. *IEEE Trans. Syst. Man Cybernet* 1, 62–66.
- [7] Lin, X., Cowan, R., Young, A., 2006. Automated detection of left ventricle in 4D MR images: experience from a large study. In: *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. No. 9 in LNCS. pp. 728-35.
- [8] Jolly, M.-P., 2008. Automatic recovery of the left ventricular blood pool in cardiac cine MR images. In: *Proc. of Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. LNCS 5241. pp. 110–118.
- [9] C .Constantinides, Y Chenoune , E Mousseaux , E Roullot, F Frouin. Automated Heart Localization for the Segmentation of the Ventricular Cavities on Cine Magnetic Resonance Images *Computing in Cardiology* 2010;37:911-914.

- [10] Pavani, S.-K., Delgado, D., Frangi, A., 2010. Haar-like features with optimally weighted rectangles for rapid object detection. *Pattern Recognition* 43 (1), 160-172.
- [11] Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. pp. 511-518.
- [12] Christopher.D, Garson, Bing Li, Scott T. Acton, John A. Hossack, Guiding automated left ventricular chamber segmentation in cardiac imaging using the concept of conserved myocardial volume, *Computerized Medical Imaging and Graphics*, 32, 321-330(2008).
- [13] Fernandez-Caballero.A José.M. Vega-Riesco, Determining heart parameters through left ventricular automatic segmentation for heart disease diagnosis, *Expert Systems with Applications*, 36, 2234-2249(2009).
- [14] M. Ammar & S. Mahmoudi & M.A. Chikh & A. Abbou. Endocardial Border Detection in Cardiac Magnetic Resonance Images Using Level Set Method. *Journal of Digital Imaging (Springer)*. DOI 10.1007/s10278-011-9404-z. (2012) 25:294-306.
- [15] Maria C Carminati, Paola Gripari, Francesco Maffessanti, Cristiana Corsi, Gianluca Pontone, Daniele Andreini, Mauro Pepi, Enrico G Caiani. Semi-Automated Border Detection for Right Ventricular Volume Estimation from MR Images. *Computing in Cardiology* 2011;38:381-384.
- [16] Suinesiaputra A, Frangi AF, Kaandorp TAM, et al. Automated detection of regional wall motion abnormalities based on a statistical model applied to multislice short-axis cardiac MR images. *IEEE Trans Med Imaging* 2009;28:595-607.
- [17] ElBaz MS, Fahmy AS. Active shape model with inter-profile modeling paradigm for cardiac right ventricle segmentation. *Med Image Comput Assist Interv*. 2012;15(Pt 1):691-8.
- [18] Yang X, Song Q & Su Y: Automatic segmentation of left ventricle cavity from short-axis cardiac magnetic resonance images. *Med Biol Eng Comput* 2017; 55:1563-1577.
- [19] Tan LK, Liew YM, Lim E & McLaughlin RA: Convolutional neural network regression for short-axis left ventricle segmentation in cardiac cine MR sequences. *Med Image Anal* 2017;39 :78-86.
- [20] Dangi S, Yaniv Z & Linte CA: Left Ventricle Segmentation and Quantification from Cardiac Cine MR Images via Multi-task Learning. *Stat Atlases Comput Models Heart* 2019;11395:21-31.
- [21] Attia A & Dayan S: Detection and segmentation of the Left Ventricle in Cardiac MRI using Deep Learning. 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.
- [22] J.C. Bezdek. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York, 1981
- [23] K.L. Wu, M.S. Yang, Alternative c-means clustering algorithms, *Pattern Recognition* vol. 35, pp. 2267-2278, 2002
- [24] L. Zhang, W.D. Zhou, L.C. Jiao. Kernel clustering algorithm. *Chinese J. Computers*, vol. 25(6), pp. 587-590, 2002 (in Chinese)
- [25] D.Q Zhang, S.C. Chen. Fuzzy clustering using kernel methods. in *Proceedings of Inter. Conf. Control and Automation (ICCA'02)*, pp. 123-128, Xiamen, China, June 16-19, 2002
- [26] K.R. Muller, S. Mika, et al. An Introduction to Kernel-based Learning algorithms. *IEEE Trans. Neural Networks*, vol. 12(2), pp. 181-202, 2001
- [27] M.-K. Hu, "Visual pattern recognition by moment invariants" *IEEE trans. on Information Theory*, vol. 8, pp. 179-187, 1962.
- [27] Huang, Z. , & Leng, J. (2010). Analysis of Hu Moment Invariants on Image Scaling and Rotation. *Proceedings of 2010 2nd International Conference on Computer Engineering and Technology (ICCET)*. (pp. 476-480). Chengdu, China. IEEE.
- [28] Conseil, S. and Bourennane, S. and Martin, L. Comparison of Fourier Descriptors and Hu Moments for Hand Posture Recognition. *European Signal Processing Conference (EUSIPCO)*. 2007
- [29] Radau P, Lu Y, Connelly K, Paul G, Dick AJ, Wright GA. "Evaluation Framework for Algorithms Segmenting Short Axis Cardiac MRI." *The MIDAS Journal - Cardiac MR Left Ventricle Segmentation Challenge*, <http://hdl.handle.net/10380/3070>
- [30] Tsukamoto, Y., An approach to fuzzy reasoning method. In: Gupta, M. M., Ragade, R. K., and Yager, R. R., (Eds.), *Advances in Fuzzy Set Theory and Applications*, 137:149. North-Holland, Amsterdam, 1979.
- [31] L. Najman, J. Cousty, M. Couprie, H. Talbot, S. Clément-Guinaudeau, T. Goissen et J. Garot. An open, clinically-validated database of 3D+t cine-MR images of the left ventricle with associated manual and
- [32] Ammar M., Mahmoudi S., Chikh M.A., Abbou A. (2014) Toward an Automatic Left Atrium Localization Based on Shape Descriptors and Prior Knowledge. In: Camara O., Mansi T., Pop M., Rhode K., Sermesant M., Young A. (eds) *Statistical Atlases and Computational Models of the Heart. Imaging and Modelling Challenges*. STACOM 2013. *Lecture Notes in Computer Science*, vol 8330. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-54268-8_5

K-NN Classification Using Two Dimensional Random Coefficients Autoregressive Model(2D-RCA)

Ayoub Abderrazak Maarouf ¹, Fella Hachouf ¹, Soumia Kharfouchi ²

¹ laboratoire d'Automatique et de Robotique

Département d'Electronique, Université des frères Mentouri
Constantine 1 Algerie

Email: ayoub.maarouf@umc.edu.dz

² Département de médecine, Bon Pasteur Chalet des Pins,
Université Constantine 3 Algerie

Email: S_kharfouchi@yahoo.fr

Abstract—In this work, a novel approach for automatic texture classification using K Nearest Neighbor (K-NN) classifier is presented. This approach is divided into two steps. First two-dimensional random autoregressive model 2D-RCA is applied in texture images to extract the relevant parameters to represent this texture. Then extracted parameters are estimated by the Generalized Method of Moment (GMM). These parameters are used, as pixel features in texture image classification using the K Nearest Neighbor (K-NN). A comparative study has been performed with other classification methods. The proposed method has exhibited great execution regarding obtained results by these methods.

Index Terms—Texture Classification, K Nearest Neighbor , Classification, 2D-RCA models, GMM.

I. INTRODUCTION

Texture is one of the most important information for image interpretation (1). Texture features reveal essential details about the primitives that characterize a texture, along with their relationship (2). In particular, texture classification has become an important topic in image processing and pattern recognition. Texture classification relies on identifying image regions using their textural properties. However, and over the years, various methods of classification have been developed among them: namely: structural(3), statistical (4) and signal processing methods and based stochastic model (5). Additionally, morphology-based methods (5). Based on prior research, it has been discovered that the majority of classification texture classification approaches have integrated texture characteristics with classifiers to improve classification accuracy for various images (6).

Texture characteristics offer crucial information about the primitives that make up a texture and their relationship (2). Several techniques to representing texture have been presented. Co-occurrence matrices(7), Weber Local Descriptor (WLD)(8), Local Binary Pattern (LBP) (9), autocorrelation-based and registration-based features (10) are examples of statistical features. Primitive measurements (3), edge measurements (11), and mor-phological operation characteristics

are examples of structural features. Spatial domain filtering (2), frequency domain analysis (12), and common spatial-frequency techniques (12) are examples of filter characteristics.

in this work, a proposed method based on spatial statistics and the K-nearest neighbor is developed. First, an image is modeled by using two random coefficients auto-regressive models (13). Then a Generalized Method of Moment (GMM) (14) is used to estimate the coefficient of this model. The 2D-RCA estimated parameters will be used as an input to the K Nearest Neighbor (K-NN) for classification. The framework of the proposed method is given in figure1.

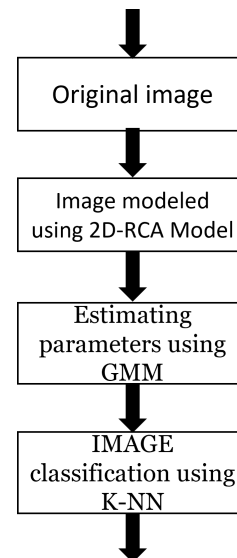


Fig. 1: block diagram of the proposed method.

The paper is arranged as follows: Section 2 introduces and describes the spatial RCA models as well as the GMM application to the 2D-RCA model. Section 3 presents the proposed algorithm. In section 4, some preliminary results are

discussed. Conclusions are given in the last section.

II. SPATIAL RCA MODEL AND PARAMETER ESTIMATIONS

The first step of this algorithm consists of modeling the texture with the 2D-RCA models introduced by (13)

The 2D-RCA model is an extension of a 2D-AR model introduced in(15) to model some non-Gaussian spatial data set such as textured images. Its construction was inspired by the famous 1D-RCA model widely used in econometric modeling (16; 17) and engineering applications (18) A 2D stochastic process follows a 2D-RCA model if:

$$x(t) = \sum_{s \in]0; P]} a_s(t)x(t-s) + e(t), t \in Z^2 \quad (1)$$

Where $a_s(t) = \alpha_s + \beta_s(t)$. For this model, we need the following assumptions:

- $e(t)$; $t \in Z^2$ is an independent second order stationary sequence of random variables with mean zero and variance σ^2
- α_s ; $s \in]0; p]$ are real constants.
- If $\beta(t) = (\beta_s(t); s \in S]0; P])'$, then $(\beta(t), t \in Z^2)$ is an independent sequence of $d \times 1$ random vectors with mean zero and $E, (\beta(t)\beta_0(t)) = C$.
- $\beta(t)$ and $e(t)$ are independent.

The 2D-RCA models proposed in (15) are drawn on a regular network. They are unilateral by construction, only most of the images we process are made of grids with irregular pixels. Fortunately, with the increasing use of computers begins center technology, in at least some situations, data with irregularly spaced pixels can be replaced by a regular grid using image interpolation techniques and re-sampling programs. In the estimation part, we need the second and fourth-order stationarity of the model. the conditions of the second and fourth orders have been established in (13). As in practice most of the spatial models are of first-order we will focus on the first order 2D-RCA model given by:

$$\begin{aligned} X(i, j) = & \alpha X(i, j-1) + \beta X(i-1, j) + \gamma X(i-1, j-1) + \\ & a_1(i, j)X(i, j-1) + a_2(i, j)X(i-1, j) + \\ & a_3(i, j)X(i-1, j-1) + \varepsilon(i, j) \end{aligned} \quad (2)$$

$$\underline{x} = (X(i, j-1); X(i-1, j); X(i-1, j-1)). \quad (3)$$

where $X(i,j)$: gray level at (i,j) .

The condition of the second order stationarity is given by:

- $\Delta = \alpha^2 + \beta^2 + \gamma^2$.
- $\Delta > 0$.
- $1/2 (\alpha^2 + \beta^2 + \sqrt{\Delta}) < 1$.

Under stationary conditions, the estimation of the 2D-RCA model given by Equation (5) is achieved by the generalized method of moments (GMM) (14). Based on the observations :

$$\underline{x} = (X(i, j-1); X(i-1, j); X(i-1, j-1)). \quad (4)$$

The GMM estimator of (α, β, γ) is given by:

$$\hat{\theta}_n = \sum_{i=1}^n \sum_{j=1}^m [\underline{x}(i, j)\underline{x}'(i, j)]^{-1} \times \sum_{i=1}^n \sum_{j=1}^m [\underline{x}(i, j)X(i, j)] \quad (5)$$

where: $\theta=(\alpha, \beta, \gamma)$

For more details see(15) .

III. PROPOSED METHOD

in this section, we present an algorithm for texture classification using 2D-RCA modeling. We have used the k-nearest neighbor approach as a classifier in this work.

K-Nearest Neighbors (K-nn) is one of the simplest algorithms used in Machine Learning for categorizing objects based on the feature space's nearest training samples(19). The k-nearest neighbor algorithm computes the distance between the query sample and each training sample and chooses the best neighbors with the smallest distance. In comparison to the other technique, the nearest neighbor is simpler to implement. It also does not need much training. It is especially useful when dealing with a small data set that cannot be effectively trained using other machine learning approaches.

The main disadvantage of the nearest neighbor method is that the calculating distance speed increases with the amount of available training samples. K is the number of nearest neighbors considered while making the prediction. The nearest pixel is determined by its proximity to the pixel under consideration.

For example, if K=6, we consider the six nearest points and use the label of the majority of these six points as the projected label. Figure 2 shows an example of how the neighbors are estimated using KNN.

The aim is to predict the label for the point marked as X. If K=3, out of 3 neighboring points of X, 2 are blue circle and 1 is black. So we predict the label for X as a blue circle. If K=6, out of 6 neighboring points of X, 4 are black circle and 2 are blue. So we predict the label for X as black circle. From this example, we can see that as K varies, the predicted label differs. Thus K is the parameter for K-nn that is to be tuned to find the optimal value. On the labeled training data, different values of K are tested, the K value that given the best result is choosing. the K value is fixed, this value can later be used for predicting unlabeled data points.

There are two sections of the proposed algorithm.

First, the image is modulated using the 2D-RCA model, and then the parameters are estimated using the GMM process, and these parameters are used as characteristic vectors for classification using the K-NN.

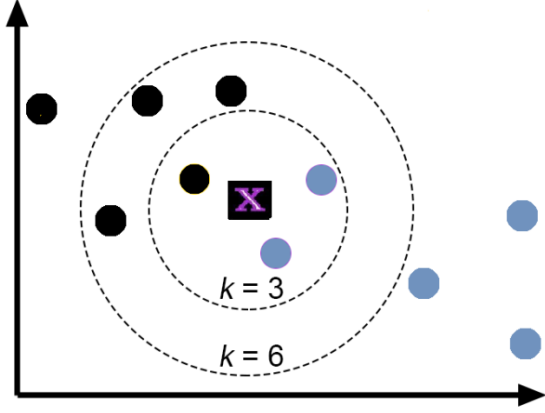


Fig. 2: k-nn classifier.

the proposed algorithm consists of the following steps:

Algorithm 1: knn classification using 2D_RCA models

Input: IMG image with N lines and M columns,
Number of class c, parameter k , windows size

Output: classified image

Step 1: Extract texture features using 2D-RCA model Do:

1- Identify the k relevant thematic classes on the IMG image // it is a supervised classification, hence the user selects representative samples for each class in the image.

1 2- For each thematic class C_i ; $i = 1; \dots; k$ select an image window of size $n \times n$, then using (4) and (5) estimate the parameters α, β, γ and σ .

2 The obtained 4 values allow constructing a 1×4 characteristic vector V_i of each thematic class C_i ; $i = 1; \dots; k$:

Step 2: classification step

1- For each pixel pix of IMG, do

a) Extract the image window W_{pix} of size $n \times n$ centred on pix .

b) Construct the characteristic vectors following previous steps

2- apply the knn algorithm using the estimated parameters of the 2D-RCA model as a vector training

IV. EXPERIMENT RESULTS AND DISCUSSION

In this section, the results of the proposed methods are presented. the algorithm has been tested on the well known brodatz database (20). The proposed method has been applied to different images from the USC-SIPI image database presented in figure(3).

The USC-SIPI image database is a collection of digitized images. It is largely maintained to support image processing,

TABLE I: Texture classification accuracy

Classification accuracy			
images	KNN	SVM	distance algorithm
image1	91.005	89.89	90.01
image2	90.11	89.84	88.87
image3	90.02	90.11	90
image4	90.1	90	89.8
image5	90	87.51	87.59
image 6	94.07	90.02	90

image analysis, and machine vision research. The USC-SIPI picture database was initially published in 1977, and since then, numerous more images have been added. the USC-SIPI image database texture includes 155 images, all grayscale, 130 512x512 and 25 1024x1024.

In the proposed K_NN classification model, the best and optimum k -value is 2, obtained results are given by fig (3), the different test has been carried out on size window of 25×25 on two and four classes with a different region.

It's noticed that most of the fine textures have been detected by the proposed approach (fig.3a).

To assess the performance of the proposed approach, obtained results have been compared to those given by support vector machine (SVM) and distance-based algorithm, SVM algorithm has used the 2D RCA parameters while the distance-based algorithm has used the first four parameters of Haralick, namely: contrast, homogeneity, correlation, and entropy extracted from the co-occurrence matrix.

Classification results illustrated by Fig.(3) show that the results obtained by the k -nn classification using 2D-RCA parameters are the best for different textures compared to other algorithms. the 2D-RCA model presents a generality in relation to textures; hence a better classification rate under varying K has been studied using the K -nearest neighbor classifier, they have been computed for different textures and methods. The results are presented in Table(1).

On most testing images, the best classification rate is given by the proposed approach. For different images, it is observed that the images classification is visually acceptable. The carried tests (table 1) have shown that the best result obtained by the proposed algorithm.

Classification results illustrated by Fig.(3) show that the results obtained by the k -nn classification using 2D-RCA parameters are the best for different textures compared to other algorithms. the 2D-RCA model presents a generality in relation to textures; hence a better classification rate under varying K has been studied using the K -nearest neighbor classifier, they have been computed for different textures and methods. The results are presented in Table(1).

On most testing images, the best classification rate is given by the proposed approach.

For different images, it is observed that the image's classification is visually acceptable. The carried tests (table 1) have shown that the best result obtained by our algorithm.

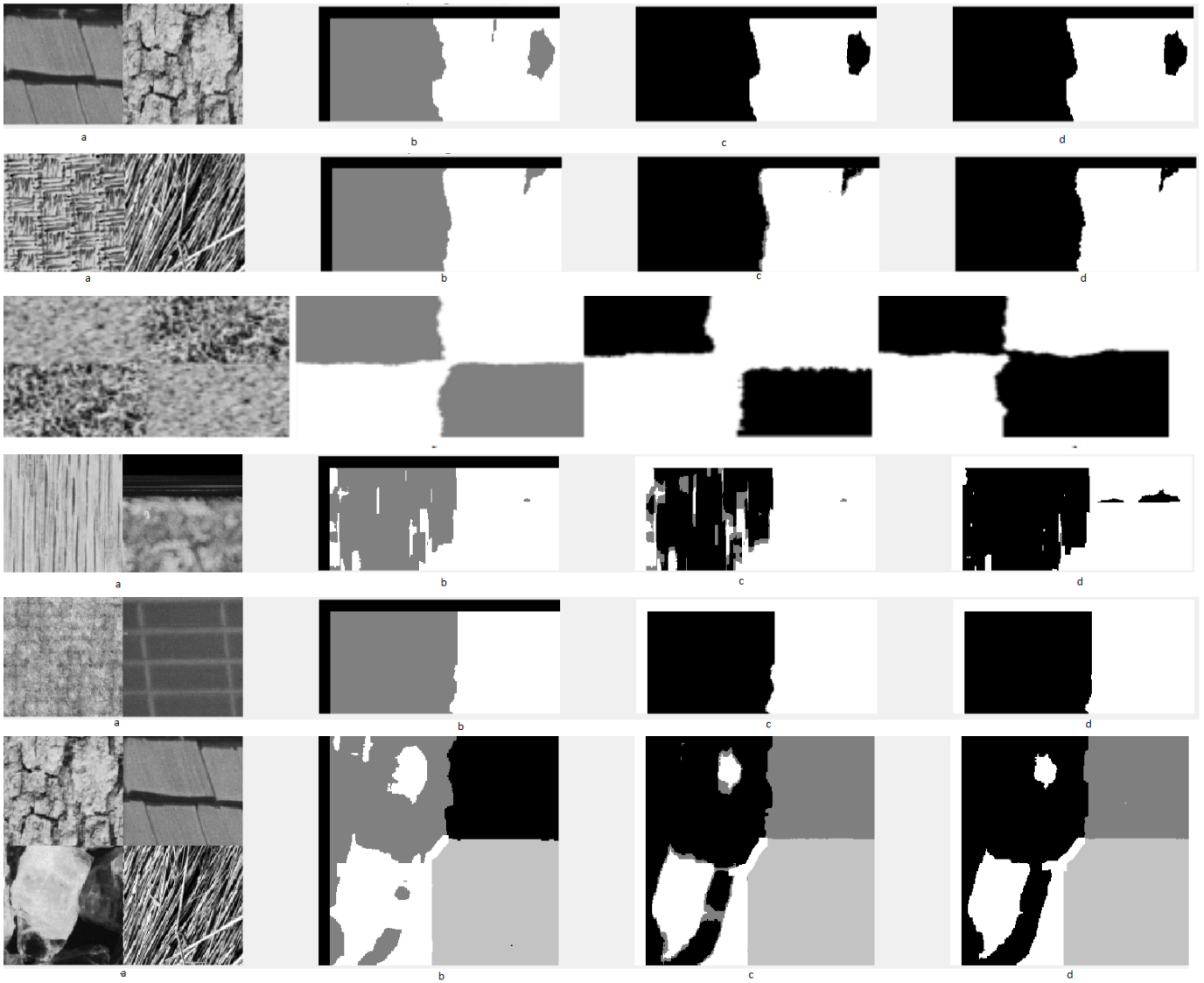


Fig. 3: (a) original images (b) 2D-RCA Knn results algorithm,(c) svm results (d) distance algorithm results .

V. CONCLUSION

The main objective of this paper is to propose a new approach for texture classification based on the estimation of the parameters of a 2D-RCA model using the GMM estimation. In the first step, mathematical definitions and the model properties have been given. The estimated parameters have been incorporated into a classification scheme. The proposed approach uses few textures features with a relatively good quality of classification. To conduct the experimentation different window sizes have been considered to study the effect on classification accuracy. The obtained results have been compared with other approaches results to evaluate the performance of the proposed algorithm. the efficiency of the proposed algorithm has been proved.

REFERENCES

- [1] H.-C. Lin, L.-L. Wang, and S.-N. Yang, "Regular-texture image retrieval based on texture-primitive extraction," *Image and Vision Computing*, vol. 17, no. 1, pp. 51–63, 1999.
- [2] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *IEEE Access*, vol. 7, pp. 8975–9000, 2019.
- [3] A. Mehnert, R. Moshavegh, K. Sujathan, P. Malm, and E. Bengtsson, "A structural texture approach for characterising malignancy associated changes in pap smears based on mean-shift and the watershed transform," in *2014 22nd International Conference on Pattern Recognition*. IEEE, 2014, pp. 1189–1193.
- [4] G. R. Cross and A. K. Jain, "Markov random field texture models," pp. 25–39, 1983.

- [5] Y. Q. CHEN, "A thesis submitted for the degree of doctor of philosophy," Ph.D. dissertation, Citeseer, 1995.
- [6] A. A. MAAROUF and F. HACHOUF, "Robust characteristics for texture classification," 2021.
- [7] R. M. Haralick *et al.*, "Statistical and structural approaches to texture," pp. 786–804, 1979.
- [8] J. Chen, S. Shan, C. He, G. Zhao, M. Pietikäinen, X. Chen, and W. Gao, "Wld: A robust local image descriptor," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 9, pp. 1705–1720, 2009.
- [9] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [10] T. Toyoda and O. Hasegawa, "Texture classification using extended higher order local autocorrelation," in *International Workshop on Texture Analysis and Synthesis*, 2005, pp. 131–136.
- [11] K. Mikolajczyk, A. Zisserman, and C. Schmid, "Shape recognition with edge-based features," in *British Machine Vision Conference (BMVC'03)*, vol. 2. The British Machine Vision Association, 2003, pp. 779–788.
- [12] A. W. Lohmann, "Image rotation, wigner rotation, and the fractional fourier transform," *JOSA A*, vol. 10, no. 10, pp. 2181–2186, 1993.
- [13] S. Kharfouchi, "Statistical inference in spatial bilinear processes," 2012.
- [14] H. H. Kelejian and I. R. Prucha, "A generalized moments estimator for the autoregressive parameter in a spatial model," pp. 509–533, 1999.
- [15] A. Boulemnadjel, F. Hachouf, and S. Kharfouchi, "Gmm estimation of 2d-rca models with applications to texture image classification," pp. 528–539, 2015.
- [16] J. Conlisk, "Stability in a random coefficient model," *International Economic Review*, pp. 529–533, 1974.
- [17] —, "A further note on stability in a random coefficient model," *International Economic Review*, pp. 759–764, 1976.
- [18] J. Anděl, "Autoregressive series with random parameters," *Mathematische Operationsforschung und Statistik*, vol. 7, no. 5, pp. 735–741, 1976.
- [19] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [20] P. Brodatz, *Textures: a photographic album for artists and designers*. New York: Dover Pub., 1966.

Automatic System Inspection of Surface defects using Gray-level Co-occurrence Matrix and Decision Tree Model

Adel Boudiaf¹, Khaled Harrar², Touba Mostefa Mohamed³, Rachid Zaghdoudi¹, Slimane Ziani¹

¹ *Research Center in Industrial Technologies CRTI, cheraga 16014 Algeria*
Address

a.boudiaf@crti.dz

R.Zaghdoudi@crti.dz

S.Ziani@crti.dz

² *LIST Laboratory, University M'HamedBougara of Boumerdes, 35000, Algeria*

khaled.harrar@univ-boumerdes.dz

³ *LI3C Laboratory University of Biskra, Algeria*

mostefa.touba@univ-biskra.dz

Abstract—As manufacturing speed increases in the sector of the steel industry, speedy and precise product inspection turns more important. Many industries such as aerospace, ships, submarines, and automotive reject the steel strips which have surface defects because a minor defect in these surfaces might result in Human and material losses in a later stage. In this context, we propose in this work to develop an intelligent recognition system of surface defects for hot-rolled steel strips images using gray-level co-occurrence matrix (GLCM). Nineteen (19) features were derived from GLCM technique. The classification system was based on the decision tree model (DTM). A total of 1800 images were analyzed and classified into six categories of defects. The results showed that our classifier model can be used easily for effective screening of surface defects for hot-rolled steel strips with very high classification accuracy up to 98.11%. In addition, we addressed through this research a comparative study between the proposed classification model and the rest of the modern classification models. This study highlighted the efficiency and effectiveness of our proposed model for the classification of surface defects.

Keywords—Decision Tree Model, Gray-level Co-occurrence Matrix, Defect classification, Steel surface inspection.

I. INTRODUCTION

Surface defects plays an important role on the performance and the quality in industry. For that reason, many industries such as aerospace, ships, submarines, and automotive reject the steel strips which have surface defects. In fact, a minor defect in these surfaces might result in Human and material losses in a later stage [1]. Inspection of surface defects by Human operators [2] is not adequate, slow, variable, and subjective. Studies [3,4] have proven that the inspection using a manual manner can only detect 65-70% of the defects. So, developing automatic inspection systems is an important issue in the metallic industry.

Currently, many defect recognition algorithms can be used in system of inspection of steel surface defects such as k-nearest

neighbor (KNN) [4], Artificial neural network (ANN) [5], fuzzy inference system (FIS) [6], support vector machine (SVM) [7], K-mean Algorithm [8], Decision tree [9], etc.

In recent years, many researchers developed automated inspection systems based on machine vision. Luiz et al. (2010) adopted Principal Component Analysis (PCA) as features extractor, and Self-Organizing Maps (SMO) as a classifier to classify six classes of the hot-rolled steel surface defects with a classification accuracy of 87% [10]. Sharifzadeh et al. (2008) used image processing algorithms for detecting four popular classes of steel defects and the result of the prediction was up to 90% [11]. Hongbin et al. (2004) and Keesug et al. (2006) applied support vector machine (SVM) as a classifier of the inspection system, SVM gave better performance than ANN for their samples on hot rolling steel and achieved an accuracy of 90% [12,13]. Decision trees are a recent and effective method of data mining to predict a qualitative variable using variables of any type (qualitative and/or quantitative). In this context, this paper aims to construct an automatic surface defect inspection system for hot-rolled flat steel (Fig.1) based on the GLCM and DTM outperforming the subjective traditional human inspection procedures. The system was successfully trained and validated using 1800 images divided into six groups of defects. These defects include scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In), and scratches (Sc). The experimental results show that most surface defects can be effectively recognized and the highest recognition rate of the prediction was up to 98.11%.

The contributions of this work are the combination of the GLCM and DTM techniques applied to hot-rolled steel strips images, and improve the performance of the classification system compared to the literature.

The remainder of this paper is organized as follows: Section 2 presents our proposed technique. Section 3, presents the

experimental results of our work. Finally, conclusions are provided in Section 4.

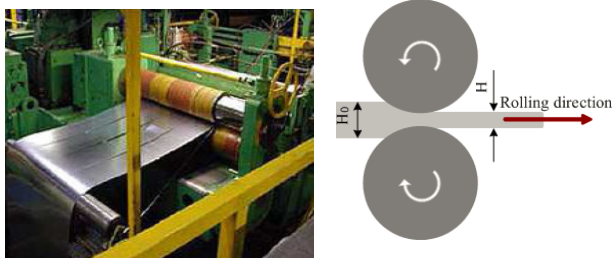


Fig.1. Rolling process

II. PROPOSED METHOD

The proposed technique consists of three major steps:

- The first step is image acquisition. In this work, we have used hot-rolled steel images represented in Fig. 2, with six classes.
- The next step is extraction the features of the images. This is done with the co-occurrence matrix. In this work, we have used the GLCM matrix to extract 19 features which are: Autocorrelation, Cluster Prominence, Cluster Shade, Contrast, Correlation, Difference entropy (Dent), Difference variance, Energy, Entropy, Dissimilarity, Homogeneity, Information measure of correlation 1, Information measure of correlation 2, Inverse difference, Maximum probability, Sum average, Sum entropy, Sum of square, Variance, Sum variance.
- In the final step, the decision tree model is used to classify the different steel surface defects.

The flowchart of the proposed technique is shown in Fig. 3.

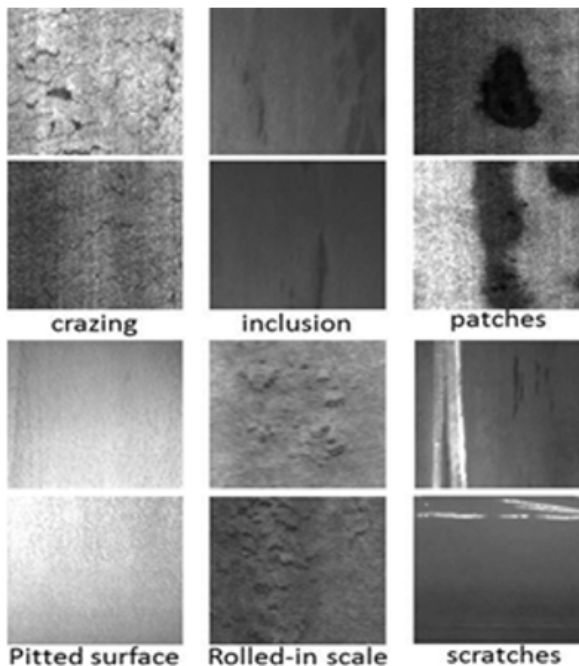


Fig.2. Samples images of six typical surface defects classes

A. Features extraction

Feature extraction is the most important step of any system of pattern recognition. Its main function is to find an appropriate representation of the pattern under study. Many methods can be used to extract the features of an image such as Histogram of Oriented Gradients (HOG), Discrete Cosine Transform (DCT), Wavelet Transform (WT), and Gray-level co-occurrence matrix (GLCM) [14-17]. In view of the capabilities and features provided by the Gray-level co-occurrence matrix (GLCM) we have used it to extract the feature of the images.

1) *Gray-level co-occurrence matrix (GLCM)*: GLCM is a matrix in which the number of rows and columns is equal to a number of gray levels in the picture. The properties of the textures in the statistical analysis are determined by the pixel density distribution at certain locations relative to each other in the image [17-20].

The distance (ℓ) between the pairs of adjacent pixels and the direction or angle (α) for some are significant in the GLCM calculation. Fig. 4 illustrates the GLCM of an image of a 5×5 matrix containing four Gray levels (0, 1, 2, and 3).

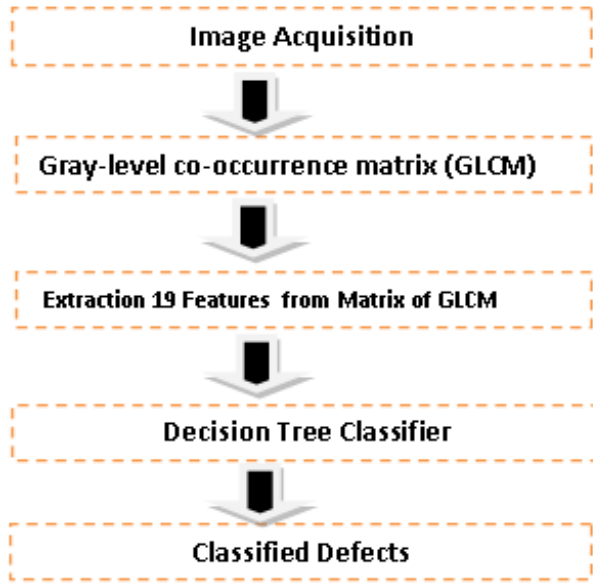


Fig.3. Block diagram of the proposed system

In Fig. 4, the pairs of pixels 0 and 1 are repeated twice at a distance of 1 and an angle of 0°. The GLCM matrix was formed from the value 2 for row 0 and column 1. In the same way the GLCM matrix of row 0 and column 0 were valued at 2. We can calculate GLCM for the directions 45°, 90°, and 135° as shown in Fig. 5.

Matlab was used to calculate the cost matrix (GLCM) or what is known as the probability density function. In this study, we have not used the matrix of gray level co-occurrence (GLMC) directly as a feature descriptor of the image. Rather than we have used this matrix to extract 19 second-order statistical texture features to describe the steel surface defects images.

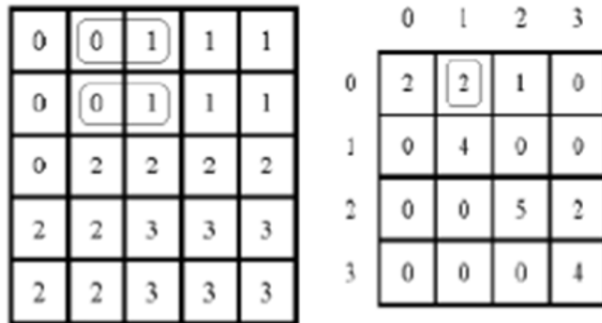


Fig.4. Example of the GLCM matrix estimated for an image

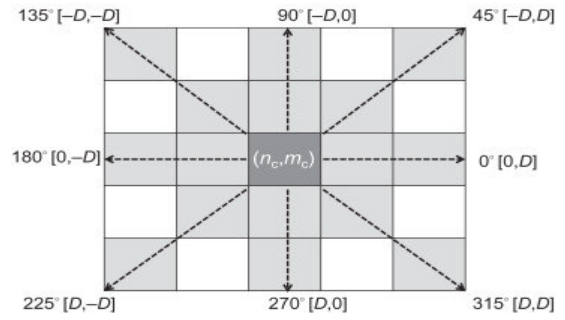


Fig.5. Different trends of pixel values when calculating GLCM

Some equations of features extracted from the GLCM method are given in the following equations:

$$Variance = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} (i - \mu)^2 * P(i, j) \quad (1)$$

$$Entropy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} P(i, j) * \log(P(i, j)) \quad (2)$$

$$Correlation = \frac{\sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{i, j\} * P(i, j) - \{\mu_x * \mu_y\}}{\sigma_x * \sigma_y} \quad (3)$$

$$Contrast = \sum_{n=0}^{G-1} n^2 \{ \sum_{i=1}^G \sum_{j=1}^G P(i, j) \}, |i - j| = n \quad (4)$$

$$Dent = - \sum_{i=0}^{G-1} P_{x+y}(i) \log(P_{x+y}(i)) \quad (5)$$

B. Features extraction

This technique is simple to use, easy to understand, visualize, implement, and offers many advantages compared to other decision-making tools. Decision trees consist of two steps. Firstly, the method constructs a tree with the training data. Secondly, the tree is examined to determine the class to which the record belongs. There are nodes in the tree which provide a test on an attribute. An outcome of the test is given by a branch, and a class is provided by a leaf node. [21-23].

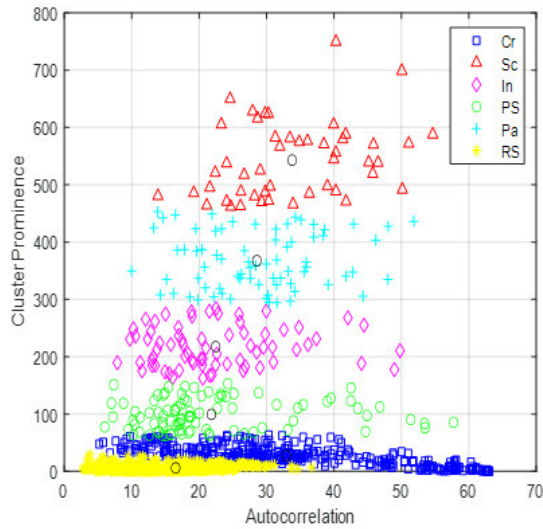


Fig.6. Visualization of the data

TABLE I. GLCM FEATURES EXTRACTED FROM THE IMAGES

GLCM feature	Cr	Pa	PS	In	RS	Sc
Autocorrelation	28.19	9.62	26.20	28.06	20.6	17.35
Cluster prominence	37.61	5.09	487.86	68.57	4.40	91.77
Cluster Shade	1.48	0.10	18.47	0.022	0.28	11.02
Contrast	0.46	0.07	0.37	0.11	0.16	0.16
Correlation	0.79	0.89	0.96	0.96	0.77	0.90
Difference entropy	0.75	0.26	0.69	0.34	0.45	0.41
Difference variance	0.28	0.06	0.25	0.09	0.14	0.15
Dissimilarity	0.43	0.07	0.34	0.10	0.16	0.13
Energy	0.12	0.45	0.08	0.18	0.33	0.35
Entropy	2.40	1.14	2.77	1.92	1.41	1.51
Homogeneity	0.79	0.96	0.83	0.94	0.92	0.93
Information measure of correlation1	-0.33	0.68	-0.57	-0.75	-0.4	-0.63
Information measure of correlation2	0.79	0.83	0.94	0.95	0.74	0.87
Inverse difference	0.79	0.96	0.84	0.94	0.91	0.93
Maximum probability	0.21	0.64	0.17	0.27	0.42	0.54
Sum average	10.46	6.11	9.39	10.30	9.01	8.14
Sum entropy	2.05	1.09	2.49	1.85	1.29	1.42
Sum of squares variance	1.07	0.33	4.33	1.56	0.36	0.85
Sum variance	3.83	1.24	16.90	6.13	1.28	3.25

III. RESULTS AND DISCUSSION

In this section, the experimental results are presented. For evaluating the system performance and to know the output unit response accuracy, the classification matrix is developed considering the following points:

- We have used in this work 1800 images, with 300 images of each class (defects).
- The 6 typical defects used as the output of the classifier model (DTM) were:scale (RS), patches (Pa), crazing (Cr), pitted surface (PS), inclusion (In), and scratches (Sc).

Table I shows the GLCM attributes extracted from the images for the six classes. To evaluate the performance of our classifier model a confusion matrix was carried out. It is a table containing information and details about the actual ratings (classified by humans) and the predictive ratings that the classifier predicts. From the confusion matrix (Table II) we can conclude the following points:

- The classifier made a total of 1800 predictions.
- The number of cases correctly predicted for class “Cr” and its actually belongs to the “Cr” class is 299 cases.
- The number of cases correctly predicted for class “In” and its actually belongs to the “In” class is 294 cases.
- The number of cases correctly predicted for class “Pa” and its actually belongs to the “Pa” class is 294 cases.
- The number of cases correctly predicted for class “Ps” and its actually belongs to the “Ps” class is 292 cases.
- The number of cases correctly predicted for class “Rs” and its actually belongs to the “Rs” class is 296 cases.
- The number of cases correctly predicted for class “Sc” and its actually belongs to the “Sc” class is 291 cases.
- The diagonal elements of the confusion matrix showsthe correctly classified groups.
- The total number of cases correctly predicted is 1766 cases.
- The overall accuracy of the classifier is 98.11%.

TABLE II. CONFUSION MATRIX

		Predicted Cass (Group)					
		Cr	In	Pa	Ps	Rs	Sc
Actual Class (Group)	Cr	299	0	1	0	0	0
	In	0	294	0	3	0	3
	Pa	4	0	294	0	0	2
	Ps	0	4	0	292	2	2
	Rs	1	3	0	0	296	0
	Sc	1	3	2	2	1	291

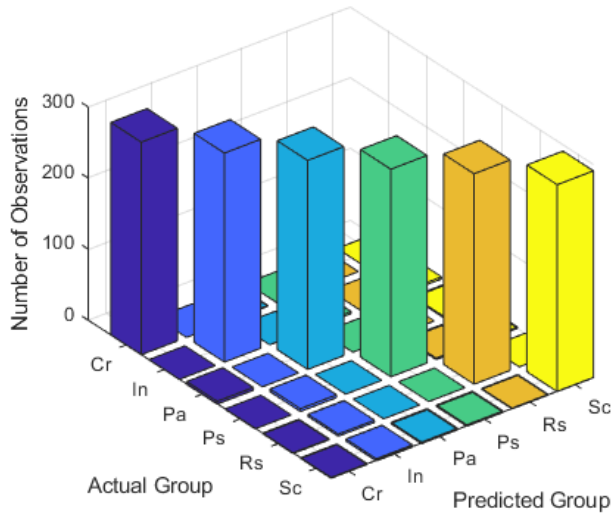


Fig.7. Confusion matrix in 3D

Figure 8 shows the visualized surface for the different classes using the decision tree model.

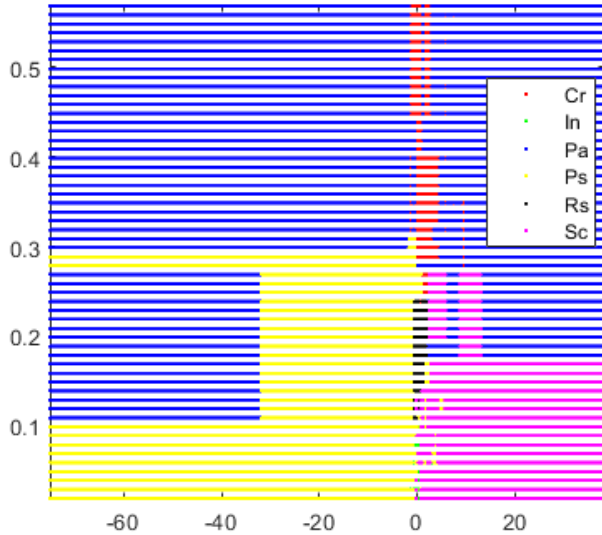


Fig.8. A decision surface for the different classes

To evaluate the performance of the classifier optimally, we have used four metrics namely: Accuracy, Precision, Recall, and F-Measure.

Accuracy: It gives the overall accuracy of the classifier, In other words, it gives the total samples that were correctly classified by the classifier. To calculate accuracy, we use the following formula:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

Precision or Sensitivity: represents the classifier's success rate in classifying the correct cases for each class (or group) of the predicted classes, and is calculated as follows:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

Recall: It represents the classifier's success rate in not classifying wrong cases to another class (or group) of the predicted classes, and is calculated as follows:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

F1-Score: This metric is derived from the precision and recall rate. The value of F1-Score ranges from 0 to 1, where 1 represents the best model output and 0 represents the worst model output. It is estimated as follows:

$$F1 - Score = \frac{2TP}{2TP + FP + FN} \quad (9)$$

Examining Table III, it can be noticed that the classifier's performance was excellent. The results of the "precision" revealed high rates of correct detection for all six classes, ranging between 96% and 98%. This reflects the classifier's success rate in classifying the correct cases for each class of the predicted groups. Same findings for the "recall". It provided high rates, ranging between 97% and 99%, which translates the classifier's success rate in not classifying wrong cases to another class of the predicted classes. Regarding "F1 Score", the results were very close to "1", as the results were limited between 0.97 and 0.98, which reflects the efficiency and effectiveness of the classifier.

TABLE III. PERFORMANCE EVALUATION OF THE PROPOSED MODEL

Classes	Precision (%)	Recall (%)	F1-Score
Cr	98.03	99.66	0.9884
In	96.71	98	0.9735
Pa	98.99	98	0.9849
Ps	98.31	97.3	0.9782
Rs	98.99	98.66	0.9883
Sc	97.65	97	0.9732

To highlight the efficiency of the proposed classification model, we conducted a comparative study with some models in the same field (Table IV). The results showed that our model achieved a good classification rate of 98.11 %, compared to other models. Luiz et al (2010) adopted Principal Component Analysis as a features extractor, and Self-

Organizing Maps as a classifier to classify six classes of the hot-rolled steel surface defects with a classification accuracy of 87%. Ashour et al. (2019) developed a model to classify the hot-rolled steel surface defects by combining the DST and GLCM for feature extraction. PCA was applied to reduce the dimension of the feature vector. The SVM classifier was finally trained to classify the surface defects. The proposed model provided a classification rate of 94.11%. Our model was able to achieve the highest performance compared to all modern classification models with a rate of 98.11% outperforming the methods existing in the literature. This demonstrates the performance of the proposed method (Table IV).

TABLE IV. COMPARISON WITH THE STATE-OF-THE-ART FINDINGS USING NEU SURFACE DEFECT DATABASE

Author	Features Extractor	Classifier	Accuracy (%)
Boudiaf et al. [5]	GLCM	KNN	91.12
Luiz et al.[10]	PCA	SOM	87
Ashour et al. [24]	DST-GLCM	SVM	94.11
Proposed model	GLCM	DTM	98.11

IV. CONCLUSION

In this paper, an automatic system inspection of surface defects on hot rolled steel using GLCM and DTM was proposed to overcome the limits of the traditional human inspection procedures. The experimental results showed that our method is effective in classifying the surface defects. Moreover, the rate of accuracy obtained by this algorithm was 98.11%. May this work will assist the engineers to classify the surface defects of the hot-rolled steel strip quickly and correctly compared with quality control of steel products by the human vision which remains tedious, fatiguing, bit fast, bit robust, sketchy, dangerous, or impossible. In future, we plan to increase the number of the classifiers to improve the performance of the system.

REFERENCES

- [1] C.C. Shao, G.W. Xu, H. Xu, M.H. Liao, H.L. Zhao, Z. J. Zhang, Q. Zhao, "Research of Recognition Method for Surface Defects of Hot-Rolled Round Steel Based on Image Processing", *Advanced Materials Research*, vol. 1090, pp 84–89, 2015.
- [2] A. Boudiaf, K. Boubendira, K. Harrar, A. Saadoune, H. Ghodbane, A. Dahane, O. Messai, "Surface Image compression of surface defects of the hot-rolled steel strip using Principal Component Analysis", *Matériaux & Techniques*, vol. 107(2), n. 203, 2019.
- [3] K. Schickanz, "Automatic fault detection possibilities on nonwoven fabrics," *Melliand Textilberichte*, vol. 74, pp. 294-295, 1993.
- [4] A. Boudiaf, R. Zaghdoudi, Z. Mentouri, S. Teleb, "An Automatic System of Surface Defect Classification of Hot Rolled Steel", *Proceedings IEEE of The 3rd International Conference on Electromechanical Engineering (ICEE'2018)*, November 20-21, Skikda, Algeria. ISBN: 978-9931-9447-4-4, 2018.
- [5] F. Pernkopf, "Detection of surface defects on raw steel blocks using Bayesian network classifiers". *Pattern Anal. Appl.* 7, pp. 333–342, 2004.
- [6] A. Borselli, "A fuzzy inference system applied to defect detection in flat steel production" *proceedings of the 2010, International Conference on Fuzzy Systems*, no.14, pp. 1–6, 2010.
- [7] R. Zaghdoudi, H. Seridi, A. Boudiaf, S. Ziani, "Binary Gabor pattern (BGP) descriptor and principal component analysis (PCA) for steel surface defects classification". *Proceedings IEEE of International Conference on Advanced Aspects of Software Engineering (ICAASE)*. pp. 1917–1927, 2021.
- [8] H. Sidqi, J. Kakbra, "Image Classification Using K-mean Algorithm", *International Journal of Emerging Trends & Technology in Computer Science*, vol. 3, no. 1, pp. 38-41, 2014.
- [9] S. R. Aghdam, E. Amid, M. F. Imani, "A fast method of steel surface defect detection using decision trees applied to LBP based features", *Proceedings IEEE of 7th IEEE Conf. Ind. Electron. Appl. (ICIEA)*, pp. 1447-1452, Jul. 2012.
- [10] A. Luiz, L. Flavio, E. Paulo, "Automatic detection of surface defects on rolled steel using Computer Vision and Artificial Neural Networks", *Proceedings IEEE of 36th Annual Conference on IEEE Industrial Electronics Society*, Glendale, USA, November 7-10, 2010.
- [11] M. Sharifzadeh, S. Alirezaee, R. Amirfattahi, S. Sadri, "Detection of steel defect using the image processing algorithms", *Proceedings of 12th IEEE International Multitopic Conference*, Karachi, Pakistan, December 23-24, 2008.
- [12] J. Hongbin, L. Yi, S. Jianjun, "An Intelligent Real-time Vision System for Surface Defect Detection", *Proceedings IEEE of 17th International Conference on Pattern Recognition*, Cambridge, UK, August 23-26, 2004.
- [13] C. Keesug, K. Kyungmo, J. Lee, "Development of defect classification algorithm for POSCO rolling strip surface inspection system", *SICE-ICASE International Joint Conference*, Busan, Korea, Oct. 18-21, 2006.
- [14] S. Karthikeyan, M.C. Pravin, B. Sathyabama, M. Mareeswari, "DWT Based LCP Features for the Classification of Steel Surface Defects in SEM Images with KNN Classifier", *Australian Journal of Basic and Applied Sciences*, vol 10, n 5, 2016.
- [15] K. Roopashri, S. Meghana, "An Efficient Face Recognition using DWT and FFT Algorithm", *IJSTE - International Journal of Science Technology & Engineering*, vol. 2, no. 12, pp. 255-258, 2016.
- [16] W. Zhou, S. Gao, L. Zhang, and X. Lou, "Histogram of oriented gradients feature extraction from raw Bayer pattern images", *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 67, no. 5, pp. 946–950, 2020
- [17] A. Borselli, "Utilizing Discrete Wavelet Transform and Discrete Cosine Transform for Iris Recognition", *proceedings IEEE of the 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, Monastir, Tunisia pp. 1–6, 2020.
- [18] K. Sharma, E. Priyanka, A. Kalsh, K. Saini, "GLCM and its Features", *International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)*, Vol.4, no. 8, pp. 2180–2183, 2015.
- [19] Y. Guo, Z. Sun, H. Sun, X. Song, "Texture feature extraction of steel strip surface defect based on gray level co-occurrence matrix", *proceedings IEEE of the International Conference on Machine Learning and Cybernetics (ICMLC)*, Guangzhou, 2015, pp. 217-221, 2015.
- [20] R. Zaghdoudi, H. Seridi, A. Boudiaf, "Detection and classification of steel defects using machine vision and SVM classifier". *Proceedings IEEE of 2nd International Conference on Automatic control, Telecommunication and Signals (ICATS'17)*, pp. 1-5, 2017.
- [21] A. M. Ahmed, A. Rizaner, A. H. Ulusoy, "A Decision Tree Algorithm Combined with Linear Regression for Data Classification", *proceedings IEEE of the International Conference on the Computer Control Electrical and Electronics Engineering (ICCEEE) 2018 International Conference on*, pp. 1-5, 2018.
- [22] A. Wajid, N.K Singh, P. Junjun, M.A. Mughal, "Recognition of ripe unripe and scaled condition of orange citrus based on decision tree classification", *proceedings IEEE of International Conference on*

Computing Mathematics and Engineering Technologies (ICOMET'18), pp. 1-4, 2018.

- [23] S. Naidu, "classification of defects in software using decision tree algorithm," International Journal of Engineering Science and Technology (IJEST), Vol. 5 no.06 June 2013.
- [24] M.W. Ashour, F. Khalid, A.A. Halin, L.N. Abdullah, S. H. Darwish, "Surface Defects Classification of Hot-Rolled Steel Strips using Multi-directional Shearlet Features", Arabian Journal for Science and Engineering (AJSE). Vol 44, pp. 2925–2932, 2019.

Deep Transfer learning for Arabic handwritten text recognition: A comparative study

Soumia Faouci, Djamel Gaceb, Mohammad Haddad

^{1 and 2} *Laboratoire d'Informatique, de Modélisation, d'Optimisation et des Systèmes Électroniques (LIMOSE), Faculté des sciences*

Boumerdès, Algérie

³ *Lab LIRIS, UMR CNRS 5205, University of Claude Bernard Lyon 1, F69622*

Villeurbanne, France

¹ s.faouci@univ-boumerdes.dz

² d.gaceb@univ-boumerdes.dz

³ mohammed.haddad@univ-lyon1.fr

Abstract— Arabic handwritten text recognition is an active research field. Various challenges facing these recognition systems most notably the freestyle and the infinite variability of human handwriting. This paper compares the performance of well-known convolutional neural network learning models for handwritten Arabic text recognition. We applicate the idea of transfer learning and we use the pre-trained models with feature extractor and fine-tuning strategies. In this study, the last fully connected layer is replaced with another fully connected layer to recognize a number of classes according to the tested dataset. To compare these models we used different databases, which are developed for the recognition of Arabic handwritten words, letters, or digits. These databases are CMARDB 3.3.1 and MADBASE for digit recognition, OIHACDB and AIA9K for character recognition, and AHDB for word recognition.

Keywords— Arabic handwritten text recognition, Deep learning (DL), Convolutional Neural Network (CNN), Transfer Learning (TL), Pre-trained models.

I. INTRODUCTION

Handwriting recognition becomes a central problem in the sector of pattern recognition and artificial intelligence in recent years. It continues to be a very challenging area due to its enormous application. Many works for different languages have been proposed in this field but most of these works were concentrated on Latin languages. Arabic is among the languages which have not had much interest by researchers in the fields of recognition. There are many levels of Handwritten text recognition; sentences recognition, word recognition, character recognition, numerals recognition, etc [1].

Handwritten digit recognition is the main component for many applications in various domains; checks verification, office automation, business, postal address reading, printed postal codes, and data entry applications [2]. The recognition of handwritten digits presents many difficulties related to the unlimited variation in human handwriting.

Arabic Handwritten text recognition is a more difficult task that faces several challenges, including the structural similarity of some handwritten Arabic samples. The main difficulties in Arabic handwriting recognition revolve around

variations in the writing between different individuals and also the intra variability of a single writer.

The handwriting recognition problem has been studied using different classifiers such as SVM [3, 4], KNN [4], Neural Network [4], and Convolutional Neural Network (CNN) [5, 6]. Great interest has been given to CNN in recent times. CNN is a category of Deep Learning methods that have shown unlimited success in image-related problems such as images classification, image segmentation, object detection, etc.

The important step in CNN is training the network. Wherever, training CNNs involves many samples, which makes computational demands on the system.

Considering the enormous resources necessary to train deep learning models or the large and challenging datasets used to train these models, transfer learning may be the best technique allowing to speed up the training and improve the performance of deep learning models.

Transfer learning (TL) can be defined as the ability of a system to recognize and apply knowledge learned from previous tasks on other tasks or similar areas. In the machine learning domain, TL is a method, which allows to reuse a model already developed for a task as a starting point for another task. The idea of the transfer learning method is described in Fig.

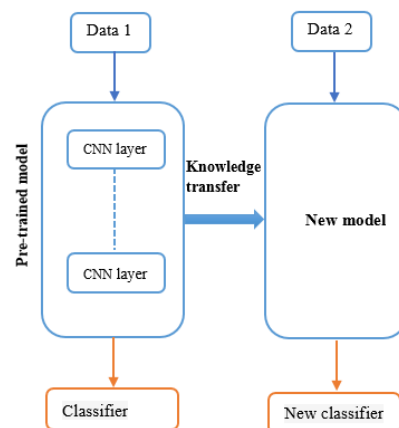


Fig. 1. Transfer learning technique

The training of these models named pre-trained models has already been carried out on a large dataset such as the famous ImageNet database with millions of images. These models can be exploited in two different ways: (1) feature extractor and (2) fine-tuning. Feature extractor allows us to use a pre-trained model without its final layer as a fixed feature extractor for other tasks. The pre-trained model's weighted layers are used to extract features and the weights of the model's layers are not updated during training with new data for the novel task. In the fine-tuning strategy we replace the final layer, we enrich the network with new layers and we selectively retrain some of the previous layers.

This work evaluates the performance of several pre-trained models for the recognition of Arabic handwritten words, letters, and digits

The rest of this paper is structured as follows. Section 2 describes the characteristics of Arabic language. Section 3 gives an overview of related works in the field of Arabic text recognition. Section 4 presents details of the proposed methods. Section 5 provides an analysis of the experimental results and discusses the results of the evaluated techniques. Finally, section 5 elaborates a conclusion of this study.

II. ARABIC CHARACTERISTIC'S

Arabic language represents the mother language in various countries. It has several characteristics that cause problems for a recognition system. These characteristics can be summarized in the following:

- The orientation of the writing is from right to left.
- Arabic alphabet has 28 letters with a variety of shapes where each character has more than two shapes depending on the position of the character in the word (at the beginning, in the middle, or at the end of the word). Six characters (ا, ب, ت, ز, ر, د) have just one shape. Fig. 2 displays the letter "taa" with different forms according to its position in the word.

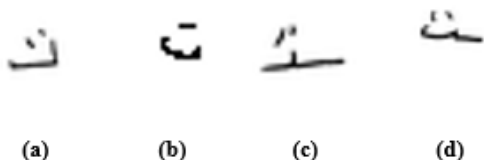


Fig. 2. Letter "taa" with different forms according to its position in the word. (a) in the beginning, (b) and (d) at the end, (c) in the middle.

- Diacritics points play an important role in the Arabic language. Several characters have the same main form and differ in the position (above or below these characters) or the number of these points. Fig. 3 shows an example of Arabic letters with the same main part and different numbers of diacritics points.



Fig. 3. Arabic letters with the same main part and different numbers of diacritics point

Fig. 4 shows an example of Arabic letters with the same main form and different positions of diacritics points.



Fig. 4. Arabic letters with the same main form and different positions of diacritics points

- Other diacritic marks that affect the recognition exist like hamza or madda which can be placed above or below characters (e.g. Fig. 5). These marks can change the meaning of a word.

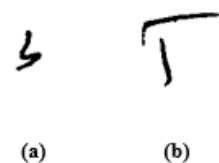


Fig. 5. Other diacritics marks in Arabic language. (a) Hamza, (b) madda.

- The combination of two or more letters with horizontal or vertical overlapping between component characters is called a ligature. Examples of ligature in Arabic language are presented in Fig. 6.



Fig. 6. Examples of ligature in Arabic language

- For Arabic handwritten digits, there is a structural similarity in writing between the different digits.

III. RELATED WORKS

Numbers of methods were improved in the literature to recognize Arabic text.

In [9], Ashiquzzaman and al. proposed a hybrid method that combined Restricted Boltzmann machine (RBM) and CNN. RBM was used as a features extractor and CNN was used to classify handwritten Arabic digits of the CMATERDB 3.3.1 dataset [7]. The authors used the data augmentation technique and also the activation function 'Elu'. The proposed method achieved an accuracy of 99.4%.

For Arabic handwritten word recognition, Alia and al. [10] presented a technique based on scale-invariant feature transform (SIFT) and SVM classifier without segmentation to sub-letters. AHDB dataset [11] was used to test this work and obtained a recognition rate of 99.08%.

For recognizing Arabic handwritten characters, Boufenar and al. proposed a bio-inspired method [12]. The proposed work is composed of three main modules. The first module is pre-processing, the extraction of primitives (structural and statistical characteristics) and recognition using an artificial immune recognition system (AISR). The authors had also collected a new database of isolated Arabic handwritten

characters called OIHACDB (Offline Isolated Handwritten Arabic Characters DataBase). The collected dataset exists in two versions; OIHACDB-28 and OIHACDB-40 (28/40 classes of Arabic handwritten letters). The recognition rate attained by this system is 93.2% on OIHACDB-28 dataset. Later, Boufenar [13] has proposed architecture based on deep learning built with three convolutional layers and pooling layers followed by a fully connected layer and regularization with the dropout optimization operation. In the same work, they tested the pre-trained AlexNet model with two different ways: training from scratch and transfer learning.

In [14], the authors proposed a simplified CNN model that includes five convolution layers and uses batch-normalization and dropout operations to recognize Arabic handwritten isolated characters. In this work, OIHACDB-28, OIHACDB-40, and AIA9K datasets were used for the test. In addition, the authors evaluated various deep transfer learning pre-trained models and study different levels of fine-tuning strategy. The proposed CNN model achieved accuracies of 94.7%, 98.3%, and 95.6% for the test set of OIHACDB-28, OIHACDB-40, and AIA9K [15] datasets respectively.

Another work based on CNN was presented in [16]. The authors proposed a deep sequential CNN model to address the problem of recognizing offline handwritten Arabic text, including isolated digits, characters, and words. The proposed model consists of 34 building blocks, including one input layer, one output layer, five stacked convolutional layers blocks, and two fully connected hidden layers for non-linear classification. The comparative experimental demonstrated a promising results with a testing accuracy of 99.91%, 99.72%, 99.91%, 99.82%, 99.86%, and 99.95% using MADBase [17], CMATERDB, HACDB, and three types of SUST-ALT databases, including digits, characters, and words, respectively.

IV. PROPOSED WORKS

The methods evaluated in this research are based on the deep transfer learning CNN models. This section presents the procedure for applying the transfer learning in the context of Arabic handwritten text recognition.

A. Transfer learning for Arabic handwritten text recognition

The idea of transfer learning is to re-use neural network model trained on a problem in another similar problem. The pre-trained models were developed for standard computer vision benchmark datasets, such as the ImageNet, which contains 1.2 million labeled images belonging to 1000 categories. Transfer learning has the advantage of reducing the training time and the consumption of hardware resources.

To demonstrate the power of transfer learning method for classifying images of Arabic handwritten text, the performance of the most popular architectures namely; VGG16 [18], InceptionV3 [19, 20, 21], ResNet50 [22], MobileNet [23], EfficientNet [24] and DenseNet121 [25] is compared in this work.

In this paper, the transfer learning technique was applied with two scenarios. First by fine-tuning each of the exploited pre-trained models. In this strategy we initialize the network with a selected pre-trained network, other layers were added and the whole model was retrained. Second, by using the pre-trained models in features extractor mode, so we replace the final layer, we add some layers then we just retrain the newly-added last layers of the network. The steps of these two scenarios are presented in Fig. 7.

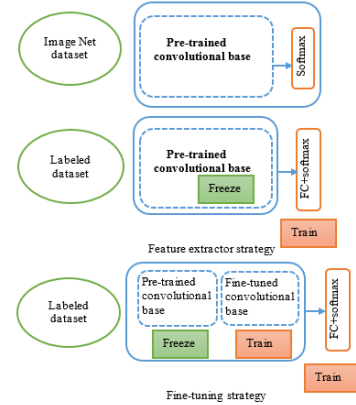


Fig. 7. Fine-tuning strategy, VS feature extractor strategy

1) *Features extractor scenario*: For all the used pre-trained models we retain the earlier layers as fixed features extractor for the tested datasets. In this mode of transfer learning technique, the last layer in the pre-trained models was replaced with a set of layers. Then, we just train the new layers without updating the weights of the pre-trained models. For ResNet50, EfficientNet, Inception V3, DenseNet121, and VGG16 models, the added layers are a fully connected layer with a filter size of 32*32, a ReLU layer to improve the non-linear problem, and another fully connected layer with N output neurons to enable the classification of N classes. In this work, the number of classes (N) differs according to the used dataset. Fig. 8 illustrates the proposed ResNet50, Inception V3, EfficientNet, DenseNet121, and VGG16 model's customization for Arabic handwritten text recognition.

For MobileNet model, the added layers are a batch normalization layer, a fully connected layer with a filter size of 32*32, a ReLU layer, a dropout operation with a probability of 0.2, and a last fully connected layer.

Fig. 9 describes the proposed MobileNet model's customization for Arabic handwritten text recognition.

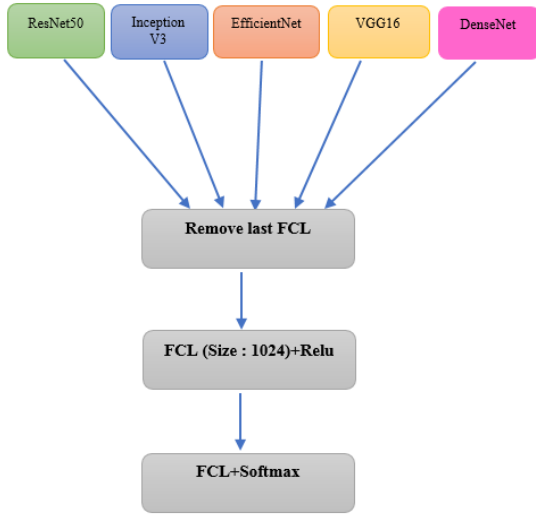


Fig. 8. Proposed model's customization for Arabic handwritten text recognition. ResNet50, Inception V3 EfficientNet, DenseNet121 and VGG16. FCL: fully connected layer.

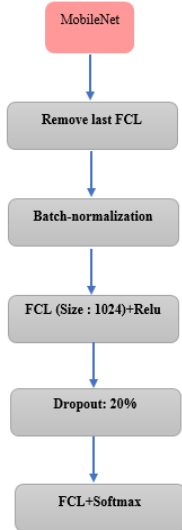


Fig. 9. Proposed MobileNet model's customization for Arabic handwritten text recognition. FCL: fully connected layer.

2) *Fine-tuning the pre-trained models*: In this mode, the final fully connected layer in each CNN architecture was swapped out and replaced with a softmax layer to match the number of categories of each tested dataset. We add a fully connected layer with filter size of $32*32$, and a ReLU layer. Finally, this new model (the pre-trained model and the added layers) is trained on the new datasets.

B. Training

When training the CNN models we used the standard cross-entropy loss function in combination with softmax as the last layer to provide probabilities of each class. As optimizer, we used Adam optimizer with a learning rate of 0.0001 for all the tested models. Each model was trained up to 200 epochs in maximum.

V. DATASETS

Several datasets were used to evaluate the proposed methods; numeral datasets, character datasets, and word dataset.

A. Numeral datasets

1) *CMATERDB 3.3.1*: The CMATERDB 3.3.1 Arabic handwritten digit dataset [7] was collected by researchers at the Jadavpur University from three different sources; class notes of students from different age groups, handwritten manuscripts of magazines, and a preformatted datasheet designed for a collection of handwriting samples [2]. This dataset contains 3000 images saved in RGB bitmap format, each image of size $32*32$. The 30,000 images are shared between 10 classes with 300 samples in each class. In this work, the CMATERDB 3.3.1 dataset is divided into two sets; 80% for the training and 20% for the testing process.

Table 1 shows the handwriting Arabic digits with their corresponding images from CMATERDB 3.3.1 DATASET.

TABLE I ARABIC HANDWRITTEN DIGITS with their CORRESPONDING IMAGES from CMATERDB 3.3.1 DATASET

Arabic digit	Latin digit	Example of image
٠	0	
١	1	
٢	2	
٣	3	
٤	4	
٥	5	
٦	6	
٧	7	
٨	8	
٩	9	

2) *MADBase*: The MADBase dataset [17] is based on MNIST dataset. It consists of 60,000 training images, and 10,000 test images, written by 700 writers from different institutions: Colleges of Engineering and Law, School of Medicine, the Open University (whose students span a wide range of ages), a high school, and a governmental institution.

B. Word dataset

The AHDB dataset [11] contains 105 folders including 315 documents written by 105 writers. It consists of words used for the numbers and quantities in checks filling. Also, it comprises some sentences that are used in writing checks in Arabic words. In addition, it contains the most popular words in Arabic writing and a free handwriting page from each writer's imagination.

C. Character datasets

- 1) *OIHACDB*: This dataset was developed by Boufenar [12, 13]. It was made from the famous IFN / ENIT dataset. The images are saved in a bitmap format with a size of 128×128 pixels. This dataset was proposed in two versions OIHACDB-28 and OIHACDB-40 (28 or 40 classes). OIHACDB-28 covers 5600 examples with 200 examples per class. OIHACDB-40, comprises 30000 images with 750 examples per class. In this work, OIHACDB-28 and OIHACDB-40 databases are divided into 75% for the training and 25% for the testing process.
- 2) *AlexU Isolated Alphabet (AIA9K) dataset*: Presented by Marwan Toriki and al. in [15]. It was collected from 107 writers where each writer wrote all the Arabic letters 3 times. It comprises 8737 images of the 28 Arabic alphabet letters. The AIA9K dataset was divided into three groups: 70% for training, 15% for validation, and 15% for testing.

Table 2 presents a description of the evaluated datasets.

TABLE II DESCRIPTION of the TESTED DATASETS

Dataset	Number of classes	Width/Height	number of images
CMATERDB 3.3.1	10	32×32	3000
MADBase	10	28×28	70000
AHDB	30	Variant	3046
OIHACDB-40	40	128×128	30000
OIHACDB-28	28	128×128	5600
AIA9K	28	32×32	8737

VI. RESULTS and DISCUSSIONS

The proposed methods were developed using Python programming languages (Keras library and Tensorflow environment).

The performance of the suggested works is estimated using the test accuracy measure. The testing accuracy measure corresponds to the fraction of the total samples that were correctly classified. It is calculated from the confusion matrix according to equation (1).

$$\text{Testing accuracy} = \frac{TN+TP}{TN+TP+FN+FP} \times 100 \quad (1)$$

Where TP is the number of true positive samples, TN is the number of true negative samples, FP is the number of false positive samples, and FN is the number of false negative samples.

The F1-score metric is also used. It is calculated with equation (2).

$$F1 - score = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (2)$$

$$\text{Where: } \text{precision} = \frac{TP}{TP+FP} \text{ and } \text{recall} = \frac{TP}{TP+FN}$$

For the evaluation strategy, we used automatic verification datasets. For that, we separate a portion of each dataset into validation data and training data and we evaluate the

performance of the tested models on the validation dataset each epoch.

For CMATERDB 3.3.1 and OIHACDB-28 datasets we have tested the two scenarios of transfer learning; fine-tuning and feature extractor. For other datasets, we have worked just with the feature extractor strategy.

A. Numeral datasets (CMATERDB 3.3.1 and MADBase)

Table 3 presents the test accuracies values achieved by the different CNN pre-trained models on the evaluated numeral datasets; CMATERDB 3.3.1 and MADBase.

TABLE III TEST ACCURACIES OBTAINED by the PRE-TRAINED MODELS and PREVIOUS WORKS for CMATERDB 3.3.1 and MADBASE DATASETS. FE: FEATURE EXTRACTOR, FT: FINE-TUNING

CNN model	CMATERDB 3.3.1		MADBase
	FE	FT	FE
ResNet50	95.6	96	98.2
EfficientNetB0	93.7	94.2	98.65
EfficientNetB1	94.9	95.1	98.6
DenseNet121	93	98.8	99
MobileNet	60.1	99.2	99.5
InceptionV3	85.3	87	97
VGG16	96.4	99.2	97.2
Deep CNN model: [16]	99.72		-
Deep CNN model: [16]	-		99.91

Table 3 illustrates that for CMATERDB 3.3.1 dataset, MobileNet and VGG16 models used in fine-tuning mode achieved the highest percentage of testing accuracy compared to other architectures with a test accuracy of 99.2%. The same MobileNet model attained the worst test accuracy when it was exploited in feature extractor mode. This result confirms what has been mentioned in the literature, that fine-tuning does not require too many images for training to produce a successful classification. It is necessary to note that MobileNet model gave an accuracy of 99.2% in fewer than 70 epochs of training (less than 1 hour). However, the VGG model is computationally expensive in comparison to MobileNet.

According to the achieved results for MADBase dataset, it can be observed that the MobileNet model is the most appropriate architecture for this dataset with a testing accuracy of 99.5 percent.

B. Word dataset

Table 4 shows the results of different pre-trained models applied on AHDB dataset with feature extractor mode.

TABLE IV TEST ACCURACIES for AHDB DATASET

CNN model	AHDB
ResNet50	87.6
EfficientNetB0	96
EfficientNetB1	95.8
DenseNet121	91.5
MobileNet	97.5
InceptionV3	93.1
SIFT+SVM: [10]	99.08

Comparing the results of different models evaluated on AHDB dataset, it can be clearly seen that MobileNet model gave the highest test accuracy value of 97.5%. Moreover, Table 4 depicts that the ResNet50 model was the worst for the AHDB dataset. EfficientNetB0 model also gave a promising result with a test accuracy of 96%. All the evaluated models did not improve the result obtained in previous works [10]. For this dataset, traditional techniques based on features extraction are currently the better solution.

C. Character datasets (OIHACDB and AIA9K)

Table 5 displays the test accuracies obtained by the evaluated pre-trained models and previous works for OIHACDB-28, OIHACDB-40, and AIA9K datasets.

TABLE V EXPERIMENTAL RESULTS on the OIHACDB and AIA9K DATASETS. FE: FEATURE EXTRACTOR, FT: FINE-TUNING

CNN model	OIHACDB-28		OIHACDB-40	AIA9K
	FE	FT	FE	FE
ResNet50	87	87.4	91	74.1
EfficientNetB0	92.8	98.29	94.4	89.9
EfficientNetB1	94.2	98.3	93.8	91
DenseNet121	91.07	93.5	94.5	81.6
MobileNet	95	97.6	96.25	91.5
InceptionV3	93.1	95.2	90.4	88.5
AlexNet : [13]	-	98.12	98.12	-
AlexNet from scratch: [14]	96		--	
CNN model: [14]	-	-	98.3	95.6
CNN model : [14]	-	-	-	95.6

As shown in Table 5, MobileNet architecture used in features extractor strategy obtained the most acceptable testing accuracy result for OIHACDB-40 and AIA9K datasets compared to other models except the previous works.

For OIHACDB-28 dataset, the best test accuracy value was 98.3% attained by EfficientNetB1 model trained with the fine-tuning scenario. This result is better than the results of previous works [13, 14]. We note that fine-tuning strategy consumes more training time compared to features extractor strategy.

As a conclusion of the obtained results, it was found that for each dataset we select a suitable model giving the highest test accuracy on the considered dataset. Some models take too long to train because they have a very high number of parameters like VGG and DenseNet models.

Some of the evaluated models in this paper did not make progress in the result attained in previous works. However, the achieved results may be a starting point for other works research focused on the use of the evaluated pre-trained models.

Table 6 displays the f1-score and the number of miss classified examples obtained by the MobileNet model for the testing set of CMATERDB 3.3.1 and Table 7 shows the f1-score and the number of incorrect examples obtained by the

EfficientNetB0 model for the testing set of OIHACDB-28 dataset.

TABLE VI F1-SCORE and NUMBER of INCORRECT EXAMPLES on TESTING DATA of CMATERDB 3.3.1 DATASET ACHIEVED by the MOBILENET MODEL

Digit	F1-score	Number of missed examples
0	0.99	1
1	0.99	1
2	0.98	2
3	0.98	2
4	0.98	2
5	1	0
6	0.99	1
7	1	0
8	1	0
9	0.99	1

TABLE VII F1-SCORE and NUMBER of INCORRECT EXAMPLES on TESTING DATA of OIHACDB-28 DATASET ACHIEVED by the EFFICIENTNETB0 MODEL

Character	F1-score	Number of missed examples
أ	1	0
ب	1	0
ت	0.98	2
ث	0.95	5
ج	1	0
ح	1	0
خ	1	0
د	0.94	6
ذ	0.99	1
ر	0.98	2
ز	1	0
س	1	0
ش	1	0
ص	0.99	1
ض	1	0
ط	1	0
ظ	1	0
ع	0.99	1
غ	1	0
ف	0.93	7
ق	0.90	9
ك	1	0
ل	0.97	3
م	1	0
ن	0.93	7
ه	0.98	2
و	0.98	2
ي	1	0

From Table 6 and Table 7, it is shown that some characters or digits have a high number of miss classified samples. For example, Arabic characters having a high number of wrong classification are: "ر", "د", "ن", "ق" and other characters. It can be clearly seen that the two letters "ز" and "ز" (also "د" and "ذ") are not confusable despite its most similar form. However, the two letters "د" and "ر" are very confusable in

the structure. In addition, the letter “ف” is confused with the letter “ق”. The difference between “ف” and “ق” is the number of dots that above the two characters. In general, the EfficientNetB0 model had the ability to distinguish between letters with the same main part and different numbers or positions of diacritic points. This is justified by the number of miss classified examples of the letters “ج”, “ح”, “خ”, “ع”, “غ”, “س”, “ط”, “ظ”, “ض”, “ص”, “ش”, “س”.

For Arabic digits, “٢”, “٣” and “٤” had the high number of incorrect classified samples. The causes of the classification errors are related to the strong shape deformation and the similar morphology of certain letters or digits in Arabic handwriting.

VII. CONCLUSIONS and FUTURE WORKS

This work aims to study the performance of transfer learning technique in the domain of Arabic handwritten text recognition.

In this work, the well-recognized pre-trained deep CNN models were compared for predicting Arabic handwritten text. These models are mainly used in two scenarios; Feature extractor and fine-tuning. To adapt the evaluated models to the tested datasets in this study, in both strategies the last layer is changed by a new fully connected layer to match the number of classes of the experienced datasets. For that, various datasets were used: numeral datasets (CMATERDB 3.3.1 and MADBase), word dataset (AHDB), and character datasets (OIHACDB and AIA9K). The obtained results demonstrate the performance of transfer learning method for Arabic handwritten text recognition. For each dataset, we selected a suitable model in terms of testing accuracy. Some models have a minimum number of parameters, which decreases the training time, and others take too long time to train. The models that gave higher accuracy can be selected to be used just as a features extractor combined by a traditional classifier like SVM. This idea can be tested in future works for the same or other datasets. In addition, future works can be done by evaluating other models or by the combination of several models. For some datasets (like AIA9K dataset), the training dataset is small for that data augmentation techniques can be applied to increase the amount of the data and reduce the overfitting.

REFERENCES

- [1] M. Silfverberg, CHAPTER 1 - Historical Overview of Consumer Text Entry Technologies Editor(s): I. Scott MacKenzie, Kumiko Tanaka-Ishii, Text Entry Systems, Morgan Kaufmann, 2007, pp. 3-25.
- [2] K.S. Dash, N.B. Puhan, and G. Panda. Unconstrained handwritten digit recognition using perceptual shape primitives. In *Pattern Analysis and Applications*; Springer: London, UK, 2016.
- [3] M. Ait Aider, K. Hammouche, and D. Gaceb, “Recognition of handwritten characters based on wavelet transform and SVM classifier”, *Int. Arab J. Inf. Technol*, vol. 15: 1082-1087, 2018.
- [4] N. Abdul Hamid, and N.N.A. Sjarif, “Handwritten Recognition Using SVM, KNN and Neural Network”, 2017, arXiv:1702.00723v1.
- [5] A. Baldominos, Y. Saez, and P. Isasi, “A survey of handwritten character recognition with mnist and emnist, *Applied Sciences*, vol. 9, 2019, pp.3169.
- [6] M. Ramzan, H.U. Khan, S.M. Awan, W. Akhtar, M. Ilyas, A. Mahmood, and A. Zamir, “A survey on using neural network based algorithms for handwritten digit recognition”, *International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 9, 2018, pp.519-528.
- [7] Handwritten Arabic Numeral Database. Google Coe Archieve—Long-Term Storage for Google Code Project Hosting. Available online: <https://code.google.com/archive/p/cmaterdb/downloads> (access ed on 9 November 2017).
- [8] A. Alani, “Arabic handwritten digit recognition based on restricted boltzmann machine and convolutional neural networks”, *Information*, vol. 8, 2017, pp.142
- [9] A. Ashiquzzaman, A.K. Tushar, A. Rahman, and F. Mohsin, “An efficient recognition method for handwritten Arabic numerals using CNN with data augmentation and dropout”, In: Balas VE, Sharma N, Chakrabarti A (eds) *Data management, analytics and innovation, advances in intelligent systems and computing*. Springer, Singapore, 2016, pp. 299–309.
- [10] K. Alia, S.M. Bashar, and A. Asmaa, “Arabic Handwriting Word Recognition Based on Scale Invariant Feature Transform and Support Vector Machine”, In *Iraqi Journal of Science*, 2019.
- [11] S. Al-Maadeed, D. Elliman, and C. Higgins, “A Dataset for Arabic Handwritten Text Recognition Research”. *Proceedings Eighth International Workshop on Frontiers in Handwriting Recognition*, 2002, pp. 485-489
- [12] C. Boufenar, M. Batouche, and M. Schoenauer, “An Artificial Immune System for Offline Isolated Handwritten Arabic Character Recognition”, *Evolving Systems* 9, 2016, pp.25–41.
- [13] C. Boufenar, A. Kerboua, and M. Batouche, “Investigation on deep learning for off-line handwritten Arabic character recognition”. *Cognitive Systems Research*, 2018, pp.180-195.
- [14] S. Faouci, D. Gaceb, and M. Haddad, “Handwritten Arabic Character Recognition: Comparison of Conventional Machine Learning and Deep Learning Approaches”, In: Saeed F., Mohammed F., Al-Nahari A. (eds) *Innovative Systems for Intelligent Health Informatics. IRICT 2020, Lecture Notes on Data Engineering and Communications Technologies*, vol. 72, 2020, Springer, Cham.
- [15] M. Torki, M.E. Hussein, A. Elsallamy, M. Fayyaz, M., and S. Yaser, ‘Window-Based Descriptors for Arabic Handwritten Alphabet Recognition: A Comparative Study on a Novel Dataset’, 2014 arXiv:1411.3519v2.
- [16] R. Ahmed, M. Gogate, A. Tahir, K. Dashtipour, B. Al-tamimi , A. Hawalah, M.A. El-Affendi, and A. Hussain, “Deep Neural Network-Based Contextual Recognition of Arabic Handwritten Scripts”, *Entropy*, vol. 23, 2021, Art. No.: 340. <https://doi.org/10.3390/e23030340>.
- [17] S. Abdleazeem and E. El-Sherif, “Arabic handwritten digit recognition” *International Journal of Document Analysis and Recognition (IJRAR)*, vol. 11, no. 3, pp. 127-141, 2008.
- [18] K. Simonyan, and Z. Andrew, “Very deep convolutional networks for large-scale image recognition”, 2014, arXiv preprint arXiv:1409.1556.
- [19] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions”, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [20] I. Sergey and S. Christian, “Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift”, *Proceedings of the 32nd International Conference on International Conference on Machine Learning*, vol. 37, 2015, pp. 448–456.
- [21] C.Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, Z. WojnaS, “Rethinking the Inception Architecture for Computer Vision”, in 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 2818-2826.
- [22] K. He, X. Zhang, S. Ren, and J.Sun, “Deep residual learning for image recognition”, 2015, arXiv preprint arXiv:1512.03385.
- [23] A.G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Weijun, W. Tobias, A. Marco, and A. Hartwig, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, 2017, arXiv:1704.04861v1 [cs.CV].
- [24] M. Tan and V.L. Quoc, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks”, 2019, ArXiv abs/1905.11946.
- [25] G. Huang, E. Liu, K.Q. Weinberger, and L. van de Maaten, “Densely connected convolutional networks”, arXiv preprint arXiv:1608.06993, 2016.

Meteosat Image Compression Based 2d Gradient Tensor

BENSAFI Noureddine, ATTAFYoucef, LAHDIR Mourad

Laboratoire LAMPA, Université UMMTO, Algeria

n.bensafi@hotmail.com

attaf2002@yahoo.fr

mlahdir@yahoo.fr

Abstract— In this paper, we present a new approach for Meteosat image compression based on the 2d gradient tensor. A quadtree decomposition is applied on the image to be compressed with a depth of 8*8. The blocks belonging to this depth are considered as heterogeneous and are quantized by the 2d tensor to which a dictionary is associated by learning. The blocks of higher dimensions are considered as homogeneous and are quantized using an averaging filter and coded with the RLE coding. Experimental results are obtained on several test images and a set of multispectral images from the MSG satellite.

Keywords— Image compression, Gradient, Tensor 2d, Gaussianfilter, Quadtree Quantification, RLE, MSG.

I. INTRODUCTION

The MSG (Meteosat Second Generation) satellite covers a large area of the globe with its geostationary orbit of 0 ° longitude and an altitude of 35 800 km from Europe, Africa and the western part of Asia. The instrument on board is the SEVIRI (Spinning Enhanced Visible and Infra-Red Imager) radiometer that provides 12 images of 12 different spectral channels every 15 minutes [1], [12]. Several studies have shown the extent of the possibilities offered by this instrument for meteorological and climatic monitoring. Also, the information it provides is necessary for the prevention of natural phenomena, such as the estimation of precipitation and the intensity of different cloud systems [5], [18], forest fire monitoring [8], detection of drought [15] and the study and monitoring of eruptive phenomena [13]. And many other areas that can benefit from the information provided by the SEVIRI radiometer.

In view of the increase in the number and size of multispectral images provided by the MSG satellite, studies have been carried out to reduce the size of the data to be transmitted or archived while preserving the quality of the reconstructed image essential for its application [20]. Compression methods are used which can be lossy or lossless, hybrid methods have been implemented exploiting the advantages offered by the two types of methods already cited [2], [9]–[11], [16], [17],[21], [23]. It is for this purpose that we offer this method of image compression by tensor 2d of the gradient.

The gradient of an image makes it possible to deduce the contours that have high amplitude values. The representation of the tensor structure 2d of the gradient is using for the local displacement of the contour locations to be minimized by the application of a Gaussian filter, element by element, to the tensor structure 2d [19]. In addition, cancellations of opposite polarity of gradient directions are avoided when the tensor structures 2d are added [7]. This form of gradient representation is particularly useful for deducing noise

[22]. Also, the diffusion process can be used to infer local contour patterns and surfaces and local structural estimates that serve applications, such as obstruction and occlusion detection in image sequences and biometric identification of fingerprints [3], [4]. The eigenvalues deduced from the 2d tensor structure are efficient for measuring and detecting angles and locating salient points.

In this work, we propose a lossy compression method based on the tensor 2d of the gradient. The main contribution of this paper is to provide a very efficient compression technique for multispectral images of MSG. We exploit the parameter of tensor 2d of the gradient of each image for quantification of heterogeneous blocks results of quadtree decomposition and a RLE encoder.

This paper is organised as follows. The tensor structure 2D of the gradient are detailed in section II. The compression algorithm based on the 2D gradient tensor is proposed in section III. Section IV shows the experimental results obtained from Lena and MSG images. Finally, we draw some conclusions in section V.

II. THE 2D GRADIENT TENSOR STRUCTURE

A. The Gradient of Image

The gradient of image $I(x, y)$ is the partial derivative of the gray level function, in the form of a continuous signal is:

$$\vec{\nabla}I(x, y) = \begin{pmatrix} \nabla_x I \\ \nabla_y I \end{pmatrix} = \begin{pmatrix} \frac{\partial I(x, y)}{\partial x} \\ \frac{\partial I(x, y)}{\partial y} \end{pmatrix} \quad \#(1)$$

Since the image function $I(x, y)$ is defined in a two-dimensional space, we can define partial derivatives with respect to the definition variables of I .

$$\frac{\partial I(x, y)}{\partial x} = \lim_{h_x \rightarrow 0} \frac{I(x + h_x, y) - I(x, y)}{h_x} \quad \#(2)$$

$$\frac{\partial I(x, y)}{\partial y} = \lim_{h_y \rightarrow 0} \frac{I(x, y + h_y) - I(x, y)}{h_y} \quad \#(3)$$

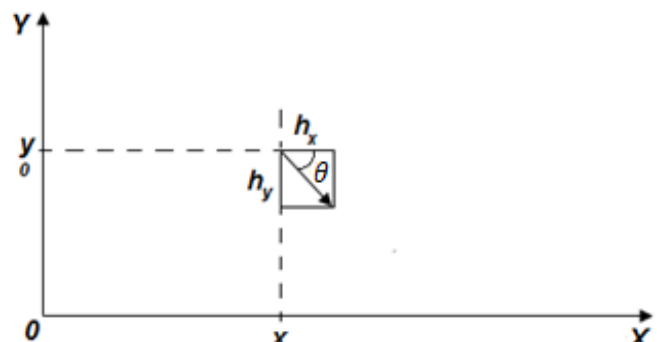


Fig.1. Gradient components.

The gradient characterized by its modulus and direction as follows:

$$|\vec{\nabla}I|(x, y) = \sqrt{\left(\frac{\partial I(x, y)}{\partial x}\right)^2 + \left(\frac{\partial I(x, y)}{\partial y}\right)^2} \quad \#(4)$$

$$\theta = \text{Arg}(\vec{\nabla}I(x, y)) = \tan^{-1}\left(\frac{\nabla_y I}{\nabla_x I}\right) \quad \#(5)$$

The variation is done with a step of a pixel, and then its gradient is:

$$\vec{\nabla}I(x, y) = \begin{pmatrix} \nabla_x I \\ \nabla_y I \end{pmatrix} = \begin{pmatrix} I(x+1, y) - I(x, y) \\ I(x, y+1) - I(x, y) \end{pmatrix} \quad \#(6)$$

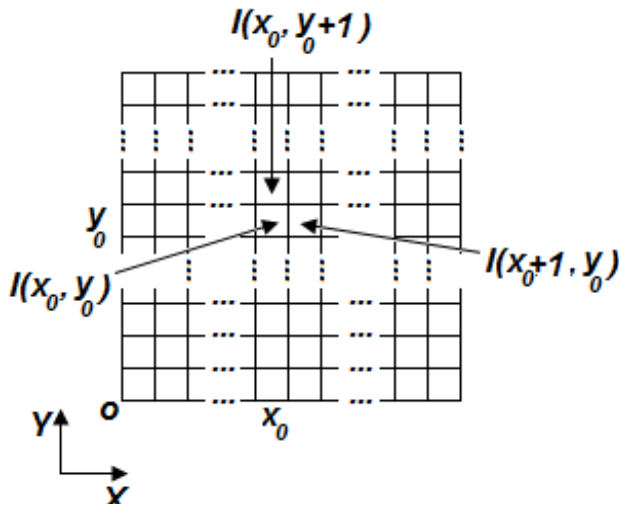


Fig.2. Gradient components of a pixel

B. Relation Between Contour and Gradient

A contour is a strong local variation of the gray level; the gradient vector is perpendicular to the contour. Since the directional derivative is zero in the direction perpendicular to the contour, the gray level variation is zero along the contour.

Among the information representing a gradient of an image, we will choose the directional derivative, which provides a vector representation. Its amplitude reflects the maximum variation of pixel values as the phase is directed along the orientation corresponding to the maximum variation.



Fig.3Lena's image, (b)

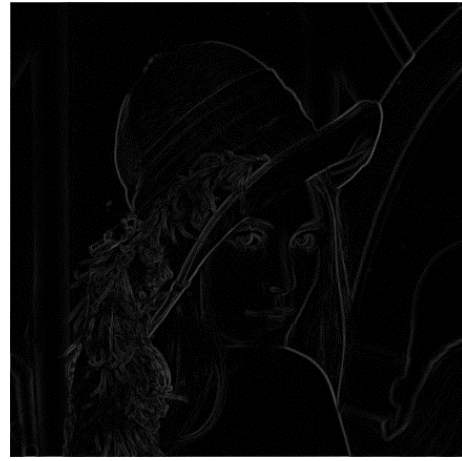


Fig.4Gradient of Lena image

C. Application of the Gradient

Since the intensity of a digital image is discrete, the derivatives of this function cannot be defined except under a continuity assumption of continuous intensity function that sampled. In practice, one can calculate more or less faithful approximations of the gradient at each point.

The gradient is a derivative filter, it is a high pass filter, because $\nabla_x I(x, y)$ gives us the details along the ox axis, and $\nabla_y I(x, y)$ along the axis oy , and $|\vec{\nabla}I|$ Gives us the details of the whole image. The gradient is sensitive to noise, it is necessary to apply beforehand a low pass filter, such as the Gaussian filter to eliminate the noise that can exist in the image.

$$G_\sigma(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad \#(7)$$

With

$G_\sigma(x, y)$: The Gaussian filter.

σ : The standard deviation.

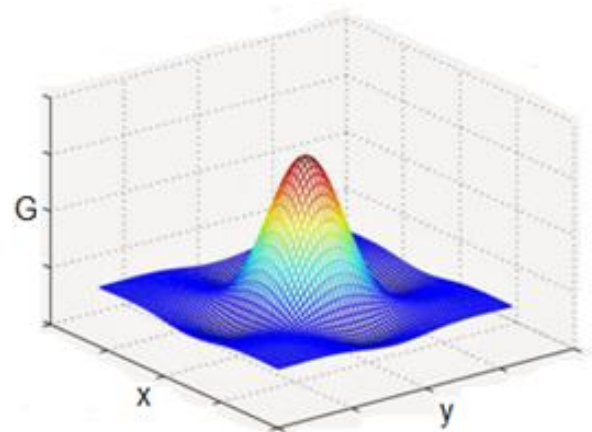


Fig.5.Graphical representation of a Gaussian filters.

The result obtained after applying the Gaussian filter is:

$$I^*(x, y) = I(x, y) * G_\sigma(x, y)$$

The gradient of the filtered image is: $\vec{\nabla}I^*(x, y) = \begin{pmatrix} \nabla_x I^* \\ \nabla_y I^* \end{pmatrix}$

Another approach is to compute the gradient of the Gaussian filter and then make the convolution product with the image processed. Such as:

$$\vec{\nabla} G_\sigma(x, y) = \begin{pmatrix} \nabla_x G_\sigma \\ \nabla_y G_\sigma \end{pmatrix} = \begin{pmatrix} G_\sigma(x) \\ G_\sigma(y) \end{pmatrix} \#(8)$$

With

$$I_x = I(x, y) * G_\sigma(x)$$

$$I_y = I(x, y) * G_\sigma(y)$$

$$\vec{\nabla} I(x, y) = \begin{pmatrix} I_x \\ I_y \end{pmatrix}$$

$$|I^*|(x, y) = \sqrt{I_x^2 + I_y^2}$$

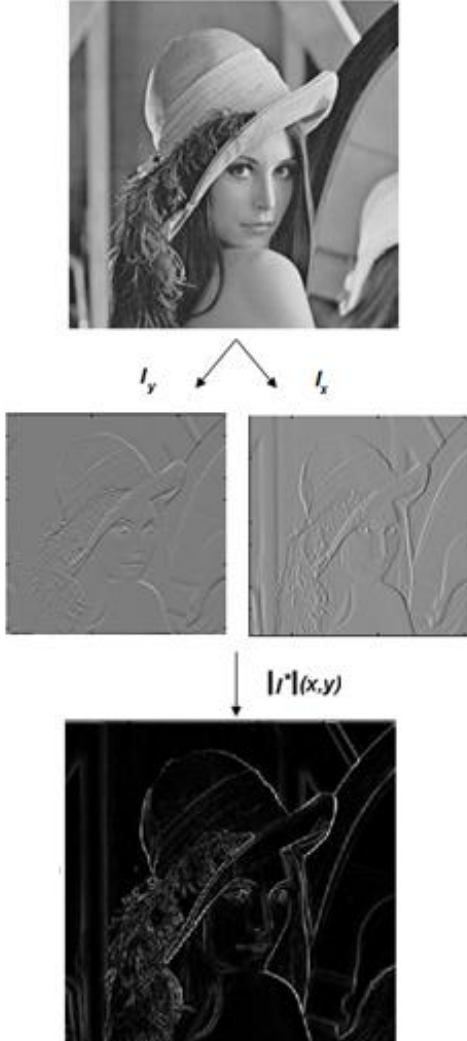


Fig.6.Gradient of the Lena image with $\sigma = 1$.

The standard deviation of the Gaussian filter σ , allows us to have more information in a proportional way.

D. The 2d Tensor Structure

Another method of representing the gradient information is the use of the tensor structure 2d, it also allows the use of the values I_x and I_y . However, in a matrix form:

$$T = \begin{bmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{bmatrix} = \begin{bmatrix} I_x^2 & I_x I_y \\ I_y I_x & I_y^2 \end{bmatrix} \#(9)$$

A proper decomposition is performed at the tensor T to form the eigenvalues (λ_1, λ_2) and the own unit vectors (\vec{e}_1, \vec{e}_2) . These new features of the gradient allow a more precise description of the local gradient characteristics. As \vec{e}_1 is a unit vector directed perpendicular to the contour while the vector \vec{e}_2 is tangent to the contour.

$$\vec{\lambda}_1 = \lambda_1 * \vec{e}_1 .$$

$$\vec{\lambda}_2 = \lambda_2 * \vec{e}_2 .$$

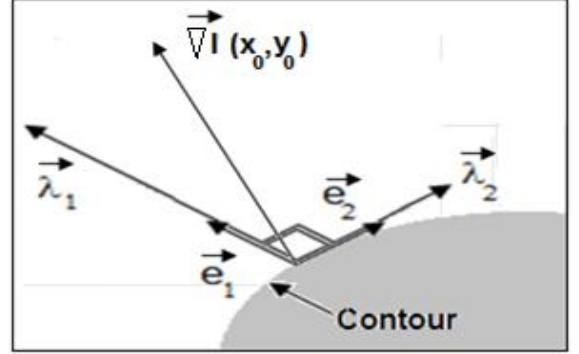


Fig.7. Decomposition of the gradient into eigenvectors.

Eigenvalues indicate the underlying certainty of the gradient structure along their associated eigenvector directions.

Consistency obtained as a function of eigenvalues.

$$C = \begin{cases} \frac{(\lambda_1 - \lambda_2)}{(\lambda_1 + \lambda_2)} & \text{if } (\lambda_1 + \lambda_2) > 0 . \\ 0 & \text{else .} \end{cases}$$

With,

C : The coherence of the tensor T .

$\vec{\lambda}_1$: The eigenvector associated \vec{e}_1 ,

$\vec{\lambda}_2$: The eigenvector associated \vec{e}_2 .

III. PROPOSED METHOD

Our compression method is composed of three parts: block partitioning and classification, quantification and coding. Fig. 8 shows the different steps of the proposed algorithm.

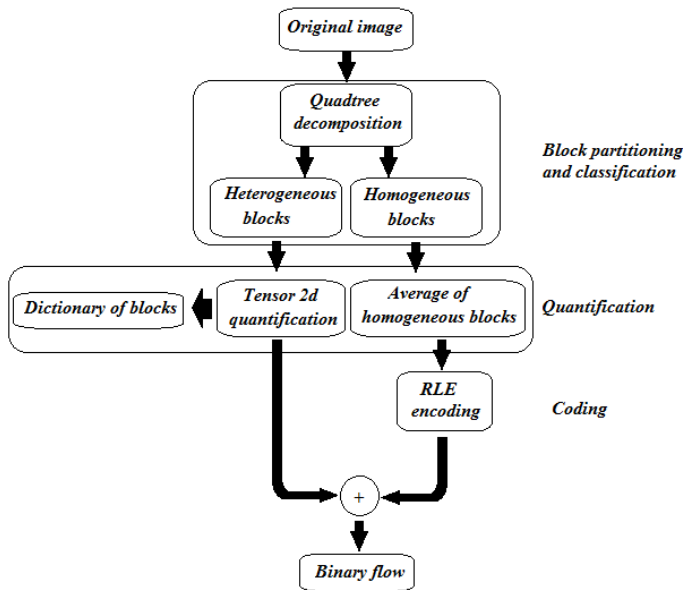


Fig.8. Block diagram of the compression method.

1) Block Partitioning and Classification

The first step of our method use a quadtree decomposition that involves subdividing an image into blocks according to a homogeneity threshold and a depth of decomposition, [6]. The blocks belonging to this depth are heterogeneous, and the blocks of higher dimensions 8*8 are homogeneous [14].



Fig.9. Homogeneous blocks

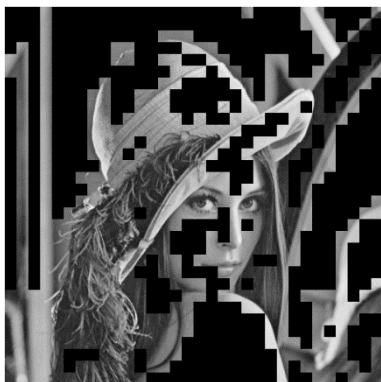


Fig.10 Heterogeneous blocks

2) Quantification and Coding of Homogeneous Blocks

The homogeneous blocks quantized using an averaging filter, which allows us to have a block of a set of pixels of the same mean value. This allows applying RLE (Run Length Encoding) for encoding the sequences of same pixel value with line-by-line scanning. The number of pixels of the block is no longer preserved because it exists in the quaternary structure.

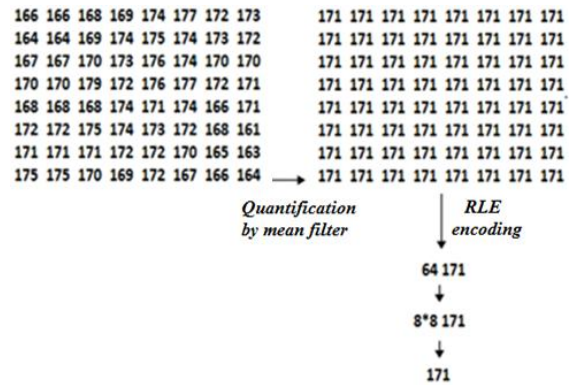


Fig.11. Quantification and coding of block.

3) Quantification of heterogeneous blocks

Let us know that the heterogeneous blocks contain the contours, i.e. the relevant information then we observe a variation of the values of the gradient from a positive value to a negative value, this variation that we will exploit to characterize these blocks. A quantization is carried out while keeping S_x and S_y coded on 8 bits, respectively, are the sign of the directional derivatives I_x and I_y of the eight significant pixels placed strategically according to a mask, such as:

$$S_{xk} = \begin{cases} 0 & \text{if } I_{xk} \geq 0 \\ 1 & \text{else} \end{cases} .$$

$$S_{yk} = \begin{cases} 0 & \text{if } I_{yk} \geq 0 \\ 1 & \text{else} \end{cases} .$$

With: $k = 1 \dots 8$.

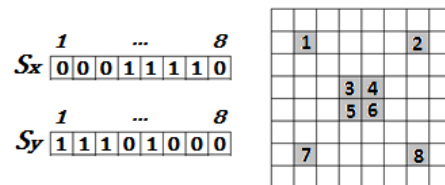


Fig.12. Sign of the directional derivatives of a block.

In the case where we have two heterogeneous blocks with the same contours but with different gray levels, we will have a variation of the same gradient, to make the difference between these two blocks we associate with S_x and S_y the two values NG_{min} and NG_{max} encoded on 8 bits respectively the minimum and maximum gray level and their position in the block P_{min} et P_{max} coded on 8 bits, such as:
The position of NG_{min} is $P_{min} = 01100111$.
The position of NG_{max} is $P_{max} = 00010000$.

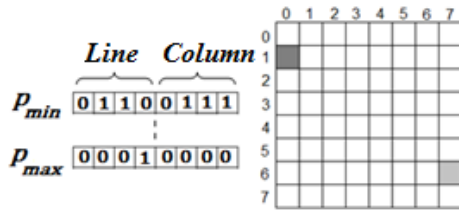


Fig.13 Minimum and maximum values and their position in the block.

4) *Block Dictionary*

A dictionary of blocks by learning is realized using the 6 values $(S_x, S_y, NG_{min}, NG_{max}, P_{min}$ and $P_{max})$ which correspond to a heterogeneous block, if this block does not exist in this dictionary, the block and the 6 values characterizing it are saved.

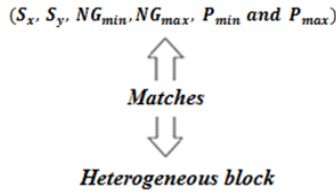


Fig.14. Correspondences of heterogeneous blocks

The binary flow is composed of the quaternary structure, the mean values of the homogeneous blocks and the six values characterizing the heterogeneous blocks.

The reconstruction of the image applied out by means of an RLE decoding for the homogeneous blocks and of an inverse quantification for the heterogeneous blocks using the dictionary of the blocks.

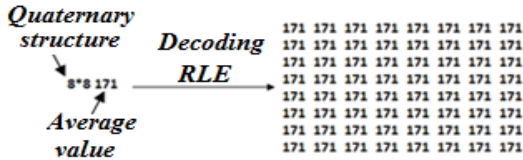


Fig.15. Decoding of a homogeneous block.

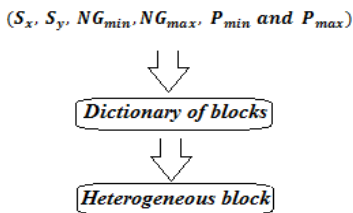


Fig.16. Inverse quantification of a heterogeneous block.

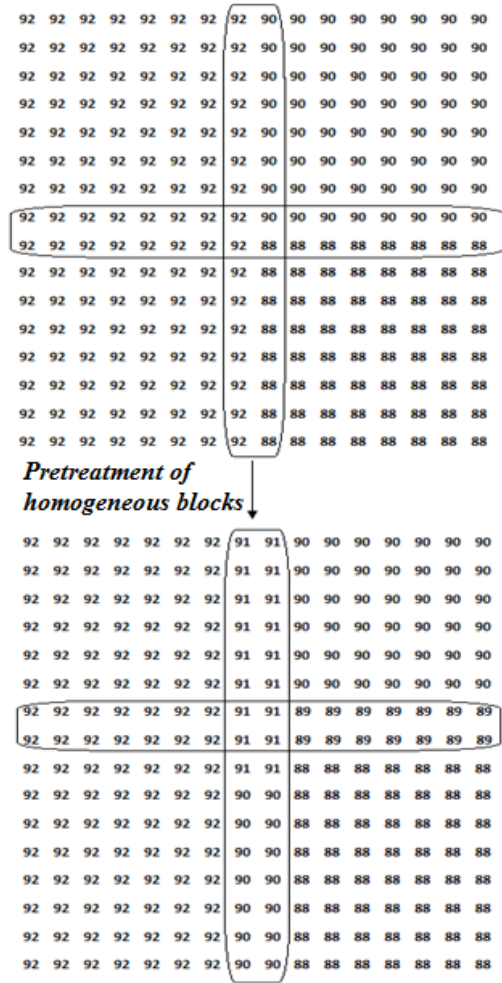


Fig.17. Pre-treatment of homogeneous blocks.

A pre-processing step applied out to the homogeneous blocks to eliminate the "paved" effect during the reconstruction, which consists in attributing to the edges of these blocks the mean value of the two adjacent blocks.

IV. RESULTS AND DISCUSSION

In order to evaluate the performance of our compression algorithm, we used as a reference the Lena image and a set of multispectral images of the MSG satellite. The MSG images provided by the meteorological station of the National Office of Meteorology (ONM) of Dar El Beida, Algeria recorded on 23.11.2011 at 11h45. The resolution of MSG images is 512x512 with eight bpp. As is shown in Fig. 25 and 27, our choice is fixed on the following MSG images: Infrared 9.7 (IR 9.7) and infrared 3.9 (IR 3.9). For a better appreciation of the results obtained during the application of our algorithm, the tests applied out by varying the homogeneity threshold of 0.1 to 0.4, which intervenes in the quadtree decomposition of the proposed method.



Fig.18 Reconstituted Lena image with a depth of 8 * 8 and a homogeneity threshold of 0.1.



Fig.19 Reconstituted Lena image with a depth of 8 * 8 and a homogeneity threshold of 0.2.



Fig.20 Reconstituted Lena image with a depth of 8 * 8 and a homogeneity threshold of 0.3.

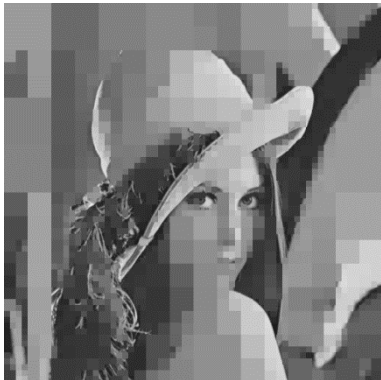


Fig.21 Reconstituted Lena image with a depth of 8 * 8 and a homogeneity threshold of 0.4.

For a threshold of homogeneity equal to or less than 0.2, the images returned are of good quality, beyond this value, the image loses its pertinent information. This assessment remains subjective. To quantify this evaluation, we calculated the performance criteria Mean Squared Error MSE , the Peak Signal to Noise Ratio $PSNR$ and the compression ratio CR .

$$MSE = \frac{1}{m * n} \sum_{x=1}^m \sum_{y=1}^n (I_o(x, y) - I_r(x, y))^2 \quad \#(10)$$

With

I_o : Original image.

I_r : Restored image.

$$PSNR = 10 * \log_{10} \left(\frac{d^2}{MSE} \right) \quad \#(11)$$

With

d : Maximum value of a pixel in image.

$$CR = \left(1 - \frac{Uncompressed\ size}{compressed\ size} \right) * 100 \quad \#(12)$$

Table 1 shows the results obtained by application on the Lena image for different homogeneity threshold.

Fig. 22 shows when the threshold of homogeneity increases, the compression ratio increases and the PSNR decreases.

TABLE 1.

PERFORMANCE CRITERIA FOR LENA IMAGE COMPRESSION.

The homogeneity threshold	MSE	PSNR (dB)	CR (%)
0.1	3.00	86.03	76.35
0.2	20.20	69.46	82.26
0.3	59.44	60.08	86.01
0.4	171.78	50.87	89.53

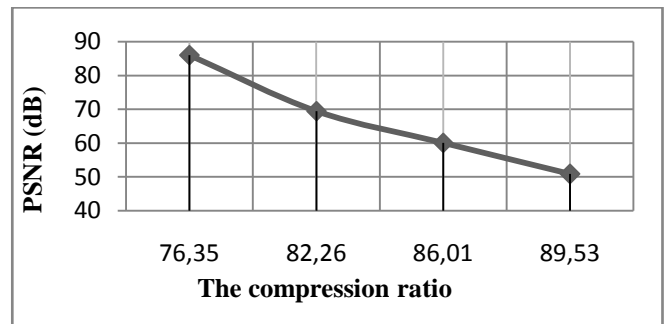


Fig.22. PSNR against different compression ratio values of the Lena image.

In the light of the subjective results and the performance evaluation criteria obtained, we chose a homogeneity threshold of 0.2, which allowed us to obtain a compression ratio of 82.26% and a PSNR of 69.46 dB.



Fig.23Original image Lena 512 * 512 pixels..



Fig.24Restored image CR =82.26% and PSNR = 69.46dB.

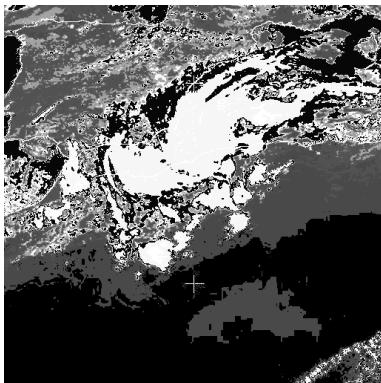


Fig.25Original infrared image IR 3.9.

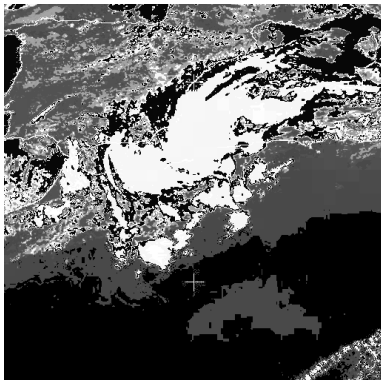


Fig.26Restored image CR =78.65 % and PSNR = 102.13 dB.

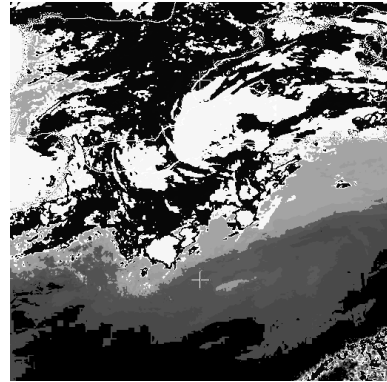


Fig.27Original Meteosat infrared image IR 9.7.

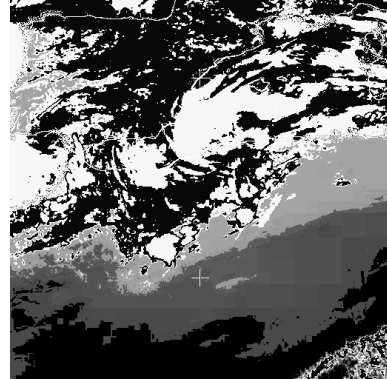


Fig.28 Restored image for CR =77.16% and PSNR = 92.19dB.

Fig. 26 and 28 give the reconstruction of the MSG images using the proposed method. We notice from these Figs that the image reconstructed by the proposed method has a best visual quality. Our method provides better preservation of the important details like the clouds and the edges of different regions.

V. CONCLUSIONS

Our approach is in line with the same requirement as hybrid methods, namely the reduction of the amount of data representing the image to be coded, keeping a visual aspect of that reconstructed. Our method based on quadtree decomposition and an adequate quantification of the homogeneous and heterogeneous blocks resulting from this decomposition.

The quadtree decomposition allowed us to define the homogeneous and heterogeneous blocks of an image according to a threshold of homogeneity and a depth. A quaternary structure generated during the decomposition that occupies a size of 60% of the data constituting the compressed image. The homogeneous blocks were quantized using an averaging filter and RLE encoding which allowed us a compression rate of 98% per block. The quantification of the heterogeneous blocks by 2d tensor enabled us to improve the contrast of the image by applying the Gaussian filter when calculating the gradient of the pixels of the heterogeneous blocks and a compression ratio of 90% by blocks.

The results obtained from our approach are very encouraging because the images reproduced after a

compression and decompression processing a good quality with a compression ratio (80 %) and a high PSNR (100dB) for hyperspectral images.

In order to improve our results, we give the following perspectives:

- Using another decomposition model that does not generate such a type of data as the quaternary structure.
- A classification of the blocks in the level of the depth, to reduce the number of blocks quantified by tensor 2d.
- Increase the depth of the quadtree decomposition by $16*16$.
- Improve the characterization of heterogeneous blocks by using other characteristics (increase the number of significant pixels ... etc) in order to avoid the phenomenon of divergence between the blocks during the reconstruction.

REFERENCES

- [1] D. Aminou, « MSG's SEVIRI instrument », *ESA Bull*, n° 111 (July 2002): 15-17.
- [2] L. Akrou, S. Ameer, M. Lahdir, R. Fournier, and A. N. Ali, « Fast Hyperspectral Image Encoder Based on Supervised Multimodal Scheme », *International Journal of Image and Graphics*, vol. 21, n° 01, Art. n° 01, janv. 2021, doi: 10.1142/S0219467821500078.
- [3] S. Arseneau and J. R. Cooperstock, « An Asymmetrical Diffusion Framework for Junction Analysis », in *Proceedings of the British Machine Vision Conference 2006*, Edinburgh, 2006, p. 71.1-71.10. doi: 10.5244/C.20.71
- [4] S. Arseneau and J. R. Cooperstock, « An Improved Representation of Junctions Through Asymmetric Tensor Diffusion », in *Advances in Visual Computing*, vol. 4291, G. Bebis, R. Boyle, B. Parvin, D. Koracin, P. Remagnino, A. Nefian, G. Meenakshisundaram, V. Pascucci, J. Zara, J. Molineros, H. Theisel, and T. Malzbender, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, p. 363- 372. doi: 10.1007/11919476_37.
- [5] N. Bensafi, M. Lazri, and S. Ameer, « Novel WkNN-based technique to improve instantaneous rainfall estimation over the north of Algeria using the multispectral MSG SEVIRI imagery », *Journal of Atmospheric and Solar-Terrestrial Physics*, vol. 183, p. 110- 119, févr. 2019, doi: 10.1016/j.jastp.2018.12.004.
- [6] I. Blanes, E. Magli, and J. Serra-Sagrasta, « A Tutorial on Image Compression for Optical Space Imaging Systems », *IEEE Geosci. Remote Sens. Mag.*, vol. 2, n° 3, Art. n° 3, sept. 2014, doi: 10.1109/MGRS.2014.2352465.
- [7] T. Brox, A. Bruhn, N. Papenber, and J. Weickert, « High Accuracy Optical Flow Estimation Based on a Theory for Warping », in *Computer Vision - ECCV 2004*, vol. 3024, T. Pajdla et J. Matas, Éd. Berlin, Heidelberg: Springer Berlin Heidelberg, 2004, p. 25- 36. doi: 10.1007/978-3-540-24673-2_3.
- [8] L. C. Carvalheiro, S. O. Bernardo, M. D. M. Orgaz, and Y. Yamazaki, « Forest Fires Mapping and Monitoring of current and past forest fire activity from Meteosat Second Generation Data », *Environmental Modelling & Software*, vol. 25, n° 12, Art. n° 12, déc. 2010, doi: 10.1016/j.envsoft.2010.06.003.
- [9] M. Cherifi, M. Lahdir, and S. Ameer, « Sorted run length coding-application to Meteosat image compression », in *Environmental Engineering and Computer Application*, K. Chan, Éd. CRC Press, 2015, p. 159- 163. doi: 10.1201/b18565-33.
- [10] M. Cherifi, M. Lahdir, and S. Ameer, « A Novel Meteosat Second Generation Image Compression Method Based on Radon Transform, Linear Predictive Coding with Filtering and Sorted Run Length Coding », *IRECOS*, vol. 10, n° 4, Art. n° 4, avr. 2015, doi: 10.15866/irecos.v10i4.5704.
- [11] M. Cherifi, M. Lahdir, and S. Ameer, « Meteosat image sequence coding in the Radon field », *Optik*, vol. 182, p. 1228- 1243, avr. 2019, doi: 10.1016/j.ijleo.2019.02.015.
- [12] EUMETSAT, monitoring weather and climate from space, <http://www.eumetsat.int>, 2021.
- [13] B. R. Hirn, C. Di Bartola, G. Laneve, E. Cadau, and F. Ferrucci, « SEVIRI Onboard Meteosat Second Generation, and the Quantitative Monitoring of Effusive Volcanoes in Europe and Africa », in *IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium*, Boston, MA, USA, 2008, p. III-374-III- 377. doi: 10.1109/IGARSS.2008.4779361.
- [14] F. Keissarian, « A new quad-tree segmented image compression scheme using histogram analysis and pattern matching », in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, Singapore, févr. 2010, p. 694- 698. doi: 10.1109/ICCAE.2010.5451332.
- [15] B. Lacaze, and J. Berges, « Contribution of Meteosat Second Generation (MSG) to drought early warning. *Remote Sensing and Geoinformation Processing in the Assessment and Monitoring of Land Degradation and Desertification: State of the Art and Operational Perspectives*, p. 2005,406-412.
- [16] M. Lahdir, S. Ameer and A. Adane, « Algorithme non itératif, basé sur les ondelettes biorthogonales et les fractales, pour la compression d'images satellitaires », *Télé-détection*, vol 6 apr. 2006, p. 345-360.
- [17] M. Lahdir, A. Nait-ali, and S. Ameer, « Fast Encoding-Decoding of 3D Hyperspectral Images Using a Non-Supervised Multimodal Compression Scheme », *JSIP*, vol. 02, n° 04, Art. n° 04, 2011, doi: 10.4236/jsip.2011.24045.
- [18] M. Lazri and S. Ameer, « Using cloud water path and cloud top temperature for estimating convective and stratiform rainfall from SEVIRI daytime data », *Arab J Geosci*, vol. 9, n° 11, Art. n° 11, août 2016, doi: 10.1007/s12517-016-2610-8.
- [19] M. Nicolescu and G. Medioni, « Motion segmentation with accurate boundaries - a tensor voting approach », in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, Madison, WI, USA, 2003, p. I-382-I- 389. doi: 10.1109/CVPR.2003.1211379.
- [20] M. He, L. Bai, Y. Dai, and J. Zhang, « Hyperspectral image lossless compression algorithm based on adaptive band regrouping », San Diego, CA, août 2009, p. 745504-745504- 6. doi: 10.1117/12.827450.
- [21] K. Sahnoun and N. Benabadjji, « Satellite Image Compression Algorithm Based on the FFT », *IJMA*, vol. 6, n° 1, Art. n° 1, févr. 2014, doi: 10.5121/ijma.2014.6106.
- [22] C.-K. Tang, M.-S. Lee, and G. Medioni, « Tensor Voting », in *Perceptual Organization for Artificial Vision Systems*, vol. 546, K. L. Boyer et S. Sarkar, Éd. Boston, MA: Springer US, 2000, p. 215- 237. doi: 10.1007/978-1-4615-4413-5_12.
- [23] N. Zikiou, M. Lahdir, and David Helbert. « Support Vector Regression-Based 3D-Wavelet Texture Learning for Hyperspectral Image Compression ». *The Visual Computer*, vol. 36, no 7 juillet 2020, 1473- 90. doi:10.1007/s00371-019-01753-z.

Wavelet learning-based super-resolution technique for hyperspectral image

Mohand Ouahioune[#], Leila Akrou[#], Boujemaa Ayouni[#], Mourad Lahdir[#], Soltane Ameer[#]

[#]Laboratory of Analysis and Modeling of Random Phenomena (LAMPA),
Mouloud Mammeri University (UMMTO)

¹ouahioune.m@gmail.com

²leilaakrou@yahoo.fr

³ayounibenjy@gmail.com

⁴mlahdir@yahoo.fr

⁵ameursoltane@yahoo.com

Abstract—To obtain images with higher resolution, the first idea is to upgrade the imaging system. However, this kind of approach has many limitations. In recent years, software processing-based image resolution enhancement known as super-resolution (SR) constituted an excellent alternative and a very attractive research topic. The SR techniques allow obtaining high-resolution (HR) images from low-resolution (LR) observations and, therefore, overcome the limits of imaging systems. It has found applications in many problems in different fields, such as remote sensing where HR images are urgently needed. In this paper, we propose a wavelet learning-based super-resolution algorithm for hyperspectral images (HSI). The 3D discrete wavelet transform (3D DWT) provides us with the approximate coefficients (AC) as well as the detail coefficients (DC) of the hyperspectral data cube. Then, we learn the mapping between the AC and the corresponding DC using a deep neural network (DNN). Finally, by assuming the LR image as the approximation coefficient of the unknown HR image, we reconstruct the desired HR image by predicting the unknown detail coefficients. The advantage of the technique is that we exploit all the spatio-spectral information of the HSI to reconstruct the HR image. The performance of the proposed technique has been evaluated and compared with some existing algorithms. Experimental results show that the proposed wavelet learning-based SR algorithm provides a promising performance and can generate high-quality images.

Keywords— *Hyperspectral image, Super-Resolution, Deep Neural Network, Multi-scale Structure, Wavelet Transform*

I. INTRODUCTION

Hyperspectral image is an attractive and efficient technology in remote sensing due to its ability to combine the power of digital imagery and the power of spectroscopy [1]. Hyperspectral image sensors collect the radiance for a large number of contiguous spectral bands of wavelengths ranging from 400 to 2500 nm. The different bands form a data cube that covers the visible, near-infrared, and mid-infrared regions. Depending on the capability of sensors, the data cube has a fixed spectral resolution and the same number of pixels in each band. Each pixel has a spatial resolution that defines the size of the area covered by the pixel. The constructed spectrum for every pixel is called a spectral signature and can be used for the identification of the land cover type and features [2]. HSI has found applications in many problems in the civilian and military sectors. All these applications need to achieve the greatest spatial and spectral resolution possible.

However, HSI has some limitations and many factors in the imaging system, such as atmospheric scattering, imperfect optics, and sensor noise degrade the spatial quality of these images. As a result, spatial resolution is one of the most difficult features to enhance in imaging systems [3, 4].

Researchers have investigated the use of super-resolution techniques as an alternative to obtain images with resolution beyond the limit of imaging systems and thus improve the spatial resolution of hyperspectral images. SR techniques are a highly active field of research, and many approaches, such as SR from single image techniques, multi-image-based SR techniques, and learning-based SR techniques, have been applied to the super-resolution challenge [5, 6]. However, many of these techniques disregard the spectral information of the HSI, which is a critical component for efficient HR images reconstruction. In this paper, we worked on the development of a novel super-resolution technique where all the spatio-spectral information of the HSI is exploited. Therefore, we developed a wavelet learning-based SR technique, and we named this novel SR algorithm the WSRnet. The 3D discrete wavelet transform (3D DWT) decomposes the image into approximation coefficients (AC) and detail coefficients (DC) in different scales [7-9]. We designed a deep neural network (DNN) [10-12] where the wavelet coefficients across all predefined scales are forming the training datasets and are used to learn the inter-scale dependency between the AC and the corresponding DC. Then, we obtain the SR image by assuming the LR image as AC and by predicting the corresponding DC using the mapping that was determined across the training datasets. The overall objective of this paper can be summarized as:

- Clarifying the SR techniques.
- Development of novel SR technique for all the spatio-spectral information of the HSI.
- Future challenges in terms of HSI super-resolution techniques.

This paper is organized as follows. Section II details the proposed architecture. The results and discussion with concluding remarks are given in Section III and Section IV.

II. PROPOSED ALGORITHM

A. Principle

In this section, we will describe the proposed super-resolution technique for the HSI that we have named WSRnet (Fig. 1). The algorithm consists of performing the following steps (Table I).

TABLE I. WAVELET LEARNING-BASED SR ALGORITHM

Algorithm	Wavelet learning-based super-resolution network (WSRnet)
3D DWT	<ul style="list-style-type: none"> Input: Low-resolution hyperspectral image (LR-HSI) $f(x,y,z)$. Output: LR-HSI $f(x,y,z)$ is converted into 3D DWT AC and DC at scale $j=3$ as follows: $AC_j[x,y,z] = \langle f, \Phi_{j,x,y,z} \rangle$ $DC_j^i[x,y,z] = \langle f, \Psi_{j,x,y,z}^i \rangle$
DNN	<p><i>Step 1:</i> Learning the inter-scale dependency.</p> <ul style="list-style-type: none"> Input: $AC_j[x,y,z]$ Output: $DC_j^i[x,y,z]$ <p><i>Step 2:</i> Predicting.</p> <ul style="list-style-type: none"> Input: Assuming LR-HSI $f(x,y,z)$ as 1-level AC of the HR-HSI. Output: Predicting the unknown 1-level DC of the HR-HSI.
Inverse 3D DWT	<ul style="list-style-type: none"> Input: LR-HSI $f(x,y,z)$ as 1-level AC and predicted 1-level DC. Output: The final high-resolution hyperspectral HR-HSI.

We implement the 3D DWT, which is an efficient multi-resolution image decomposition technique to exploit the inter-image correlations of the hyperspectral data cube [7, 8]. The DWT can be formulated as a filtering operation. The filters transform a discrete-time signal into an approximation coefficients (the large-scale features of the signal) and detail coefficients (the texture details and structure information of the signal). Multi-resolution analysis feeds the approximation into another set of filters, which produces the approximation and detail coefficients at the next level. We compute the 3D DWT by extending this 1D multi-level decomposition to the axes x , y , and z [9]. For a hyperspectral data cube $f(x,y,z)$, the approximation and detail sub-bands at scale $j \in \mathbb{Z}$ ($j > 0$) are defined as follows:

$$AC_j[x,y,z] = \langle f, \Phi_{j,x,y,z} \rangle \quad (1)$$

$$DC_j^i[x,y,z] = \langle f, \Psi_{j,x,y,z}^i \rangle \quad (2)$$

Where wavelets are generated from translation and dilations of a 3D scaling function $\Phi_{x,y,z}$ and 3D wavelet functions $\Psi_{j,x,y,z}^i$ (for $i = 1, 2, \dots, 7$).

Therefore, the principle of the wavelet transform is to hierarchically decompose the image into a series of consecutively sub-bands. At each level, the approximation and detail sub-bands have the necessary information to reconstruct the approximation sub-band of the next higher resolution level (Fig 2). The SR can be viewed as the problem of restoring the DC of the image given an AC input. Therefore, we assume that the LR image is the approximation sub-band of the unknown HR image. To up-sample and obtain the HR image, we design a deep neural network to predict the missing detail sub-bands [13]. All levels of decomposition form the training datasets to take advantage of the spatio-spectral correlation between the same and the separate sub-bands. Each set consists of one input that contains the approximation sub-band and seven outputs that contain the detail sub-bands. We trained the DNN for finding the complex nonlinear mapping between these sub-bands. Then, we predict the missing detail sub-bands of the LR image. Finally, the inverse

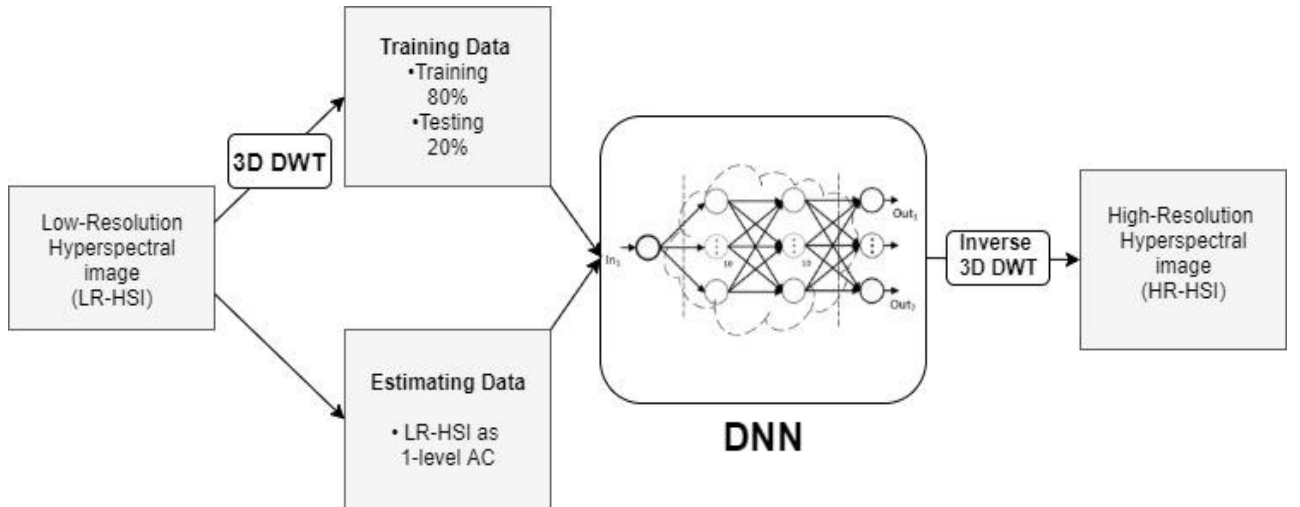


Fig.1. Illustration of the wavelet learning-based super-resolution network

3D DWT is applied to the predicted details and generates the SR results.

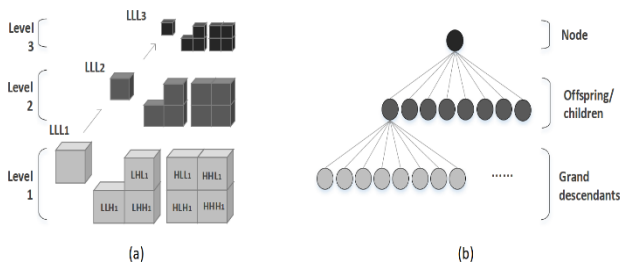


Fig.2. (a) 3D wavelet decomposition, (b) Inter-scale relationship

B. Wavelet Learning-Based SR network (WSRnet)

As we have seen, the learning-based SR method in the wavelet domain reconstructs the desired HR image by treating the LR image as an AC, then estimating the unknown DC [13, 14]. To predict the unknown DC, we develop a learning function based on a deep neural network [15-18]. A deep neural network is a highly connected network of elementary functions (neuron) operating in parallel. It is organized in layers where signals travel from the input to the output layer. In the learning process, the neuron’s weights are adjusted, which increases or decreases the strength of the signals. The trained network is then used in new situations to provide projections. Developing a neural network begins with the selection and preparation of the data samples. This step is crucial because how the samples are presented determines the type of network and the number of input and output cells. It also determines how the learning process will take place. In our case, blocks of AC and DC are obtained for the entire image and are used as training sets. Therefore, for each training input AC, the corresponding DC are identified and form targets for the learning system. This approach allows us to select the datasets that best correspond to the ones that will be estimated. The most common structure used for prediction and nonlinear function fitting is the feed-forward networks. The multi-layer perceptron (MLP) is the simplest kind of feed-forward networks, and it at least consists of three layers (input, hidden, and output layer). Several methods exist to choose the number of neurons in the hidden layer. A popular one is to take the average input and output neurons. In practice, it has been proved that a relatively simple neural network can approximate any function. After investigating several possibilities, we built a feed-forward network. The number of hidden layers is not high to don’t have a complex network and

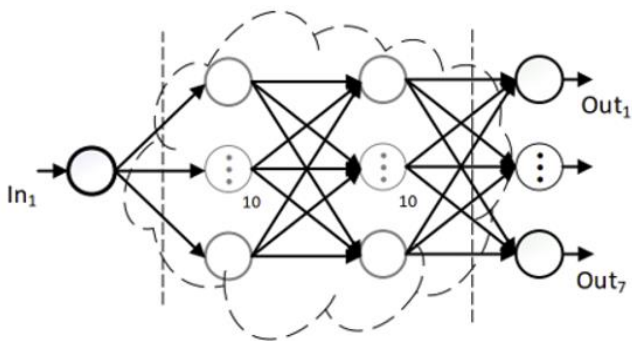


Fig.3. Deep neural network structure

not small so that the training is carried out correctly. The learning algorithm, Levenberg-Marquardt, is used to determine the best weights parameters, and the Mean Squared Error function to minimize the network error. We give an illustration of the built network in Fig. 3.

We have selected the training AC and DC datasets. The learning process consists of presenting the datasets to the input and outputs of the neural network, then calculating the optimal weights of the different neurons. Depending on the error obtained at the output, weights are corrected. We repeat this cycle until the network error curve is no longer increasing. Once the neural network is trained, we conducted tests to verify that responds correctly. After several tests, we obtained a DNN that provides output very close to the original values and allows us to produce reasonable estimates. Then, we estimate the 1-level DC from the 1-level AC datasets.

III. EXPERIMENTAL RESULTS

A. Datasets

Several hyperspectral datasets are available for the scientific community. For our tests, we consider the following hyperspectral Datasets: Kennedy Space Center, Cuprite, and Urban [3, 4, 19, 20]. The Kennedy Space Center and Cuprite are two datasets captured with the AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) sensor (Fig 4 and Fig 5). On the other hand, Urban is captured by the HYDICE (Hyperspectral Digital Imagery Collection Experiment) sensor (Fig 6). We present in detail all the used hyperspectral datasets in the following while their characteristics are listed and compared in Table II.

- **Kennedy Space Center (KSC)**

This scene was collected by the NASA AVIRIS instrument over the Kennedy Space Center (KSC) in Florida. The area covered comprises 512 lines by 614 samples in 224 bands of 10 nm width with wavelengths from 400 to 2500 nm. We get 176 bands after removing the noisy and water absorption bands. The spatial resolution of the KSC data is 18 m. Most parts of the image are composed of wetlands and various types of wild vegetation.



Fig.4. KSC (False greyscale image)

- **Cuprite**

Cuprite is one of the most popular hyperspectral datasets. Captured also using the AVIRIS sensor, it covers the cuprite mining areas in Nevada. Most parts of the image represent various minerals. It is acquired from an altitude of approximately 20 km and has a spatial resolution of 20 m. The cuprite data has 512x614 pixels and 224 bands. We get 188 after the bands 1-2, 104-113, 148-167, and 221-224 are removed because of dense water vapor and atmospheric effects.

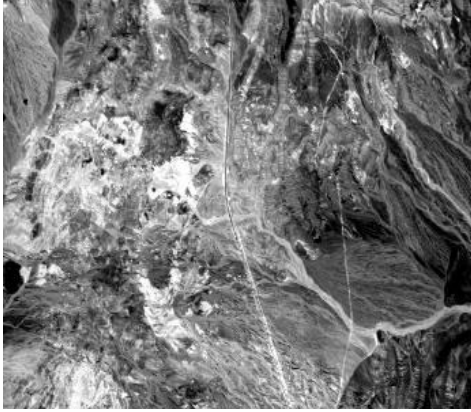


Fig.5. Cuprite (False greyscale image)

- **Urban**

Urban was collected by the HYDICE sensor and correspond to an urban image scene at Copperas Cove in Texas. The sensor works in 210 spectral bands, in the region from 400 to 2500 nm, with a spectral resolution of 10 nm. The data consist of 307x307 pixels, which are characterized by high spatial resolution (2 m). After removing the bands 1-4, 76, 87, 101-111, 136-153, and 198-210 due to dense water vapor and atmospheric effects, 162 bands remain available.



Fig.6. Urban (False greyscale image)

TABLE II. HYPERSPETRAL DATASETS AND THEIR INFORMATION

Datasets	<i>Cuprite</i>	<i>Urban</i>	<i>KSC</i>
Instrument	AVIRIS	HYDICE	AVIRIS
Image Size	512 x 614	307 x 307	512 x 614
Number of bands	224	210	214
Number of bands (selected)	188	162	176
Range	0.4-2.5 μm	0.4-2.5 μm	0.4-2.5 μm
Spectral Dimension	10 nm	10 nm	10 nm
Spatial Dimension	20 m	2 m	18 m

B. Results and Discussions

In this section, we compare the proposed WSRnet with the Bicubic and Spline interpolation benchmark. To do this, we adopt two quality indexes: the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index. The PSNR measures the proximity of the original image to its approximation one, and the SSIM measures the visual quality of the approximation image. A good PSNR indicates that a part of noise has been removed from the image. However, it not always corresponds to the best visual quality because the algorithm eliminates a part of the texture and not only the noise in the image. Therefore, we also use the SSIM metric that considers the similarity of the texture in the image. For a good SSIM, the algorithm must remove the noise and preserve the texture of the image.

However, the use of quality metrics is possible since we need a reference image to compare the results of the algorithms. For this reason, we down-sampled the HSI dataset by factor $\times 2$, and we considered the original sequence as a reference dataset. The comparison of PSNR and SSIM values over KSC, Cuprite, and Urban is shown in Table 3. We observe that our WSRnet surpasses the Bicubic and Spline with a larger margin. Because the wavelet transform has the function of restoring texture details and sharp parts of images, the WSRnet significantly boosts the results accuracy, respectively by 2% and 4% compared to the Bicubic and Spline for all datasets.

TABLE III. WSRNET RESULTS, PSNR/SSIM FOR SCALE FACTOR $\times 2$ ON DATASETS KSC, CUPRITE, AND URBAN

Datasets	Metrics	Bicubic	Spline	WSRnet
KSC	PSNR	30.88	25.93	31.22
	SSIM	0.827	0.781	0.81
Cuprite	PSNR	28.56	27.88	30.22
	SSIM	0.824	0.801	0.841
Urban	PSNR	32.99	26.39	33.22
	SSIM	0.871	0.728	0.901

The evaluation of the SR results is not only based on metrics results. Fig.7, Fig.8, and Fig.9 show the visual results of different techniques for an upscaling factor of 2. As we can observe, the WSRnet produces much sharper images than the two other approaches. In addition, we can observe that we have obtained images without any artifacts.

will concentrate on how to efficiently represent the spatio-spectral data with other wavelet transforms and how to increase the computational performance of the deep learning algorithm. Also, we will propose a modified algorithm to work on other super-resolution applications, such as medical image enhancement.



Fig.7. From left to right: a) The original “Urban” image from AVIRIS, b) Bicubic Interpolation, c) Spline Interpolation and d) WSRnet algorithm (all with an upscaling factor of x2)

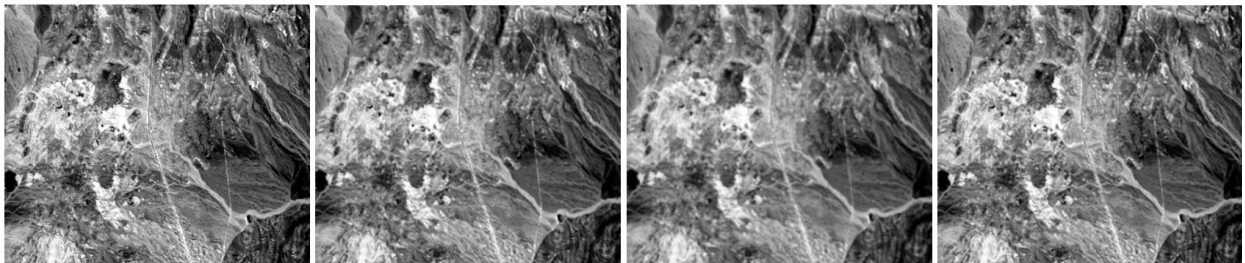


Fig.8. From left to right: a) The original “Cuprite” image from AVIRIS, b) Bicubic Interpolation, c) Spline Interpolation and d) WSRnet algorithm (all with an upscaling factor of x2)

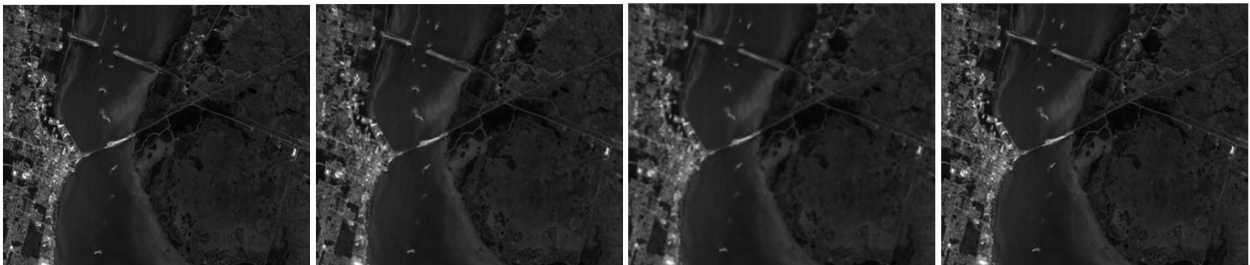


Fig.9. From left to right: a) The original “KSC” image from AVIRIS, b) Bicubic Interpolation, c) Spline Interpolation and d) WSRnet algorithm (all with an upscaling factor of x2)

IV. CONCLUSION

Deep learning has already been shown to be effective with hyperspectral images. In this work, a wavelet learning-based super-resolution technique (WSRnet) for HSI is proposed where a multi-scale wavelet decomposition is integrated into a deep neural network model. By exploiting the spatial and spectral features of the HSI and the properties of the deep neural networks, such as high-learning capacity and high performance, we developed a robust SR technique. The performance of the proposed technique has been evaluated on several hyperspectral images and compared with the classical algorithms. Based on the performance metrics such as PSNR and SSIM, the WSRnet produces good results. Moreover, all tests have shown significant improvement in the performance compared to some classical techniques. Our future research

REFERENCES

- [1] R. Pu, Hyperspectral remote sensing. CRC Press, 2017.
- [2] M. Eismann. Hyperspectral remote sensing. SPIE, 2013.
- [3] J. Chapman, "Spectral and Radiometric Calibration of the Next Generation Airborne Visible Infrared Spectrometer (AVIRIS-NG)", *Remote Sensing*, vol. 11, no. 18, p. 2129, 2019.
- [4] D. McKeown, S. Cochran, S. Ford, J. McGlone, J. Shufelt and D. Yocum, "Fusion of HYDICE hyperspectral data with panchromatic imagery for cartographic feature extraction", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 37, no. 3, pp. 1261-1277, 1999.
- [5] FE. Abd El-Samie, MM. Hadhoud and SE. El-Khamy, Image super-resolution and applications. CRC Press, 2019.
- [6] K. Nasrollahi and T. Moeslund, "Super-resolution: a comprehensive survey", *Machine Vision and Applications*, vol. 25, no. 6, pp. 1423-1468, 2014.

- [7] X. Tang and W. Pearlman, "Three-dimensional wavelet-based compression of hyperspectral images", in *Hyperspectral data compression*, G. Motta, F. Rizzo and A. Storer, Ed. Springer, 2006.
- [8] H. Yu, H. Zhang, B. Lei and C. Wang, "Hyperspectral image compressing using wavelet-based method", in *optical spectroscopy and imaging*, 2017.
- [9] M. Ouahioune, S. Ameer and M. Lahdir, "Enhancing hyperspectral image compression using learning-based super-resolution technique", *Earth Science Informatics*, vol. 14, pp.1173–1183, 2021.
- [10] M. Hassaballah and A. Awad, *Deep learning in computer vision*. CRC Press, 2020.
- [11] K. FUKUSHIMA, "Neocognitron: Deep Convolutional Neural Network", *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, vol. 27, no. 4, pp. 115-125, 2015.
- [12] S. Khan, H. Rahmani, S. Shah and M. Bennamoun, *A guide to convolutional neural networks for computer vision*. Morgan and Claypool, 2018.
- [13] N. Kuma, RN. Kumar, N. Sethi, "Learning to predict super-resolution wavelet coefficients", Tsukuba, Japan: Proceedings of the 21st international conference on pattern recognition, 2012, p. 3468.
- [14] Q. Chen, H. Song, J. Yu and K. Kim, "Current Development and Applications of Super-Resolution Ultrasound Imaging", *Sensors*, vol. 21, no. 7, p. 2417, 2021.
- [15] F. Sahito, P. Zhiwen, J. Ahmed and R. Memon, "Wavelet-Integrated Deep Networks for Single Image Super-Resolution", *Electronics*, vol. 8, no. 5, p. 553, 2019.
- [16] C. Dong, C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295-307, 2016.
- [17] Z. Wei and Y. Liu, "Construction of super-resolution model of remote sensing image based on deep convolutional neural network", *Computer Communications*, vol. 178, pp. 191-200, 2021.
- [18] N. Singh and M. A.N, "Image Super-Resolution Using Deep Learning Technique", *International Journal of Computer Sciences and Engineering*, vol. 6, no. 7, pp. 150-155, 2018.
- [19] U. Sara, M. Akter and M. Uddin, "Image Quality Assessment through FSIM, SSIM, MSE and PSNR - A Comparative Study", *Journal of Computer and Communications*, vol. 07, no. 03, pp. 8-18, 2019.
- [20] L. Akrou, S. Ameer, M. Lahdir, R. Fournier, and A. Nait Ali. "Fast Hyperspectral Image Encoder Based on Supervised Multimodal Scheme", *International Journal of Image and Graphics*, vol 21, no. 01, p.2150007, 2021.

Hyperspectral images compression based multimodal schemes

Akrour Leila[#], Ouahioune Mohand[#], Lahdir Mourad[#], Ameer Soltane[#], Fournier Régis^{*}, Nait Ali Amine^{*}

[#] *Laboratoire d'Analyse et de Modélisation des Phénomènes Aléatoires (LAMP A)*

Département d'Electronique, Université Mouloud Mammeri de Tizi Ouzou (UMMTO) Algérie

leilaakrou@yahoo.fr, ouahioune.m@gmail.com, mlahdir@yahoo.fr, ameursoltane@yahoo.com

^{*} *Laboratoire Images, Signaux et Systèmes Intelligents (LiSSi)*

Université Paris-Est Créteil ; Créteil France

rfournier@u-pec.fr, naitali@u-pec.fr

Abstract— In this paper, we present the multimodal approach for the compression of hyperspectral images. The technique highlights the contribution of the multimodal approach, whether supervised or unsupervised, to the field of hyperspectral images which is very much in demand nowadays in several fields of computer vision. The size of the initial hyperspectral sequence firstly reduced by applying the multimodal concept based on the insertion function. Some selected images of the initial hyperspectral sequence inserted into the other one of the same sequence to generate a multimodal sequence. Then, the new multimodal sequence is encoding efficiently by any 3D-Encoder. At the decoding step, a separation function is necessary to extract the inserted images from the decoded multimodal sequence in order to reconstitute the approximation of the initial sequence. The performance of the Multimodal Compression Approach encoder are evaluated on AVIRIS hyperspectral images. Experimental results indicate that the proposed approach provides very promising performance at low bit-rates while reducing the encoding/decoding time.

Keywords— Hyperspectral image, multimodal compression, supervised insertion, non-supervised insertion, Quadtree decomposition, 3D-SPIHT.

I. INTRODUCTION

In recent years, hyperspectral imagery has attracted a great deal of interest hence its use in several research areas such as target detection, classification, anomaly detection, object discrimination, material quantification, and spectral unmixing due to its ability to provide spatial and spectral information details. Hyperspectral sensors produce daily a large amount of data that is considered as 3D data (hundreds of contiguous spectral bands are collected from the visible and the near infrared spectrum light and each band has the same number of pixels and represent a specific bandwidth of the electromagnetic spectrum.). The large capacity of memory required to store these hyperspectral images (HSI) implies the use of compression and decompression algorithms allowing reducing as much as possible the amount of data to store or to transmit.

HSI compression is a technique through which the size of HSI can be reduced. Ideally, the compression should be lossless to preserve the scientific value of data. However,

lossless compression techniques allow for very limited compression ratios. To achieve much higher compression ratios, the lossy compression is used introducing some distortion between the reconstructed and the original image.

Several compression algorithms more suitable for this hyperspectral data are proposed in the literature, most of these compression techniques are two-dimensional (2D) transform-based such as discrete cosine transform (DCT), discrete wavelet transform (DWT) and Karhunen-Loève transform (KLT), that has extended to three-dimensional (3D). Some state-of-the-art algorithms in these fields are 3D-DWT [1], 3D-KLT [2], 3D-Set partitioning in hierarchical tree (3D-SPIHT) [3], etc.... However, it is very rare that these methods take into consideration the complexity cost nor the time needed for the compression and decompression steps, especially for this type of data with large amounts of information to process. Knowing that today we are in the era of speed and real-time transmission, it is of the utmost importance to focus primarily on the performance that a compression method can achieve, but also on the time required for its execution.

In this context, several works are developed improving the performances to be achieved as well as the times to be respected in order to be among the best possible in the hyperspectral compression domain [4, 5-6, 7]. The wavelet-based zero-tree algorithms such as the 3D-SPIHT encoder is a good candidate for their performance in HSI compression [8-10]. However, application of this algorithm on HSI induces a great computational complexity and consequently a great compression/decompression times given the volume of data to be compressed. In order to reduce the amount of data to be compressed, thus reducing the computational complexity and the compression time, we have exploited the principle of Multimodal Compression (MC). As defined in reference [11], the idea of MC consists in merging different form of biomedical data using an insertion function before encoding process to compress the fusion data with the same encoder. In this paper, the synthesis of multimodal compression is performed to highlight the contribution of the multimodal approach, whether supervised or unsupervised, for the field of hyperspectral images which is very solicited nowadays in several fields such as computer vision. For the multimodal

compression of HSI some selected images from the initial hyperspectral sequence are inserted in the remaining images of the same sequence producing a compacted multimodal sequence to compress using a 3D-Encoder which outperforms the others compression standards. This paper is organized as follows: section II presents the methodology of Hyperspectral images Multimodal Compression (HSI-MC). The HSI-MC

schemes, the results and the performance analysis evaluated on AVIRIS database (Airborne Visible Infrared Imaging Spectrometer) are presented in section III. Section IV provides a conclusion and a perspective for the multimodal compression of the HSI.

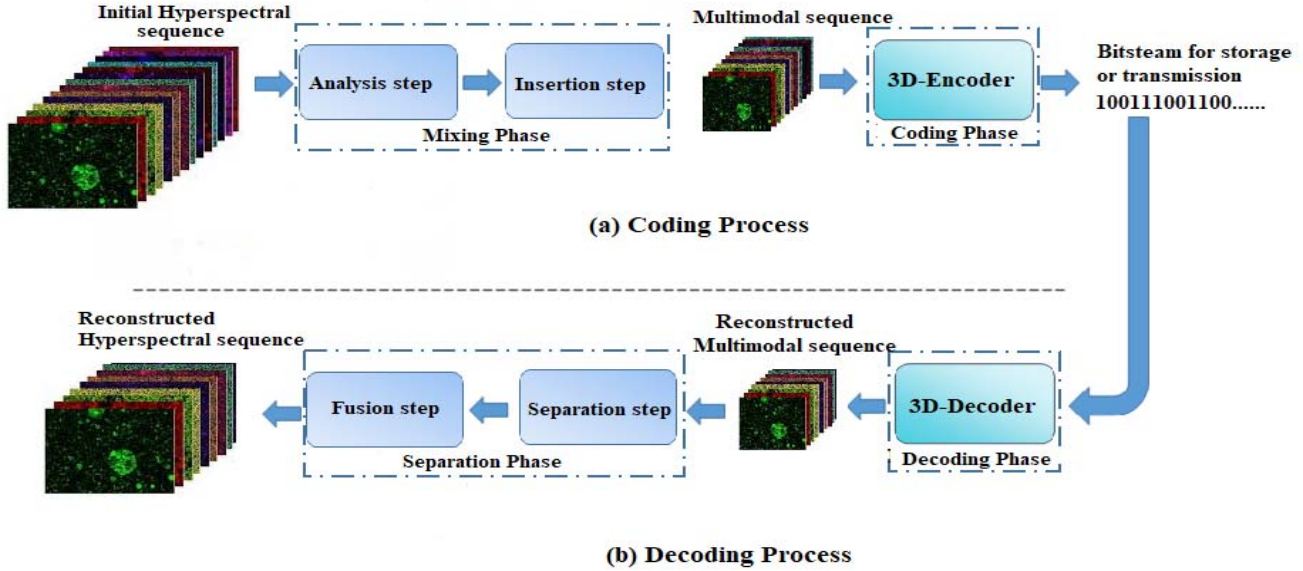


Fig. 1. Multimodal compression scheme

II. METHODOLOGY

The HSI-MC is the coding of a mixture of images, which is at the core of the compression/decompression process. The multimodal sequence formed is the result of inserting some selected images from the initial sequence into other images of the same sequence. Then, an efficient 3D-Encoder applied to the compacted sequence in order to generate the bitstream to store or to transmit. The multimodal compression of HSI requires various steps, namely: Mixing phase and Coding phase for the Coding Process. Decoding phase and Separation phase for the Decoding Process (Fig. 1).

A. Coding Process

The first steps of this process allow generating a multimodal sequence from the initial hyperspectral sequence to be compressed reducing by this way the size of the initial sequence.

- **Mixing phase**

We consider an initial hyperspectral sequence S_k where $k = 1, \dots, Z$ refers the bands. Each band represent an image of $M \times N$ pixels. In this phase, the initial sequence S_k is reduced by applying the analysis step followed by the insertion step.

1) Analysis step

In this step, the initial hyperspectral images classified into Host images and Guest images. For this purpose, a simple statistical analysis is used to perform this selection. So, the variance of each image $I(i, j, k)$ of initial sequence is used as an objective criterion to define those containing less information rate as Guest images to be insert in the other ones classified as Host images. The variance is estimate using the following equation:

$$v(k) = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (I(i, j, k) - \overline{I(k)})^2 \quad (1)$$

Where $k=1, \dots, Z$ (Z represent the number of hyperspectral bands), the average of the image (band) of size of $M \times N$.

According to the [12] and [13], the variance of some images of the sequence is much higher than the other one. Thus, we can define that these values show a greater concentration of information in these images. Therefore, the images with the lowest variance value of the initial sequence are listed as the one to be inserted into the other.

2) Insertion Step

An insertion function is used to generate the multimodal sequence from the initial sequence.

Definition: “an insertion function is a procedure that consists in replacing some pixels of a host image by other useful pixels provided by another source according to a rule defined by the user and corresponding to the application” [12].

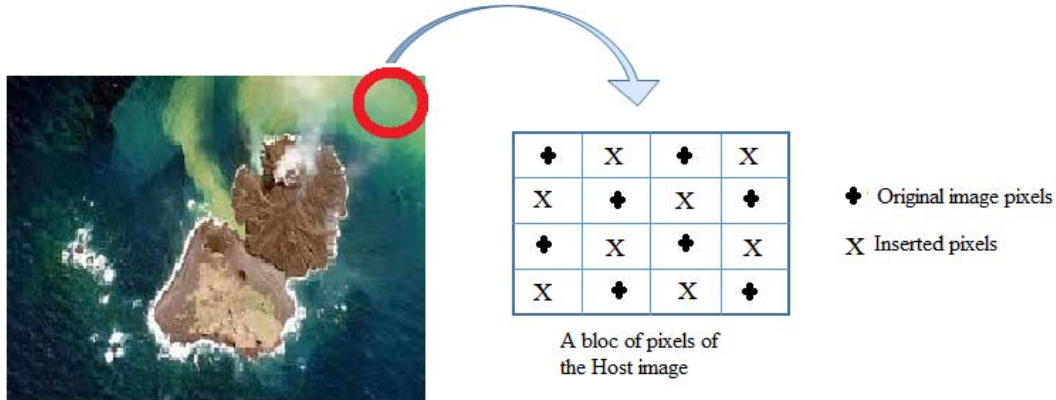


Fig. 2. Insertion pattern

- Coding Phase

The multimodal sequence thus formed is compressed using 3D-Encoder based on wavelet transform [1] to generate the bitstream for storage or transmission and thus in order to exploit efficiently the spectral and the spatial redundancies in hyperspectral images [12]. Consequently, a fast and very efficient compression technique for hyperspectral images achieved while keeping the images features intact.

B. Decoding Process

In this phase and according to Fig. 1(b), the 3D-Decoder based on the inverse 3D-Wavelet transform applied to the received bitstream and a reduced volume containing the data mixture obtained which represent the decoded multimodal sequence. The separation function of the decoding step involves extracting the inserted pixels previously in the insertion step from the reconstructed Host images of the decoded multimodal sequence to reconstruct the Guest images. The replaced pixels of the reconstructed Host images cannot be recovered to their exact values. These pixel values need to be estimated using a cubic spline interpolation [13]. The fusion step is used to reconstruct a decoded hyperspectral sequence, which is an approximation of the initial sequence.

The reconstructed Host images are reorganized with the reconstructed Guest images to generate the decoded sequence.

The adopted insertion used in this step is to replace each pixel over two in the Host images by the one of the Guest images according to the Fig. 2. The number of the Guest images to eliminate from the initial sequence depends on the reception capacity of the Host images. At the end of these steps, a multimodal sequence with a reduced size is formed.

III. THE SCHEMES OF THE MULTIMODAL COMPRESSION OF HYPERSPECTRAL IMAGES

A- Non-Supervised Multimodal Compression scheme

The basic idea of this scheme is to insert the pixels of the Guest images into reception regions using a spiral insertion. These reception regions defined by the final user of HSI on the contours of the Host images delimited by a Region of Interest (ROI). The defined ROI contain the main information that should not be distorted. Therefore, the number of Guest images that can potentially be merged depends on the global size of the ROI. After the Non-Supervised insertion step based on the spiral insertion, a multimodal sequence is formed thus is compressed by the 3D-SPIHT encoder based on the 3D-wavelet transform [12].

To evaluate the performance of the Non-Supervised Multimodal Compression based on 3D-SPIHT encoder, experiences have been performed on several hyperspectral sequences from AVIRIS database (Fig. 3). We take specifically scene 0 of the Yellowstone dataset acquired in 2006. Fig. 3 shows some images (bands) of the taken sequence. This set of data can be download from the official website. In addition, the quality of each image of the decoded sequence is estimated by using various image quality metrics as PSNR metric (Peak Signal-to-Noise Ratio).

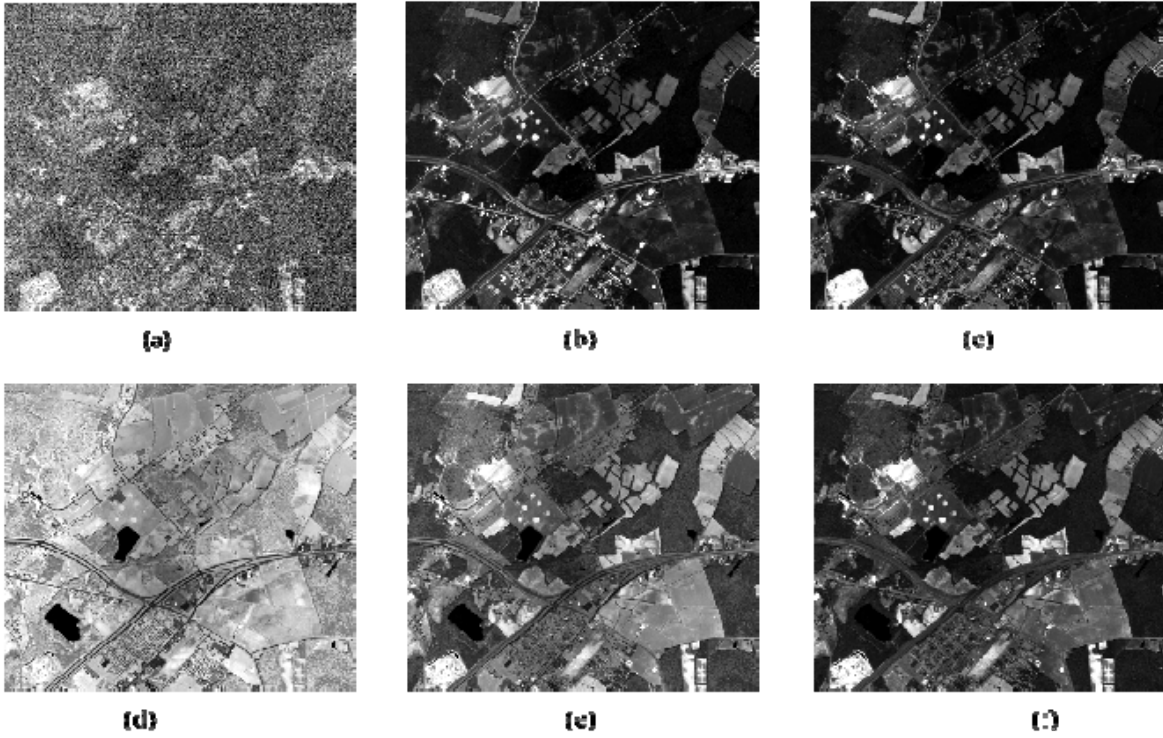


Fig. 3. Dataset of Yellowstone scene 0. (a) Band 2, (b) Band 10, (c) Band 30 (d) Band 90, (e) Band 150 and (f) Band 200.

Fig. 4 shows of the reconstructed multimodal sequence for Yellowstone scene (band 30) for 0.5 bpp and 1 bpp. In terms of quality, one can conclude that the non-supervised multimodal compression preserves the information without any significant distortion.

According to the Fig. 5, one should note that for low bit-rate, the quality of the decompressed images is objectively almost the same, whereas for high bit-rate, the quality remains subjectively the same (based on the visual quality). This can be explained by the fact that when dealing with PSNRs greater

than 50 dB (which is the case here), it becomes very difficult to distinguish, visually between image qualities. Consequently, one can conclude that the non-supervised multimodal compression method preserves the information without any significant distortion. In terms on computing time, the non-Supervised Multimodal Compression technique becomes particularly interesting since the encoding or the decoding steps are both achieved on a reduced number of images.

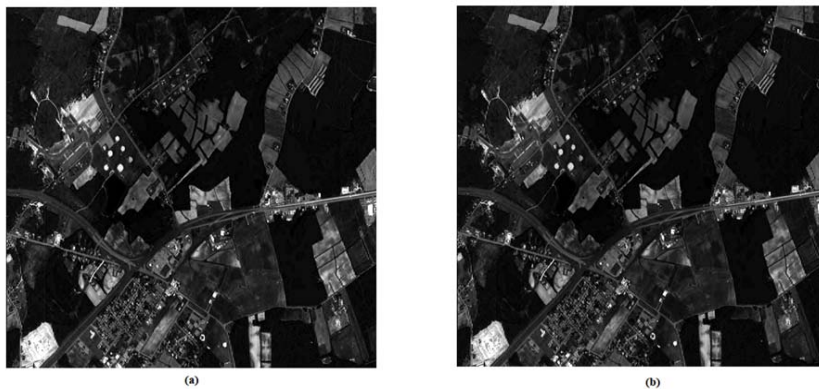


Fig. 4. Decoding results for Yellowstone scene images (Band 30). (a) 0.5 bpp, (b) 1 bpp

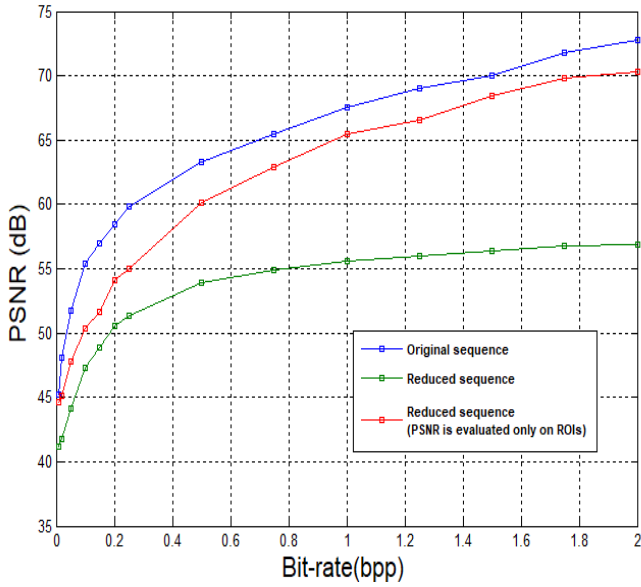


Fig. 5. Non-supervised multimodal compression results.

B- Supervised Multimodal Compression scheme

In the scheme cited in [13], the selected Guest images from initial hyperspectral sequence are inserted in the remaining images defined as the Host one of the same sequence by using the supervised insertion step based on Quadtree decomposition [14]. This decomposition is used to find the homogeneous blocks that will receive the pixels of the selected Guest images. The end of this operation produces a compacted sequence, which compressed efficiently by the 3D-SPIHT encoder. In this case, the number of Guest images to be inserted in the Host one, is defined according to the reception capacity represented by the different homogeneous blocks. Which allows us to define several insertions depending on this reception capacity. Ins2, Ins5, Ins8, Ins11 and Ins14 represent respectively the number of the Guest images to be inserted on the Host one (2 images, 5 images, 8 images, 11 images and 14 images).

In order to evaluate the performance of the Supervised Multimodal Compression based on 3D-SPIHT encoder (SMC-3D-SPIHT encoder), the quality of each decoded images estimated by using various image quality metrics as the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index (SSIM) [13, 15-16]. The results show that for low bit-rate, the SMC-3D-SPIHT encoder achieves good image quality with PSNR values neighboring 45 dB while computing times are much reduced (Fig. 6). The results are improved significantly by the reduction of the initial sequence accomplished by the optimized insertion step based on supervised way. Consequently, higher compression rates and significant reduced computing time achieved with SMC-3D-SPIHT encoder. Moreover, the variations in the spectrum of the compressed images are almost identical to those of the

original images (Fig. 7). Thus, the supervised multimodal compression preserves most part of spectral data signature that is used to identify and/or classify materials, objects, etc. [13].

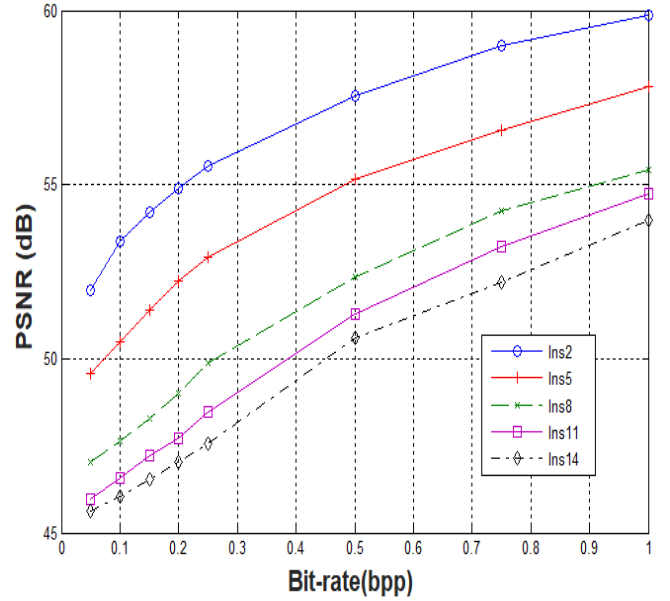


Fig. 6. Supervised multimodal compression results.

IV. CONCLUSION

In this work, the multimodal approach for hyperspectral image compression is presented. This technique highlights the contribution of the multimodal approach in the field of hyperspectral images which is very much in demand today in several fields of computer vision. Two schemes developed for multimodal compression of hyperspectral images, namely non-supervised and supervised multimodal compression. In order to exploit efficiently the spectral and the spatial redundancies in hyperspectral images, a lossy compression method-based 3D-SPIHT encoder defined. With this method, the amount of initial hyperspectral data reduced by applying the multimodal concept. The 3D-SPIHT encoder compresses the multimodal sequence thus obtained. All the results shows that the multimodal compression improve the performances of the hyperspectral images compression. One can conclude that the multimodal compression of the hyperspectral images is computationally a very fast encoder for low bit-rates compression; this technique offers a low computational complexity that is highly desirable at on-board system, with limited computational and power capabilities. The multimodal compression approach of hyperspectral images presents spectral fidelity, which is substantially useful in specific applications such as anomalies detection or data analysis.

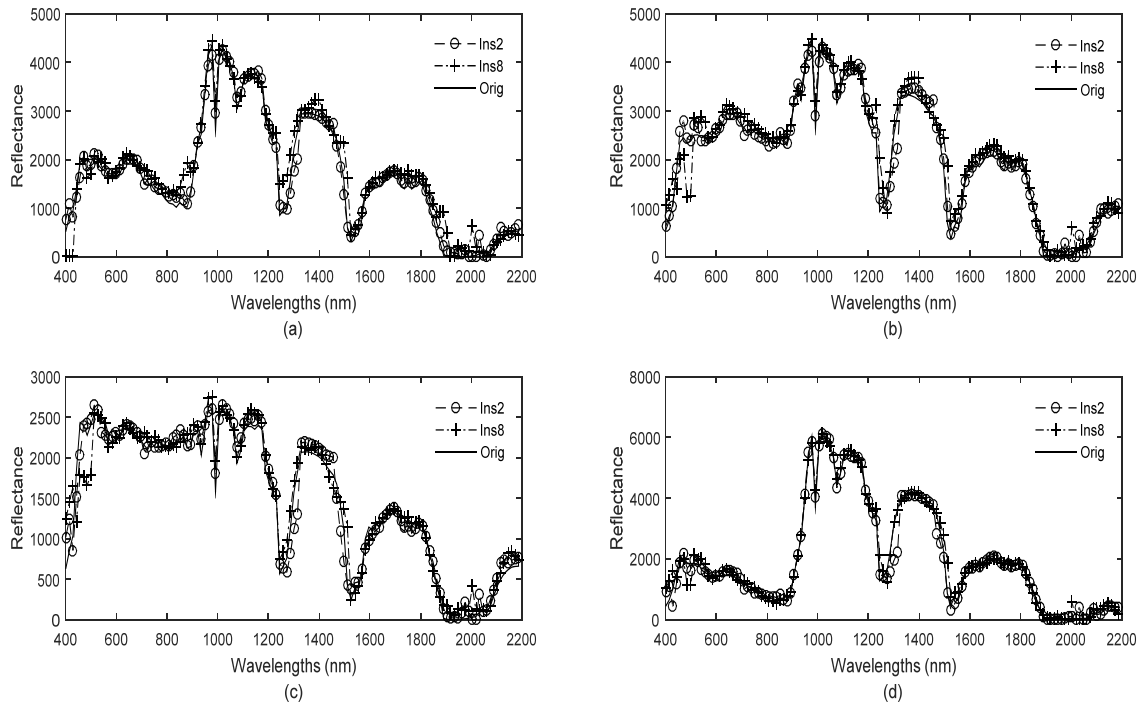


Fig. 7. Spectra generated from original and compressed reflectance for Yellowstone scene of AVIRIS database (0.05 bpp), (a) road, (b) field, (c) building, and (d) ground.

REFERENCES

- [1] [X. Tang, W. A. Pearlman and J. W. Modestino, Hyperspectral image compression using three-dimensional wavelet coding, *Proc. SPIE Image Video Communication and Processing*, pp. 1037–1047, 2003.
- [2] I. Blanes, E. Magli and J. Serra-Sagrista, A tutorial on image compression for optical space imaging systems, *IEEE Geosci. Remote Sens. Mag.* 2(3), pp. 8–26, 2014.
- [3] A. Said and W. A. Pearlman, A new, fast, and efficient image codec based on set partitioning in hierarchical trees, *IEEE Trans. Circ. Syst. Video Tech.* 6(3), 1996.
- [4] Y. Dua, V. Kumar and R. S. Singh, Comprehensive Review of Hyperspectral Image compression algorithms, *SPIE Optical Engineering Journal*, 59(9), 2020, 39p.
- [5] R. Pizzolante and B. Carpentieri, On the compression of hyperspectral data, *IT Convergence Practice (INPRA)*, 1(4), 2012, 24 p.
- [6] R. Pizzolante and B. Carpentieri, Band clustering for the lossless compression of aviris hyperspectral images *Int. J. Signal Image Process.* 5(1), 14 p, 2014.
- [7] K.S. Babu, V. Ramachandran, K. K. Thyagarajan, G. Santhosh, *Hyperspectral Image Compression Algorithms - A Review*. Book chapter: Artificial Intelligence and Evolutionary Algorithms, In *Engineering Systems. Advances in Intelligent Systems and Computing*, Vol. 325, pp. 127-138, https://doi.org/10.1007/978-81-322-2135-7_15.
- [8] E. Christophe, Hyperspectral data compression tradeoff, *Remote Sens.* 3, pp. 9–29, 2011.
- [9] M. Ouahioune, S. Ameer and M. Lahdir, enhancing hyperspectral image compression using learning-based super-resolution technique, *Earth Science Informatics*, 14, pp. 1173–1183, 2021.
- [10] N. Zikiou, M. Lahdir and D. Helbert, Support vector regression-based 3D-wavelet texture learning for hyperspectral image compression, *Vis Comput.* 36:1473–1490, 2019. <https://doi.org/10.1007/s00371-019-01753-z>.
- [11] A. Nait-Ali, E. H. Zeybek and X. Drouot, *Introduction to Multimodal Compression of Biomedical Data*, In: A. Nait-Ali, Ed., *Advanced Biosignal Processing*, Springer, Berlin, pp. 353-375, 2009.
- [12] M. Lahdir, A. Nait Ali and S. Ameer, Fast encoding-decoding of 3d hyperspectral images using a non-supervised multimodal compression scheme, *J. Signal Inf. Process.* 2, pp 316–321, 2011.
- [13] L. Akrou, S. Ameer, M. Lahdir, R. Fournier and A. Nait-Ali, Fast Hyperspectral Image Encoder Based on Supervised Multimodal Scheme, *International Journal of Image and Graphics*, 21(1), 23 p, 2021. <https://doi.org/10.1142/S0219467821500078>.
- [14] G. M. Hunter and K. Steiglitz, Operations on images using quad tree, *IEEE Trans. Pattern Anal. Mach. Intell.* 1(2), pp. 145–153, 1979.
- [15] Z. Wang and A. C. Bovik, Image quality assessment: From error visibility to structural similarity, *IEEE Trans on Image Processing* 13(4), 2004.
- [16] Z. Wang and A. C. Bovik, Modern image quality assessment, *Synth. Lect. Image Video Multimedia Process.* 2(1), pp. 1–256, 2006.

Design of a New Multi-sensor System Probe For Remote Wind Parameters Monitoring

B. Idjeri^{1*}, M.T. Bensidhoum², M. Laghrouche¹

¹ Mouloud Mammeri University of Tizi Ouzou, Algeria, LAMPA Laboratory

² Mouloud Mammeri University of Tizi Ouzou, Algeria, LVAAS Laboratory

^{1*} boussad.idjeri@ummto.dz

¹ larouche_67@yahoo.fr

² bensidhoumtahar@gmail.com

Abstract—Assessment of the wind potential in large regions accurately needs a considerable number of devices distributed over several remote locations, which is inappropriate with the classical wind sensors. It should be noted that MEMS (Micro Electro Mechanical System) sensor based anemometers are more suitable for such applications thanks to multiple advantages, among others, they are compact sized, light in weight, making them well adapted for large-scale integration. Furthermore, due to the possibility of embedding several sensors in one chip and their mass manufacturing, they are low in cost. On the other hand, their major drawbacks lie in their great sensitivity to changes in other physical parameters such as air temperature and their incapacity to detect the flow direction. In this regard, this paper presents the design and testing of a novel multi-sensor system probe based on MEMS hot-film anemometer for wind speed and direction measurement. The system probe consists of a cylindrical support on which a hexagonal MEMS hot-film sensor chip is disposed and calibrated in a wind tunnel. After that, an algorithm based on ANNS (Artificial Neural Networks) is elaborated and implemented on an Arduino board to calculate the corresponding values of wind speed and direction while compensating the influence of changing in air temperature. Finally, in order to evaluate the performances of the system, outdoor tests are carried out at wind monitoring points measured by a classical anemometer. The obtained results are consistent with each other over the entire measurement period.

Keywords— Hexagonal MEMS sensor chip; Hot film anemometer; ANNS (Artificial Neural Networks); Air flow speed; Air flow direction; Temperature correction; Arduino.

I. INTRODUCTION

Estimation of wind conditions plays an important task in many fields such as aeronautic, marine, and renewable energies which have aroused the need to develop effective systems to meet the demands of the research industry. In this context, hot wire anemometry especially thermal flow MEMS (Micro Electro Mechanical System) sensors are more appropriate when it comes to estimating the wind in large-scale outdoor areas with high spatial and temporal resolution. With the advent of micromachining technologies, a miniature hot wire sensor was fabricated by means of integrated silicon-based technologies [1,2]. One of the main assets of this type of

sensor is that it can be used to estimate other physical parameters such as air temperature and humidity [3]. However, its major shortcomings lie in its dependence on air temperature [4,5] and its inability to detect flow direction [1]. In order to circumvent the temperature drift effect, there are basically two methods based either on: hardware or software scheme. A standard hardware procedure uses a compensation resistor with negative temperature coefficient in the bridge circuit, which can compensate for certain temperature range with a precise component selection and circuit structure [4], [6-7]. The software one requires temperature measurement and includes the temperature value in the wind computing algorithm to attain correct wind speed with temperature drift correction as it has been established for the first time in [8]. As for the wind direction measurement, the solution relies on using several sensor chips [4-6] or multiple sensors configured on a single chip [9]. It is worth mentioning that, in the last few decades, wind related industries have invested a great deal of money in researches to improve the qualities of such wind sensors. These researches aim to optimize their costs, power consumptions, responses time, accuracies and enhance their sensitivities that have been obtained through various advanced schemes, tools and wind monitoring techniques [5], [10-15]. Thus, in this paper we intend to develop a 2D low-cost, wearable multi-sensor system with economical power consumption to measure wind speed and direction with high precision. To that end, we opted for a MEMS chip with hexagonal configuration, which is a good tradeoff between precision and power consumption, and suggest placing it on a probe structure. After testing different probe structures our choice fell on a cylindrical shape which is proven to be the most suitable form. After that, the probe, its conditioning and data acquisition system were calibrated in a wind tunnel. With the obtained results a data set was constructed and an algorithm based on ANNs was elaborated to calculate the corresponding speed and direction while accounting for the influence of variations in air temperature. Afterwards, the elaborated algorithm was implemented on an Arduino board. In the end, tests in the open air were carried out at high altitude on the designed prototype anemometer together with a classical anemometer at the same places. The obtained results in the same time frame were compared. The concordance of

the results, such as the mean absolute error (MAE) between the two instruments is in the order of 0.019 m/s for speed and 4° for direction, confirmed the feasibility of the hexagonal MEMS based anemometer probe. The detailed steps of the multi-sensor system probe design and tests carried out are detailed in the following sections.

II. EXPERIMENTAL

A. MEMS flow sensor

The sensor under study is composed of a thin polycrystalline silicon film forming the sensitive element which is suspended on an air cavity, two metallic electrodes, a silicon nitride layer, and a substrate as shown in Figure.1. The different steps of its realization are detailed in [1].

- 1 Polysilicon suspended wire
- 2 Metallic electrodes
- 3 Air cavity
- 4 Silicon nitride film
- 5 Silicon substrate

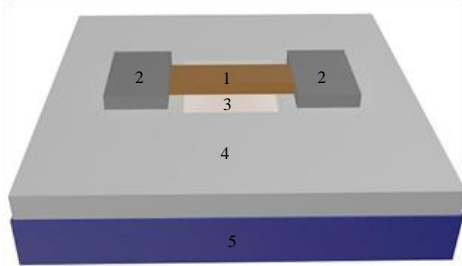


Fig.1. Structure of the hot wire sensor.

For a given ambient temperature (T_a), the silicon film is crossed by a dc current (I) which heats it at temperature (T). As air blows on the hot wire, air cooling causes resistance changes which are detected so as to make wind speed measurements possible. A scanning electron microscopy (SEM) picture of the fabricated MEMS hexagonal based hot-film sensor is shown in Fig. 2.

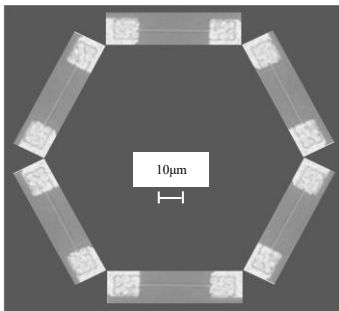


Fig. 2. SEM image of the fabricated hexagonal MEMS hot-film sensor chip.

Let R be the resistance of the hot wire sensor and R_a its resistance at ambient temperature. At temperature T , the sensor resistance is given by:

$$R = R_a [1 + \alpha(T - T_a)] \quad (1)$$

$$\text{With } \alpha = \frac{1}{R_0} \frac{\partial R}{\partial T} \quad (\text{Where } R_0 = 2.9 \text{K}\Omega \text{ is the resistance at } 0^\circ\text{C}).$$

α is the TCR (Temperature Coefficient resistance) of the hot wire sensor and R_a its resistance at ambient temperature T_a .

Figure.3 shows the $R(T)$ curve obtained for a $50 \times 2 \times 0.5 \mu\text{m}^3$ hot wire. The sensor exhibits a high degree of linearity as a function of temperature permitting to determine the TCR value equal to $0.142\%/^\circ\text{C}$.

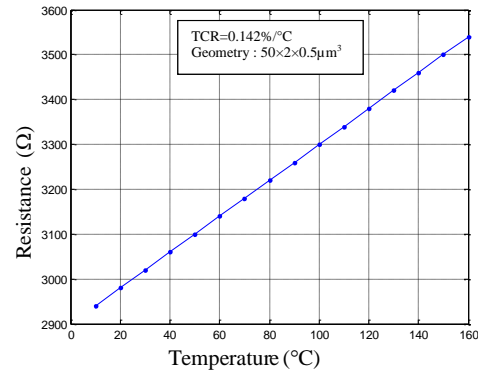


Figure.3. Resistance variations against temperature for a $50 \times 2 \times 0.5 \mu\text{m}^3$ polycrystalline silicon wire.

The relation governing the resistance variation is given by equation (2) [1]:

$$R = \frac{(T - T_a)dw(a + bv^{\frac{1}{2}})}{\left[1 - \frac{2}{d} \left(\frac{\lambda h}{a + bv^{\frac{1}{2}}} \right)^{\frac{1}{2}} \right] \left[\frac{d}{2} \left(\frac{a + bv^{\frac{1}{2}}}{\lambda h} \right) \right] I^2} \quad (2)$$

Where (d) the length, (w) the width and (h) the thickness of the polysilicon film. k and λ are respectively the material emissivity and thermal conductivity of the polysilicon material film, v is the flow velocity.

The coefficients a , b are evaluated empirically from the current–voltage and resistance–temperature curves of the sensor under study.

B. Sensor conditioning circuit

Considering Eq.2, the sensor resistance depends on both the variations of the electric current and the fluid temperature. Therefore, two measurement methods can then be used. The first method consists in using the electronic circuit of Fig. 4 to supply the hot wire sensor with a constant current which is the Constant Current Anemometer (CCA). Whereas the second one is based on the regulation of the sensor temperature by means of an electronic control circuit which is the Constant Temperature Anemometer (CTA). In this paper, we opted for the Constant Current Anemometer because of its simplicity. As shown in Figure.4 the conditioning circuit is composed of a Wheatstone bridge fed by a constant current source of 2 mA, an INA128 operational differential amplifier whose differential gain is adjusted via an adjustable resistance R_{out} .

The Wheatstone bridge includes the hot wire sensor, two identical resistances R_1 and R_2 of 0.1% accuracy, and a variable resistance to adjust the initial current flowing through the sensor.

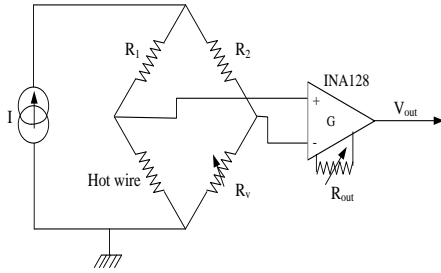


Figure.4. Constant current anemometer conditioning circuit principle.

Considering the supply current for each CCA equal to 2mA, the total power consumption of the six MEMS hot-film sensors is estimated to be in the range of 30mW.

C. Structure of the multi-sensor system probe

The probe support is built from hard plastic material because of its long lifetime, excellent heat and water resistance with high electric insulation. A hexagonal concave region was milled on the top surface of the cylinder for embedding the sensor chip as shown in Figure.5. A PCB board embedding the sensors' conditioning circuit is disposed in the base of the support, and connected to the sensors via bond wires.



Figure.5. The probe embedded with the hexagonal MEMS hot-film sensor chip in its cavity.

D. Data acquisition system

A data acquisition system built in the laboratory is precisely devised to measure air temperature, wind speed and direction. Its block diagram is shown in fig. 6. It is composed of several hot wire MEMS sensors (six hot wires to measure wind speed and direction and one hot wire to measure air temperature), their conditioning circuits and a microcontroller based on Arduino Due with Atmel 32 bits ARM Cortex-M3 CPU coupled with an XBee module. The obtained data are then transferred to a distant computer.

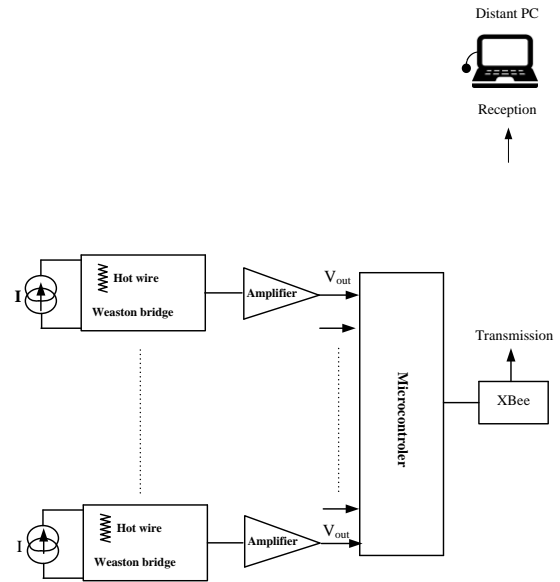


Figure.6. Block diagram of the data acquisition system.

E. Wind Tunnel Test and Calibration Setup

The Multi-sensor System Probe with its conditioning circuit and the data acquisition system were tested and calibrated in a in a closed-circuit wind tunnel as shown in Figure.7.



Figure.7. The anemometer probe in the wind tunnel.

The probe was disposed firstly such that the sensor1 will be facing the wind direction marked as 0° . Different wind speeds were used from 0 m/s to 30 m/s with a step of 1 m/s and wind direction from 0° to 360° at 15° increment. For each speed and direction, the readings of the sensors were registered and transferred to a distant PC via an XBee module. After that the same experience was repeated at different air temperatures: 15°C , 20°C , 25°C , and 30°C respectively, to account for the influence of variation in fluid temperature and address it. Figures 8 and 9 show some measured curves of the sensors versus flow direction angle at 14m/s with a direction interval of 15 degree at air temperature equal to 15°C and 20°C respectively.

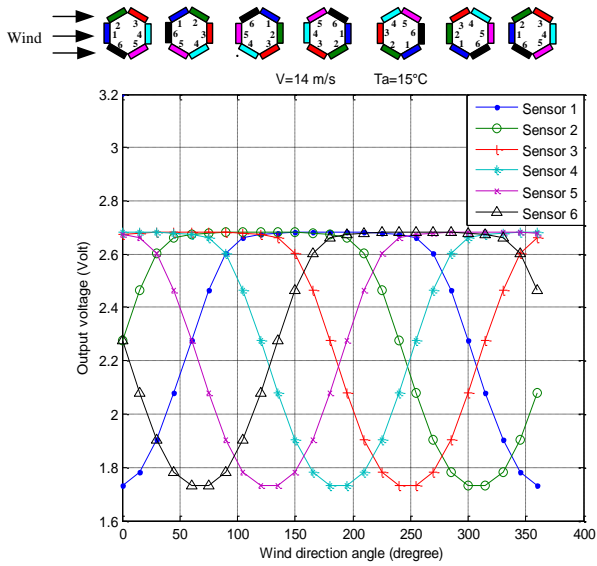


Figure.8. Output voltage according to wind direction in the wind tunnel at wind speed 14 m/s and fluid temperature 15°C.

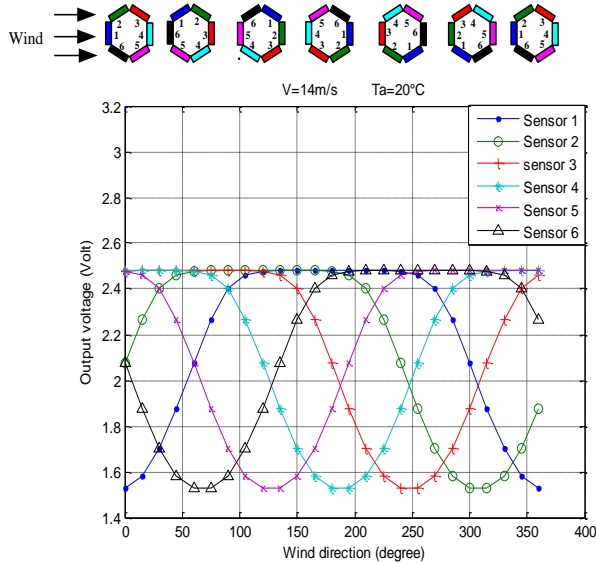


Figure.9. Output voltage according to wind direction in the wind tunnel at wind speed 14 m/s and fluid temperature 20°C.

The differences of the sensors output voltage between figures 8 and 9 as it can be noticed, is due to the influence of the changing in air temperature which causes a shift of the curves along all the range of the wind direction. Therefore, a correction method is necessary to obtain a unique response for each speed and direction irrespective of air temperature.

F. Speed and Direction calculation on the basis of ANNs

The flow direction measurement is based on the relative output difference of the six sensing elements in response to temperature variation induced by airflow. The principle of airflow detection speed and direction is illustrated in Fig. 8. The relationship between the sensors' readouts and the flow

parameters is a multiple input multiple output coupling system, which can be formulated by:

$$[V, \alpha] = f(E_1, E_2, E_3, E_4, E_5, E_6, E_7) \quad (3)$$

In this regard, a procedure based on ANNs was adopted to solve this equation system, as it had been theoretically proved that three layers of neural network could solve arbitrarily complicated nonlinear mapping problems [16]. The structure of the network is shown in Figure 10, which composed of 3 layers (input layer, hidden layer and output layer) with 7 input neurons equal to the number of sensors, one hidden layer with 9 neurons containing sigmoid function experimentally determined and 3 output neurons containing linear function corresponding to the flow parameters. The neural network was elaborated with the data base constructed in the calibration process which is divided into 2 parts (training and test sets). The whole data included 2480 patterns for training and 496 patterns for testing. The weights and biases of the network were determined through training data which is performed by the Levenberg–Marquardt back-propagation algorithm. After the learning stage, the network was submitted to the test. The number of neurons in the hidden layer was selected experimentally on the basis of minimum Mean Square Error (MSE) for the learning and testing sets. In the present case, this choice fell on 9 because this was the smallest number of hidden neurons with acceptable least square errors over the training and test regions which are both in the order of 10^{-5} .

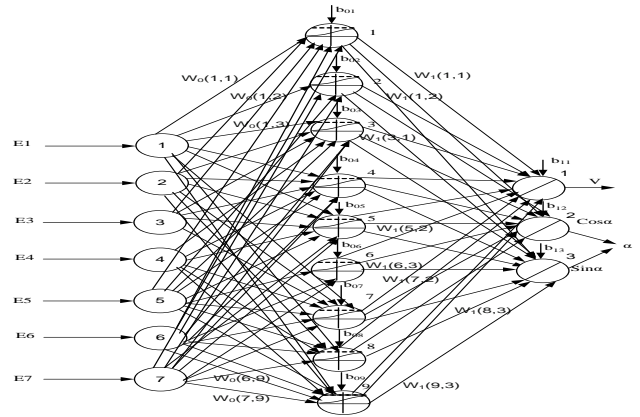


Figure.10. Structure of the neural network.

After designing the neural network, it was implemented on Arduino DUE board and applied immediately to the wind tunnel test measurements to calculate speed and direction. The scheme of its implementation is detailed in our previous work [17].

H. Field-test of the designed system

Field-test of the designed system was presented in our previous method [17], where the anemometer probe was mounted beside a classical anemometer for comparison over a period of 60 seconds. After the tests, wind speed and direction were calculated for every second in the time frame. As shown

in Figures 11 and 12 the measurement results obtained by the MEMS-based anemometer were compared with those recorded by the classical anemometer. The obtained results are consistent with each other over the entire measurement period.

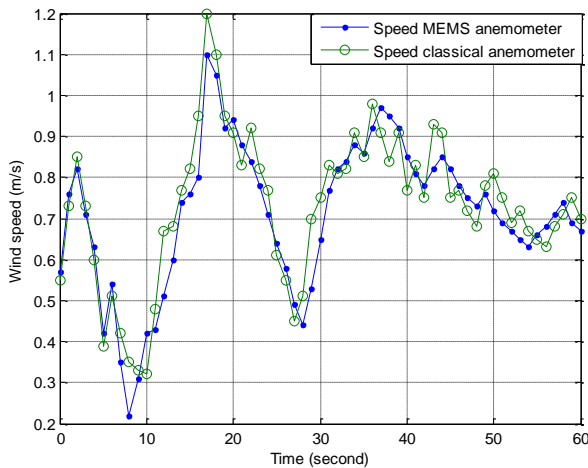


Figure.11. Field-test comparison of wind speed between the MEMS anemometer probe and a classical anemometer.

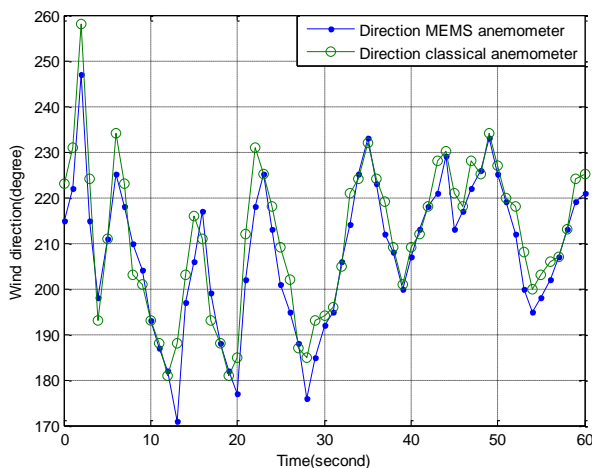


Figure.12. Field-test comparison of wind direction between the MEMS anemometer probe and a classical anemometer.

The mean absolute error (MAE) between the two instruments is in the order of 0.019 m/s for speed and 4° for direction. These differences can be explained mainly by the difference between instruments characteristics such as the responses time which may cause a delay and the drastic changes in wind speed and direction leading in some cases to instability.

III. CONCLUSION

In this paper, a probe embedded with a hexagonal MEMS hot wire sensor chip for measuring wind speed and direction was developed. The designed system exhibits compact size, low cost, high precision and low power consumption. The anemometer probe was calibrated in a wind tunnel where a database was constructed and registered by a data acquisition system. This data acquisition system consists in an Arduino

platform connected with XBee Module transferring the measured data to a distant computer. With the constructed database, a model based on Artificial Neural Networks was elaborated and implemented on the Arduino platform. The elaborated model is charged to calculate the corresponding flow speed and direction from sensor readings with correcting the influence of change in air temperature. The prototype anemometer was tested outdoor and the obtained results are consistent with those registered by the classical anemometer over the entire measurement period. The future scope of this scheme may involve the design of a sensor network for large spatial measurement.

REFERENCES

- [1] M.Laghrouche, A.Adane, J.Boussey, S.Ameur, D.Meunier, S.Tardu, "A miniature silicon hot wire sensor for automatic wind speed measurements," *Renewable Energy* 30 (12), pp. 1881–1896, 2005.
- [2] M. T. Bensidhoum, M. Laghrouche, A. Sidi Said, L. Montes, J. Boussey, "Fabrication flaws and reliability in MEMS thin film polycrystalline flow sensor," *Microsyst Technol* 20, pp. 1-7, 2014.
- [3] S.Makhlouf, M.Laghrouche and A.E.H, "Hot wire sensor-based data acquisition system for controlling the laminar boundary layer near plant leaves within a greenhouse," *IEEE Sensor J.* Vol.16, Issue: 8, pp.2650- 2657, 2016.
- [4] S.Liu, Z. Yang, Y.Zhang, F.Xue,S.Pan, J.Miao, L.K.Norford "Micro triple-hot-wire anemometer on small sized glass tube fabricated in 5 DOF UV lithography system" Proceedings of the IEEE International Conference on Micro Electro Mechanical Systems (MEMS):714-717, February 2015, pp. 714-717, Estoril, Portugal, 2015.
- [5] H.B, Liu, N. Lin, S.S. Pan, J.M. Miao, and L.K. Norford, "High sensitivity, miniature, full 2-D anemometer based on MEMS hot-film sensors", *IEEE Sensors J.* vol.13, pp.1914-1920, 2013.
- [6] Z.Yang, S.Liu, F.Xue, Y.Zhang, X.Zhao, J.Miao and L.K.Norford, "Micro anemometer by a mems compatible lab-on-a-tube technology", 2015 Conference, DOI: 10.1109, *Transducers*. 2015.
- [7] R.Que, R.Zhu, "A Two-Dimensional Flow Sensor with Integrated Micro Thermal Sensing Elements and a Back Propagation Neural Network," *Sensors*, 14, pp. 564-574, 2014.
- [8] M. Laghrouche, B. Idjeri, K. Hammouche, M. Tahanout, J. Boussey, S. Ameur "Temperature compensation of micromachined silicon hot wire sensor using ANN technique", *Microsyst Technol* 18, pp. 237–246, 2012.
- [9] M. Pioletti, G. Pennelli, and P. Bruschi, "Fabrication and characterization of a directional anemometer based on a single chip MEMS flow sensor," *Microelectro. Eng.*, vol. 88, no. 8, pp. 2214–2217, 2011.
- [10] Y.Zhu, B.Chen, M. Qin, J.Huang and Q. Huang, "Development of a self packaged 2D MEMS thermal wind sensor for low power applications" *Journal of Micromechanics and Microengineering*.25, 85011, pp.1-14, 2015.
- [11] O.Yi, Q. Furong, W. Guanya, N. Mengyan, L. Zhigang, O. Wen, " A MEMS thermal shear stress sensor produced by a combination of substrate-free structures with anodic bonding technology" *Applied physics letters* 109: pp. 1-5, 2016.
- [12] I.H. Rusli, S. Aleksandrova, H. Medina, S.F. Benjamin "Using single-sensor hot-wire anemometry for velocity measurements in confined swirling flows", *Measurement* 129, pp. 277–280, 2018.
- [13] A. Ghahramani, M. Zhu, R.J. Przybyla, M.P. Andersen, P.J. Galicia, T.E. Peffer, H. Zhang, E. Arens, "Measuring Air Speed With a Low-Power MEMS Ultrasonic Anemometer via Adaptive Phase Tracking", *IEEE Sensor J.* Vol.19, Issue: 18, pp. 8136 - 8145, 2019.
- [14] P. Ligeza, "Modification of Hot-Wire Anemometers Frequency Bandwidth Measurement Method", *Sensors*, 20, pp.1-7, 2020
- [15] C. Hao, W. Zhang, B. Wu, Z. Zhang, J. He, R. Wang, C. Xue, "A novel two-dimensional high SNR MEMS shear stress sensor for ocean turbulence", *Sensors and Actuators A* 330, pp.1-8, 2021.
- [16] Lau, C. *Neural Networks: Theoretical Foundations and Analysis*; IEEE Press: Piscataway, NJ, USA, 1991.
- [17] B. Idjeri, M. Laghrouche, J. Boussey, "Wind Measurement Based on MEMS Micro-Anemometer With High Accuracy Using ANN Technique", *IEEE Sensor J.* Vol.17, Issue: 13, pp. 4181 - 4188, 2017.

Laser shaping by Digital Micromirrors for multiple targets tracking

Fares Karaoui¹, Hamida Bougherira^{1*}, Nadja Benblidia¹, Abderrezak Guessoum¹, Nadia Khetib¹, Djamila Naceur¹, Zoubir Benselama¹, Zoheir Meliani, Abdoulkarim Mamoudou Moumouni¹, Bouba Tangara¹

¹Laboratoire Latsi,
Faculté de Technologie
Département d'Electronique
Université de Blida1
Blida, Algeria

* Corresponding author: bougherirahamida@yahoo.fr

Abstract—Laser shaping by digital micromirrors is one of the various applications of Digital Micro mirror Devices (DMD's) since their fabrication for display purposes in the 80's. It is used in integrated circuit lithography process, high precision 3D printing, biomedical instrumentation, industrial fabrication processes and more. We have implemented an experimental optical system around the micro-opto-electromechanical system (MOEMS), the Texas Instruments Digital Micromirror Device DLP7000, to demonstrate laser beam shaping and positioning by the digital micromirrors, and discuss their possible use in multiple targets high precision tracking. Periodically, a camera captures and provides images of the targets to the image processing system in order to generate new binarized images to, adaptively, reshape the laser beam and reposition it as needed, according to the targets motion. To generate the initial laser beam and position the shaped beams on targets, an optical system has been implemented, and programs to command and communicate with the DMD and its controller, the VIALUX 7000, have been written using the ALP42 VIALUX library API modules. Laser beam shaping, by the B/W images displayed by the DMD to match the targets, has been successfully achieved, and demonstrate the feasibility and effectiveness of the proposed idea.

Keywords

Digital Micromirror Device—Multiple target Tracking—image processing—ALP42—V-Module

I. Introduction

Over the past decades, beyond the invention of digital micromirror devices for projection display in 1987 [1], new DMD applications have been continuously emerging.

In industry [2], DMD's are used for 3D printing [3][4], digital imaging lithography [5], laser marking, 3D scanning [6], allowing fast prototyping, additive manufacturing, high accuracy.

In medicine, DMD's are used in ophthalmology [6], phototherapy, surgery, or for direct manufacturing of biological tissues, organs, and prostheses [7][8], or for confocal microscopes[9], and many more [10].

In military laser beam shaping can be used for defense, and weapon development.

In many of these applications, laser beam shaping [11] is obtained by coupling the DMD to a laser beam, and an optical system. These technologies rely on the DMD ability to rapidly display new images in order to correct the needed shape until

satisfactory precision is reached. Any segmented and binarized image, displayed by the DMD, shapes an initial projected laser beam into two complementary laser beams. Thus if a segmented image of multiple objects spread over a wide field, is displayed on a DMD screen, the laser beam projected on the DMD, will be reflected into shaped laser rays, matching all the objects position in the scene, simultaneously, by an appropriate adaptive optical system [12][13][14]. No mechanical system nor targets pursuit algorithms are needed. On the other hand, efficient image processing for segmentation and binarization is required. The same reasoning holds for very small scenes like tracking unstable retina blood vessels during photocoagulation process [15]. Our work aims to implement a laboratory experiment to demonstrate laser beam shaping and positioning by the digital micromirrors, and discuss their possible use in multiple targets high precision tracking. We first describe briefly the DMD properties relevant to our work. Then we present our strategy and the experiment set up. Next we show and discuss our experimental results.

II. DMD mechanism

The (dlp7000) is a Texas Instruments DMD circuit that has been initially designed in the 80's for projection display purposes. We describe in this section the DMD properties that will be used in our experiment.

A. DMD description

The DLP7000 is an array of 1024x768 micromirrors mounted on the top of a CMOS RAM cells with other electrical and mechanical elements. An inactive micromirror is in a flat position, and rotates by +12° if it is in the ON or -12° if it is in the OFF state (Fig. 1.). Each mirror pixel of a DMD rotates with respect to the main diagonal, and is addressed individually. Its angular position is determined by the binary state (logic 0 or 1) of the corresponding CMOS memory cell content [2]. Grey level colors are obtained by switching the micromirrors between their on and off states at appropriate frequency rates. The white and black color pixels correspond to the on state, and off state respectively.

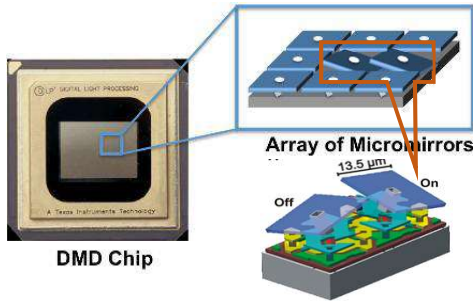


Fig. 1. DMD chip and micromirrors

B. DMD control

A typical application of the DLP7000 requires the use of a high speed V-Module (from VIALUX) to control the display of an image on the digital micromirrors (Fig. 2.). Vialux provides Texas Instrument DMD chips with an electronic board to send and display image sequences at high speed. The V-Modules provide complete subsystems containing a controller board, a DMD, and an ALP-4 controller suite software [16]. The DLP user software uses the ALP-4 library functions to allocate, deallocate and communicate with the DMD.

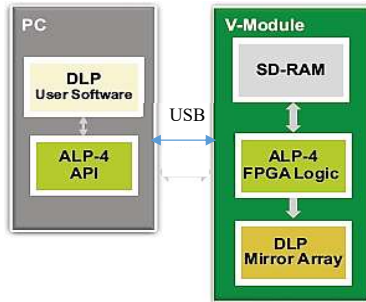


Fig. 2. Bloc diagram of the ALP Controller Suite: scheme of control and data flow [16]

III. LASER SHAPING EXPERIMENT SETUP

A. Strategy

It is based on black and white (B/W) image display by the DMD, via the V-Module DMD controller. The initial laser beam is projected perpendicularly onto the DMD screen. The mirrors corresponding to white pixels rotate at $+12^\circ$ and reflect that part of the laser beam at $+24^\circ$, while the mirrors corresponding to the image black pixels rotate at -12° , and reflect the other part of the laser beam in the -24° direction. Thus, the initial laser beam is split into a negative, and a positive part. The shape is determined by the B/W image, and the shape size, by the optical system placed in the direction of the shaped laser beam. A camera attached to the optical system, periodically, captures successive images of the targets scene, and transmits them to the computer for image processing. The program generates a new B/W high resolution image, where the targets are represented in one of the two colors, and located with a high precision (fig. 3.). If the targets move in the next captured image, a new B/W image of the targets is displayed by the DMD, generating a new shaped laser beam to instantly reposition it on the targets, accurately, in an adaptive manner. Thus, no power

electromechanical system is needed, allowing fast and precise multiple targets tracking.

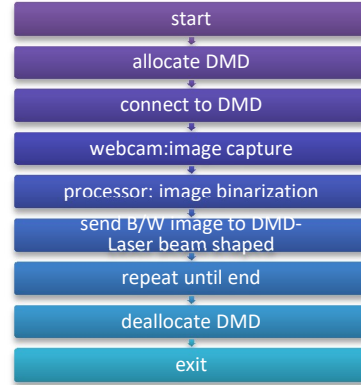


Fig. 3. DMD control: strategy laser shaping steps

B. Experiment setup

- Collimated laser beam

To obtain an incident parallel laser beam on the DMD, a diode laser source is placed one focal length from the lens L1 (fig. 4.). The collimated beam diameter D (the lens clear aperture) is chosen to be greater than the size of the B/W image to be displayed on the DLP7000 screen.

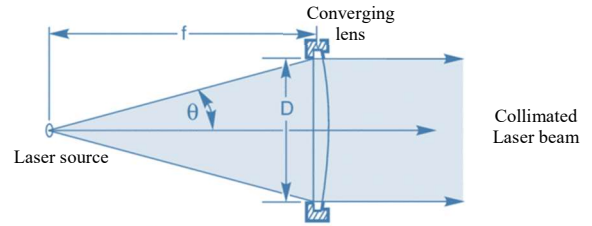


Fig. 4. Collimated laser beam generation

- The optical system:

The collimated laser beam, coming out of lens L1, faces the DLP7000 perpendicularly so that the fashioned beams are reflected $+24^\circ$ left and -24° right by black pixel, and white pixel DMD mirrors, toward targets1, and targets2, respectively. Fig. 5 shows the optic schematic diagram, and fig.6 shows the implemented system.

- Camera

The camera is a Webcam set to face the targets we want to track; in our case it is set in the direction of the fashioned laser beam reflected by the white pixels. It is programmed to capture and send images to the PC, at a variable, user defined rate

- Computer

The computer (PC) processes the images it receives iteratively, in real-time, to generate B/W images where the targets are represented in white. It sends them to the DLP via the V-Module controller to be displayed. Images are processed using image processing algorithms, and transmitted to the DMD using the ALP-42 library functions. The ALP function library encapsulates all functionality required to

control the pattern sequence data loading and the high-speed display of the sequences. It communicates with the ALP driver, to transfer image sequence data via the USB interface into the on-board SRDAM, or to read and write control registers in the FPGA controller section [21]. Finally, the FPGA logic controls the communication with the DMD mirrors. it is necessary to develop personal high level language software to specify the rate, the number of images, and the images to send to the DMD

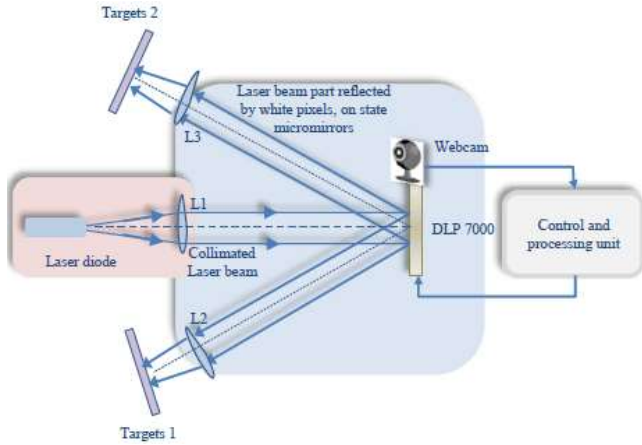


Fig. 5. optical schematic diagram

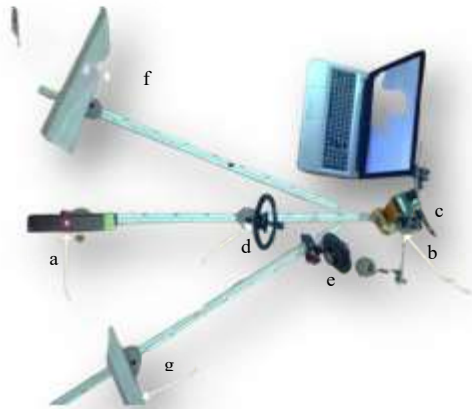


Fig. 6. experiment set up: a-laser source, b-DMD, c-V-module, d-lens.L1, e- lens.L2, f- positive targets, g- negative targets, h- PC control unit

- Image processing

The image captured by the camera and transmitted to the computer at time t , is segmented then displayed, to automatically update positioning of the laser rays on the targets. No comparison with image at time $t-1$ is necessary.

IV. RESULTS AND DISCUSSION

A. Results

The DMD displays a B/W image of the targets (fig. 9a.), The laser beam reflected by these white pixels has the shape of the targets (even if they have moved the ones with respect to the others). Fig. 9.b shows that the shaped laser beam is not

properly projected on the targets. After the processing of this last image, a new B/W image displayed by the DMD, reflects the laser beam and projects it exactly on the targets (Fig. 9.c).

Fig. 7 and Fig. 8 show the functionality of our software to address and control the DMD.

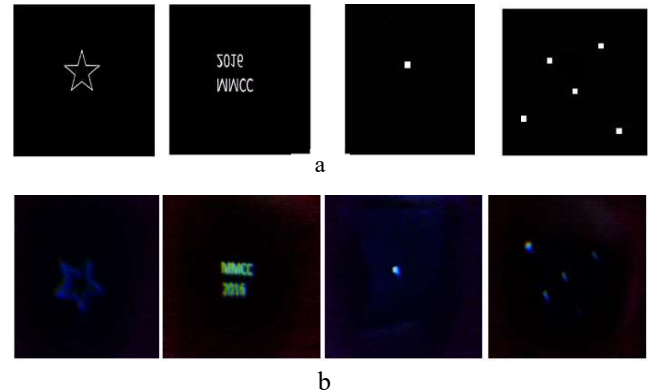


Fig. 7 a- B/W images transmitted to the DMD, b- reflected shaped beam

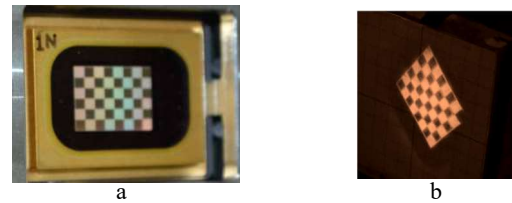


Fig. 8 a- B/W image displayed by the DLP, b- shaped laser beam reflected on positive target

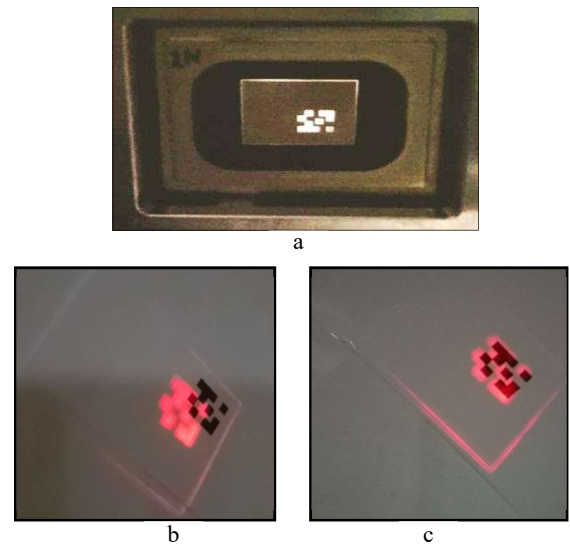


Fig. 9-The DMD displays a B/W image, b- targets and shaped laser beam do not match, c-the laser rays position is corrected by the micromirrors

B. Discussion

In our personal experiments we focused on the DMD capability to create a laser beam that matches the targets positions, by displaying binarized images on the DMD; we reached our goal, by implementing a simple laboratory experiment. Two multiple targets tracking issues were solved

at once by the use of the DMD laser beam shaping: the first one is the fast simultaneous positioning of the shaped beam on the targets, due to the high switching speed of the micromirrors, and the high resolution of the displayed image; the second one is the fact that no algorithm for object detection is needed, since the binarized image is sufficient to detect, and locate the targets. However, depending on the application, initial incidence laser beam angle, laser beam top shape, laser beam path, various optics aberrations, diffractions by the micromirrors, and laser power compatibility with the DMD material, might need to be considered for correct beam shaping and projection.

V. CONCLUSION

Images displayed by the DMD micromirrors shape an input laser beam with high accuracy because of the DMD micromirrors high resolution. We have used a simple optical system, combined with the DMD, to project the shaped output beam on the targets. To correct the positioning of the shaped laser beam, displaying a new binarized image of the targets is sufficient, and neither detection nor recognition algorithm is a priori needed. Adaptive optical technologies are already providing a variety of increased capabilities for lasers. Depending on the application, optical systems using adaptive optics models, can be developed to track either close or far targets, by expanding, narrowing, or correcting the path of the shaped laser beam, and to provide the needed laser rays power. Typically a closed loop system including a camera and an iterative efficient image processing algorithm for segmentation and binarization, coupled to the DMD and the appropriate optical system, will allow high speed, high precision multiple targets tracking, thanks to the DMD's resolution and micromirrors switching speed. Such systems could be embedded on cars, planes, medical, military or industrial instruments, or on connected MEMS (IOT's).

REFERENCES

- [1] L. J. Hornbeck: Digital light processing for high-brightness, high-resolution applications, Proc. SPIE, 3013 (1997) 27-41J. M. Smith and A. B. Jones. Book Title. Publisher, 7th edition, 2012.
- [2] Texas Instruments. DLP7000 DLP® 0.7 XGA 2x LVDS Type A DMD datasheet, august 2012 [Revised June. 2019]
- [3] Mott, Eric & Busso, Mallory & Luo, Xinyi & Dolder, Courtney & Wang, Martha & Fisher, John & Dean, David. (2015). Digital Micromirror Device (DMD)-based 3D printing of poly(propylene fumarate) scaffolds. *Journal of Materials Science and Engineering: C*. 61. 2015.11.071.
- [4] Jiumeng Zhang , Qipeng Hu , Shuai Wang , Jie Tao , Maling Gou Digital Light Processing Based Three-dimensional Printing for Medical Applications
- [5] R.C.Y. Auyeung · H. Kim · N.A. Charipar · A.J. Birnbaum · S.A. Mathews · A. Piqué. Laser forward transfer based on a spatial light modulator *Appl Phys A* (2011) 102: 21–26
- [6] Kari V. Vienola , Mathi Damodaran, Boy Braaf, Koenraad A. Vermeer, Andjohannes F. De Boer Parallel line scanning ophthalmoscope for retinal imaging Vol. 40, No. 22 / November 15 2015 / *Optics Letters*
- [7] S.P. Grogan, P.H. Chung, P. Soman, P. Chen, M.K. Lotz, S. Chen, D.D. D’Lima, Digital micromirror device projection printing system for meniscus tissue engineering, *Acta Biomaterialia* 9 (2013) 7218–7226.
- [8] J. Ni, H. Ling, S. Zhang, Z. Wang, Z. Peng, C. Benyshek, R. Zan, A.K. Miri, Z. Li, X. Zhang, J. Lee, K.-J. Lee, H.-J. Kim, P. Tebon, T. Hoffman, M.R. Dokmeci, N. Ashammakhi, X. Li, A. Khademhosseini, Three-dimensional printing of metals for biomedical applications, *Materials Today Bio*, Volume 3, 2019, 100024, ISSN 2590-0064.
- [9] *Adaptive Optics and Wavefront Control for Biological Systems II*, edited by Thomas G. Bifano, Joel Kubby, Sylvain Gigan, Proc. of SPIE Vol. 9717, 971706 · © 2016 SPIE · CCC code: 1605-7422/16/\$18
- [10] Taerim Yoon, Chang-Seok Kim, Kyujung Kim, Jong-ryul Choi, Emerging applications of digital micromirror devices in biophotonic fields, *Optics & Laser Technology*, Volume 104, 2018, Pages 17-25, ISSN 0030-3992.
- [11] P. Delaporte, D. Karnakis, I. Zergioti, Beam Shaping - an overview *Laser processing of flexible organic electronic ScienceDirect Topics materials*, Editor(s): Stergios Logothetidis, *Handbook of Flexible Organic Electronics*, Woodhead Publishing, 2015, Pages 285-313, ISBN 9781782420354.
- [12] Neumann, J. G. *DMD based hyperspectral augmentation for multi-object tracking systems. Emerging Digital Micromirror Device Based Systems and Applications*. (2009).
- [13] Zhu Han, Zhang Zi-li, Tang Ya-jun, Guo Xi-qing, Zhou Wei-hu. Design of collaborative target tracking system using digital micro-mirror device[J]. *Journal of Applied Optics*, 2014, 35(5): 750-755. Shu
- [14] Anton Travinskya, and al Evaluation of Digital Micromirror Devices for use in space-based Multi-object Spectrometer application arXiv:1708.06241v1 [astro-ph.IM] 21 Aug 2017
- [15] Vienola, Kari & Damodaran, Mathi & Braaf, Boy & Vermeer, Koenraad & Boer, Johannes. (2018). In vivo retinal imaging for fixational eye motion detection using a high-speed digital micromirror device (DMD)-based ophthalmoscope. *Biomedical Optics Express*. 9. 591. 10.1364/BOE.9.000591.
- [16] Hoefling, R., & Ahl, E. *ALP: universal DMD controller for metrology and testing. Liquid Crystal Materials, Devices, and Applications X and Projection Displays*(2004).

A Novel Image Watermarking Scheme In DWT Domain Based on Discrete-Time Chaotic Systems Of Fractional and Integer Orders

Katia Hannoun*, Ouerdia Megherbi*, Hamid Hamiche*, Mourad Lahdir**, Mourad Laghrouche**

*Laboratoire de Conception et de Conduite des Systèmes de Production, UMMTO Tizi-Ouzou, Algeria

** Laboratoire d'Analyse & Modélisation des Phénomènes Aléatoires, UMMTO Tizi-Ouzou, Algeria

hannounkb@gmail.com

hamid.hamiche07@gmail.com

meg_ouer@yahoo.fr

mlahdir@yahoo.fr

larouche_67@yahoo.fr

Abstract: In this paper, we present a novel robust image watermarking scheme in the DWT domain based on two discrete-time chaotic systems. Two modified-Hénon maps of integer and fractional orders are considered as the transmitter of this scheme, whereas a couple of step-by-step observers of integer and fractional orders are considered as the receiver. This transmission scheme allows an exact synchronization and in consequence an exact recovery of the system's states.

The encrypted image is inserted by inclusion method in the fractional order discrete-time chaotic system's dynamics; the resulting cipher image is used as a watermark to a host image. This host image is watermarked by a DWT watermarking function and then at the receiver level will be extracted and decrypted. Simulation results are presented to highlight the feasibility and robustness of the proposed method.

Keywords: chaotic systems; fractional order systems; watermarking; DWT domain; modified-Hénon map; step by step observer.

I. INTRODUCTION

Preventing unauthorized access or use of data in order to keep the information safe is a critical issue that is faced daily. Digital imaging holds several benefits, thus making the demand for its applications astonishing. However, keeping information safe by preventing unauthorized access or use of data is a daily ordeal. Effective security tools need to be put in place to insure the safety of important information in the images making the transmission applications truly confidential. And with the several benefits of digital imaging, the demand for its applications is overwhelming. Confidential data transmission applications have shown the need for reliable and effective security tools to protect important information in the images from attackers. Those security tools include several techniques, among which is digital watermarking [1], [2] and chaotic encryption [3]-[8]. The attention of cryptographers was drawn by the advantages of chaotic systems (Sensitivity to initial conditions, ergodicity, non-periodicity, non-convergence) to design robust encryption techniques. Plenty of chaotic systems can be generated with small perturbation of parameters. A small variation of parameters allows to generate a multitude of chaotic systems. The chaotic signal can be reproduced easily

by keeping the initial conditions and the chaotic parameters as the secret keys. For these reasons, chaos based encryption systems are considered a center of attention for researchers [9]. Systems of encryption based on chaos, encrypt the message by a signal generated by a nonlinear dynamical chaotic system (transmitter), generating an encrypted information, which is transmitted to the receiver through the communication channel.

When the synchronization transmitter-receiver occurs, the message is recovered. As shown in [7], [8], synchronization of chaotic systems is equivalent to an observer design problem. If so, the state and the transmitted message considered as an unknown input ought to be estimated. When it comes to security of information, copyright protection is also an important topic. For this reason, digital watermarking is an effective method for copyright protection and proof of ownership [10]. The watermarking techniques can be classified in two vast categories [11]:

Spatial domain digital watermarking: Methods to embed the watermark by directly changing the intensity of certain pixels.

Transform domain digital watermarking: Methods to operate on particular coefficients of a specifically selected transform domain (DCT domain, DFT domain, etc.).

In this paper, we propose a novel robust watermarking scheme in the DWT domain based on a fractional-order discrete-time chaotic system. The advantage of using such systems is explained in [3], [7], [8]. An original image is encrypted by two Modified-Hénon maps: integer and fractional orders discrete-time chaotic systems, in order to watermark it in a host image in the DWT domain [12]. The original image is recovered using a transmission scheme based on a synchronization process of discrete-time chaotic systems [3], [7], [8]. The synchronization of the transmitter and the receiver is established between a fractional-order discrete-time chaotic system and a fractional-order discrete-time observer to recover the watermark. The inclusion method, which has shown its reliability as proven in [5], is used to inject the message in the dynamics of the state variable that is not transmitted in order to increase the robustness of the proposed transmission scheme against malicious attacks.

The main contribution of this paper is to combine fractional-order chaotic system with a watermarking scheme, in order to make it more robust and secure.

The use of a hybrid chaotic system consisting of two chaotic systems of different orders (fractional and integer) is also one of the contributions of this paper.

The remaining of the paper is organized as follows. In Section 2, the proposed transmission scheme is introduced and different parts of both the transmitter and the receiver are reviewed. In section 3, the simulation results illustrating the synchronization and the reconstruction of the transmitted message are illustrated. Section 4 provides the security analysis. Section 5 contains concluding remarks and some perspectives to improve the proposed scheme.

II. DESCRIPTION OF THE METHOD

The fractional-order discrete-time chaotic system of modified-Hénon map is used to design a communication system which is composed of two blocks: transmitter and receiver/observer that will be used to:

Step 1: Create an encryption function using integer-order discrete-time chaotic system.

Step 2: Encrypt an original image at the encryption function level, then insert it by inclusion method in the fractional order discrete-time chaotic system's dynamics. The chaotic sequence resulting from this fractional order system is used as the watermark.

Step 3: Insert the chaos based watermark in a host image by a DWT domain watermarking algorithm.

Step 4: Extract the chaos based watermark by a DWT domain watermarking algorithm.

Step 5: Decrypt the watermark at the level of two observers.

The general scheme for the communication system and the watermarking process is shown below:

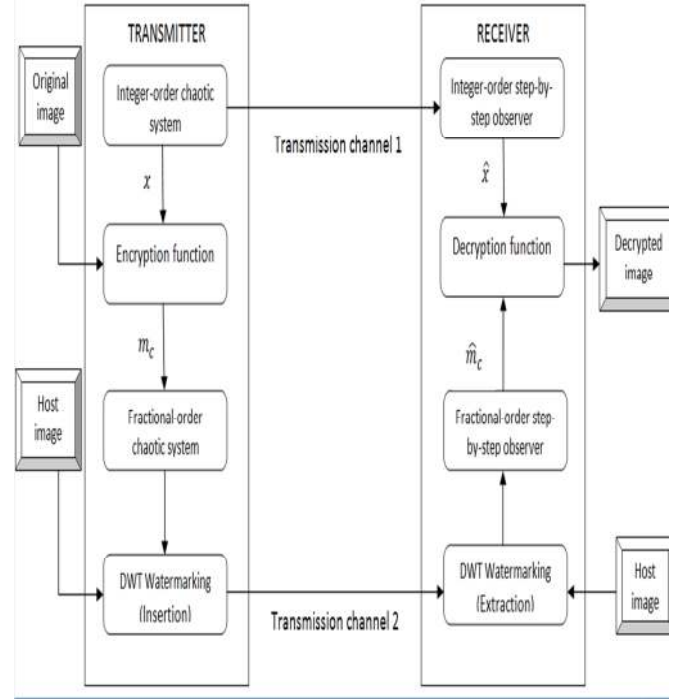


Fig1. Scheme of the proposed system.

A. Transmitter presentation

In this section, both modified-Hénon chaotic systems (integer and fractional orders) are presented.

1) *Integer-order modified-Hénon map:* Consider the following discrete-time modified-Hénon map, given in [8]:

$$\begin{cases} x_{11}(k+1) = a - x_{22}^2(k) - bx_{33}(k) \\ x_{22}(k+1) = x_{11}(k) \\ x_{33}(k+1) = x_{22}(k) \\ y(k) = x_{22}(k) \end{cases} \quad (1)$$

Where $x = [x_{11}, x_{22}, x_{33}]^T \in \mathbb{R}^3$ represents the state vector and $y(k) = x_{22}(k)$ is the output. The chaotic behavior is exhibited for $a = 1.76$, $b = 0.1$ and the initial conditions chosen inside the basin of attraction are: $x_{11}(0) = 0.1$, $x_{22}(0) = 0.1$, $x_{33}(0) = 0.1$.

As given below, simulation results illustrate the chaotic behavior of System (1). The phase portrait of the states (x_{11}, x_{33}) is presented in Fig.2 and the states response in Fig.3.

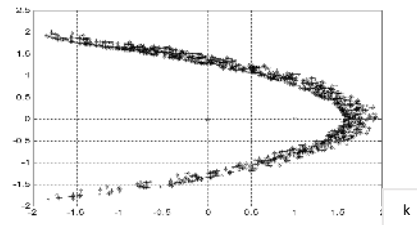


Fig.2. Phase portrait of the states $(x_{33}(k)- x_{11}(k))$ of the modified-Hénon map.

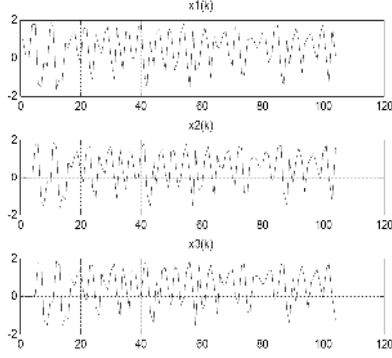


Fig.3: State responses of the modified-Hénon system.

At this stage, the original image is encrypted by creating a key through an encryption function (XOR operation) using the integer order chaotic system's states. The resulting encrypted image is then inserted by inclusion method in the third component of the fractional-order chaotic system to improve the encryption.

2) *Fractional-order modified-Hénon map: Starting from the previous integer order modified-Hénon system (1)*

To introduce the fractional discrete derivatives, the Grünwald-Letnikov formula is used [13]:

Definition1: The α -order Grünwald-Letnikov difference for a state vector x is given as follows:

$$\Delta^\alpha x(k) = \sum_{j=0}^k (-1)^j \binom{\alpha}{j} x(k-j) \quad (2)$$

Where $\alpha > 0$ denotes the fractional order, $x(k) = [x_1(k) \ x_2(k) \ \dots \ x_n(k)] \in \mathbb{R}^n$ is the state vector at the current time and $x(k-j)$ is the delayed state vector.

The term $\binom{\alpha}{j}$ can be computed from the following relation:

$$\binom{\alpha}{j} = \begin{cases} 1 & \text{for } j = 0 \\ \frac{\alpha(\alpha-1)\dots(\alpha-j+1)}{j!} & \text{for } j > 0 \end{cases} \quad (3)$$

In order to represent a fractional-order discrete-time system, let us consider the following integer discrete-time system:

$$\begin{aligned} x(k+1) &= f(x(k)) \\ y(k) &= h(x(k)) \end{aligned} \quad (4)$$

Where $y(k)$ is the output and f is a smooth nonlinear function. We denote the 1-order difference for the discrete time system (4) as:

$$\Delta^1 x(k+1) = x(k+1) - x(k) = f(x(k)) - x(k) \quad (5)$$

From (4), we define the α -order difference as:

$$\Delta^\alpha x(k+1) = f(x(k)) - x(k) \quad (6)$$

Noting the α -difference (1), we obtain:

$$\begin{aligned} \Delta^\alpha x(k+1) &= f(x(k)) - \alpha x(k) \\ &+ \sum_{j=2}^{k+1} (-1)^j \binom{\alpha}{j} x(k-j+1) \end{aligned} \quad (7)$$

By introducing a new variable $p = j - 1$, the previous relation becomes:

$$\begin{aligned} \Delta^\alpha x(k+1) &= x(k+1) - \alpha x(k) \\ &+ \sum_{p=1}^k (-1)^{p+1} \binom{\alpha}{p+1} x(k-p) \end{aligned} \quad (8)$$

Let us define the parameter $C_p = (-1)^{p+1} \binom{\alpha}{p+1}$. By replacing (8) in (6), we obtain the following relation:

$$x(k+1) = f(x(k)) + (\alpha - 1)x(k) - \sum_{p=1}^k C_p x(k-p) \quad (9)$$

By applying a truncation to the infinite memory of the fractional-order chaotic system (9), we obtain:

$$x(k+1) = f(x(k)) + (\alpha - 1)x(k) - \sum_{p=1}^L C_p x(k-p) \quad (10)$$

Remark 1. Note that System (10) is obtained in the commensurate case since the differentiation order α is chosen the same for all the state variables [18].

However, this is not always the case. To deal with the non-commensurate case, we should consider all the differentiation orders $\alpha_1, \alpha_2, \dots, \alpha_n$.

According to System (10), the corresponding fractional-order discrete-time system of (1) is:

$$\begin{aligned} x_1(k+1) &= a - x_2^2(k) - bx_3(k) + (\alpha_1 - 1)x_1(k) + \beta_1(x_1(k)) \\ x_2(k+1) &= x_1(k) + (\alpha_2 - 1)x_2(k) + \beta_2(x_2(k)) \\ x_3(k+1) &= x_2(k) + (\alpha_3 - 1)x_3(k) + \beta_3(x_3(k)) \\ y(k) &= x_2(k) \end{aligned} \quad (11)$$

Where $\beta_1 = -\sum_{j=1}^L C_{1j} x_1(k-j)$; $\beta_2 = -\sum_{j=1}^L C_{2j} x_2(k-j)$; $\beta_3 = -\sum_{j=1}^L C_{3j} x_3(k-j)$, $0 < \alpha_1 \leq 1$, $0 < \alpha_2 \leq 1$, $0 < \alpha_3 \leq 1$ are the fractional orders.

The chaotic behaviour is exhibited for $a = 1.6$, $b = 0.1$, and $\alpha_1 = 0.85$, $\alpha_2 = 0.9$, $\alpha_3 = 0.75$. The initial conditions chosen inside the basin of attraction are: $x_1(0) = -0.1$, $x_2(0) = 0.2$, $x_3(0) = 0.1$.

To preserve the chaotic behavior, the original image is introduced in the third dynamic component of system (11).

As given below, simulation results illustrate the chaotic behavior of System (11). The phase portrait of the message (encrypted image) is presented in Fig.5.

Then, we obtain:

$$\begin{aligned}
x_1(k+1) &= a - x_2^2(k) - bx_3(k) + (\alpha_1 - 1)x_1(k) + \beta_1(x_1(k)) \\
x_2(k+1) &= x_1(k) + (\alpha_2 - 1)x_2(k) + \beta_2(x_2(k)) \\
x_3(k+1) &= x_2(k) + (\alpha_3 - 1)x_3(k) + \beta_3(x_3(k)) + m_c(k) \\
y(k) &= x_2(k)
\end{aligned} \tag{12}$$

Where $m_c(k)$ represents the resulting encrypted image from the previous XOR operation, and $y(k) = x_2(k)$ is the output.

The message m_c (encrypted watermark) is introduced in the third component of the system (11) and the state variable $x_2(k)$ considered as the system output is, in fact, the fractional-order chaotic sequence converted into an image, then inserted in a host image by applying a DWT watermarking insertion scheme. A single-level 2-D wavelet decomposition (DWT) is done with respect to a Haar filter to embed the cipher image (encrypted watermark) in a host image [12].

This DWT domain algorithm is summarized as follows [14]:

Decompose the host image and the cipher image into four levels (LL, LH, HL, HH) and (L-L, L-H, H-L, H-H), respectively. These levels represent the frequency districts, with LL being the low frequency and HL, LH and HH being the high frequencies.

Apply the DWT to both the images on the sub-band LL and L-L, leading to (LL1, LH1, HL1, HH1) and (L-L1, L-H1, H-L1, H-H1), respectively.

Apply the DWT to both the images again on LL1 an L-L1, leading to (LL2, LH2, HL2, HH2) and (L-L2, L-H2, H-L2, H-H2), respectively.

Embed the watermark image considering the LL sub-band of the DWT using the formula $LL + 0.01 \times L_L$

Apply the IDWT (Inverse Discrete Wavelet Transform) using the new sub-band values to obtain the watermarked image.

Once the watermark is inserted, it is considered as the system's output, transmitted to the receiver through the transmission channel, in order to extract the watermark as in the previous process.

The same process is applied to extract the watermark.

B. Receiver presentation

In this part, the watermark is extracted following the same process as in the insertion method:

Decompose the watermarked image into four levels (a, b, c, d). These levels represent the frequency districts, with d being the low frequency and c, b and a being the high frequencies.

Apply the DWT to the watermarked image on the sub-band a leading to (aa, bb, cc, dd), then again on aa leading to (aaa, bbb, ccc, ddd)

Extract the watermark using the formula $\frac{a-LL}{0.01}$.

This process is called a non-blind watermark extraction, for it requires that the host image is part of the receiver's entries, as shown in Fig.1.

Once the watermark is well extracted, it is then decrypted at the level of two observers to obtain the original image.

Two discrete delayed step-by-step observers are designed. The fractional-order step-by-step observer perfectly recovers the states and the encrypted image [7], [8]. The integer-order discrete step-by-step observer reconstructs the key in order to decrypt the said image.

Quite a few results of the observability matching condition and left invertibility property for using delayed discrete observer are provided in [13]. The step-by-step method starts by applying one step delay on the output to recover the first starting system state. The second step is to apply two steps delay on the output to recover the second state. The delays are applied on all the system states until the last information containing the system's input. Each recovered state in k contributes to the recovery of the state in $k-1$ as given in [7], [8].

1) Fractional-order step-by-step observer: State reconstruction:

- Reconstruction of state $\widehat{x}_1(k)$:

$$\widehat{x}_1(k) = x_{10}(k)$$

From the second equation of System (1), and by applying one step delay, we deduce the state $\widehat{x}_{11}(k) = x_{110}(k)$ as follows:

$$\begin{aligned}
x_{10}(k-1) &= \widehat{x}_1(k-1) \\
&= y(k) - (\alpha_2 - 1)y(k-1) - \beta_2(y(k-1))
\end{aligned} \tag{13}$$

With $\beta_2 = -\sum_{j=1}^L C_{2j}x_2(k-j)$, then:

$$\beta_2(x_{20}(k-1)) = -\sum_{j=1}^L C_{2j}x_{20}(k-j-1)$$

- Reconstruction of state $\widehat{x}_3(k)$:

By applying two steps delay to the first equation of System (12), we obtain:

$$\begin{aligned}
x_{30}(k-2) &= \widehat{x}_3(k-2) \\
&= \frac{a - \widehat{x}_{10}(k-1) - \widehat{y}^2(k-2) + (\alpha_1 - 1)\widehat{x}_{10}(k-2) + \beta_1(\widehat{x}_{10}(k-2))}{b}
\end{aligned} \tag{14}$$

With $\beta_1 = -\sum_{j=1}^L C_{1j}x_1(k-j)$, then:

$$\beta_1(x_{10}(k-2)) = -\sum_{j=1}^L C_{1j}x_{10}(k-j-2)$$

Where the state $\widehat{x}_1(k)$ is given in (13).

Reconstruction of the message $\widehat{m}_c(k)$:

$$\widehat{m}_c(k) = m_{c0}(k)$$

From the equation of System (12):

$$\begin{aligned} m_{c0}(k-3) &= \widehat{m}_c(k-3) \\ &= \widehat{x}_3(k-2) - \widehat{x}_2(k-3) - (\alpha_3 - 1)\widehat{x}_3(k-3) - \beta_3(\widehat{x}_3(k-3)) \end{aligned} \quad (15)$$

$$\beta_3 = - \sum_{j=1}^L C_{3j} x_3(k-j), \text{ then:}$$

$$\beta_3(x_{30}(k-3)) = - \sum_{j=1}^L C_{3j} x_{30}(k-j-3)$$

Where the state $x_{30}(k)$ is given in (14).

At this point, the cipher image is recovered. Its decryption is done at the level of the second observer (integer-order), where the key is to be reconstructed.

2) *Integer-order discrete step-by-step observer: To reconstruct all the states of System (1), the same observer chosen previously is used:*

- Reconstruction of state $\widehat{x}_{11}(k)$:

$$\widehat{x}_{11}(k) = x_{110}(k)$$

From the second equation of System (12), and by applying one step delay, we deduce the state $\widehat{x}_1(k) = x_{10}(k)$ as follows:

$$\widehat{x}_{11}(k-1) = y(k) \quad (16)$$

Reconstruction of state $\widehat{x}_{33}(k)$:

$$\widehat{x}_{33}(k) = x_{330}(k)$$

From System (2):

$$\begin{aligned} \widehat{x}_{33}(k) &= \frac{a - \widehat{x}_{11}(k+1) - \widehat{x}_{22}^2(k)}{b} \\ x_{330} &= \widehat{x}_{33}(k) \end{aligned}$$

By applying two steps delay to the first equation of System (1), we deduce the state $\widehat{x}_{33}(k)$ as follows:

$$\widehat{x}_{33}(k-2) = \frac{a - y(k) - y^2(k-2)}{b} \quad (17)$$

Finally, the message (original image) is recovered by applying the decryption function, as represented in Fig.1.

III. SIMULATION RESULTS

In this section, some simulation results of the proposed image cryptosystem are given. First, the simulation results on the synchronization of the integer-order and fractional-order chaotic systems in the transmitter and their respective observers in the receiver are presented. Then, the simulation results on the watermarking scheme are illustrated.

A. Simulation results on the synchronization

1) *Integer-order discrete-time chaotic system: In the following, we present the simulation results for the*

synchronization of the integer-order discrete-time chaotic system (1) and its observer given by the equations (16) and (17). The original image to encrypt is the standard Peppers test image of size 256×256 pixels, and the host image is the standard Lena test image of size 512×512 .

As shown below, fig.4 illustrates the phase plane of x_{11} versus \widehat{x}_{11} and x_{33} versus \widehat{x}_{33} . The reconstruction of these states is done step by step and is perfect.

2) *Fractional-order discrete-time chaotic system: In the following, we present the simulation results for the synchronization of the fractional-order discrete-time chaotic system (11) and its observer given by (13), (14) and (15). The original image to encrypt is the standard Peppers test image of size 256×256 pixels, and the host image is the standard Lena test image of size 512×512 .*

As shown below, fig.5 illustrates the phase plane of the message m_c versus \widehat{m}_c . The reconstruction of these states is done step by step and is perfect.

The straight lines displayed on fig.4 and fig.5 confirm that the synchronization is well established.

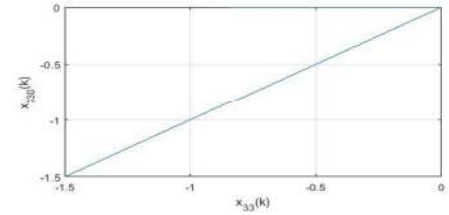
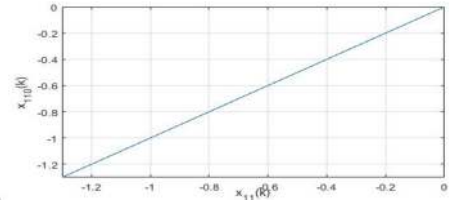


Fig.4: Phase plane of x_{11} versus \widehat{x}_{11} , x_{33} versus \widehat{x}_{33}

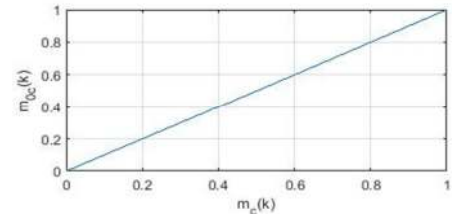


Fig.5: Phase plane of m_c versus \widehat{m}_c

B. Simulation results on the watermarking scheme

Fig.(6.a) and Fig.(6.b) show the original image and its corresponding histogram. Fig.7 shows the encrypted Peppers image and its corresponding histogram. Fig.8 shows the host image, the encrypted watermark and the watermarked image. Fig.9 shows the extracted watermark and Fig.10 the decrypted Peppers image and its corresponding histogram. Comparing Fig (6.a) and (6.b) to Fig (10.a) and (10.b), the distribution of histograms for the decrypted watermark Peppers image presented in Fig.(10.a) is approximately the same as the original Peppers image presented in Fig (6.a).

These results indicate that the watermarked image is secure with this encryption scheme and the original image (watermark) is recovered perfectly using our encryption scheme.

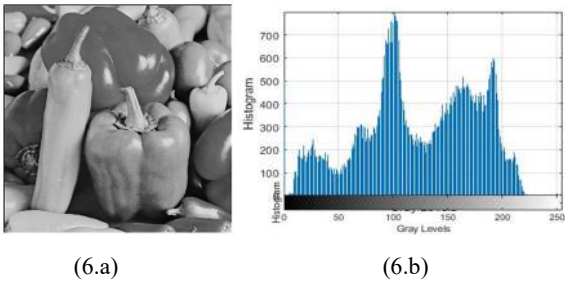


Fig.6: Original Peppers image and its corresponding histogram.

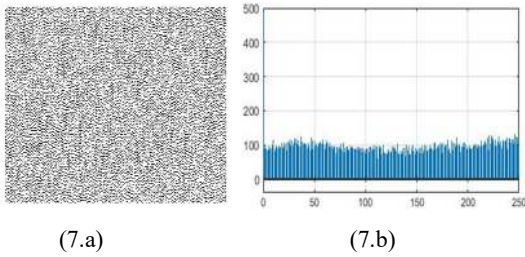


Fig.7: Encrypted Peppers image and its corresponding histogram.

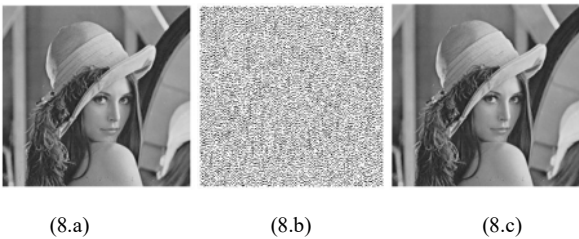


Fig.8: Host image, encrypted watermark and watermarked image.

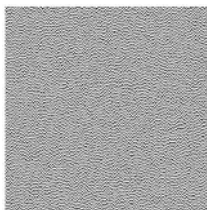


Fig.9: Extracted watermark.

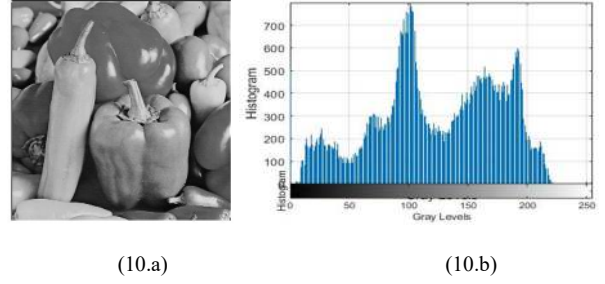


Fig.10: Decrypted Peppers image and its corresponding histogram.

IV. ANALYSIS OF THE PROPOSED TRANSMISSION SCHEME

In this section, the robustness of the proposed transmission scheme is tested. To do this, some analysis is presented:

Statistical Analysis [7]: Adjacent pixels with high correlation are one of the intrinsic characteristics of digital images without compression. To test the correlation between the gray scale values of two adjacent pixels, the correlation coefficient in each direction is calculated by:

$$cov(x, y) = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)(\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2)}} \quad (18)$$

Where $\bar{x} = \frac{1}{N} \sum_{i=1}^N (x_i)$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N (y_i)$, (x_i, y_i) represents the i^{th} pair of adjacent pixels in the same direction, N is the total number of pixels pairs.

TABLE I: CORRELATION COEFFICIENTS IN THE ORIGINAL AND ENCRYPTED IMAGES.

	Original Image	Encrypted Image
Horizontal	0.9549	0.0336
Vertical	0.9671	-0.0208
Diagonal	0.9116	-0.0117

The results of the correlation coefficients are presented in Table I, which are far apart. Therefore, the proposed algorithm has high security against statistical attacks.

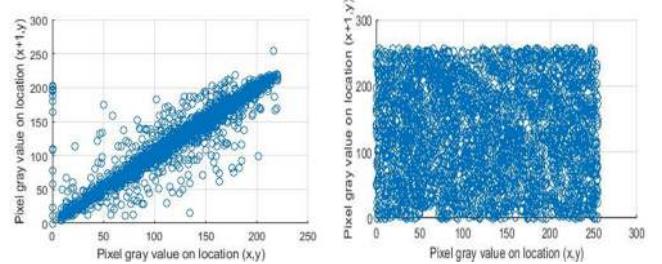


Fig.11: Correlations of two horizontal adjacent pixels in the original image and encrypted image.

Fig.11 shows the correlation distribution of two horizontally adjacent pixels in the plain image and that in the encrypted image.

V. CONCLUSION

In this paper, we have proposed a novel robust image DWT domain watermarking scheme using two chaotic discrete-time

systems: Integer-order discrete-time chaotic system and Fractional-order discrete-time chaotic system. First, the message (Peppers image) was encrypted, at the encryption function level, using states of an integer-order discrete-time modified-Hénon chaotic map. The encrypted Peppers image was then inserted, by inclusion method in the Fractional-order discrete-time modified-Hénon chaotic system's dynamics. The resulting chaotic image is watermarked in a host image (Lena image) by single-level 2-D wavelet decomposition (DWT algorithm). After the extraction process, the encrypted watermark is then decrypted at the level of two observers to obtain the original image.

The application of the proposed scheme for the transmission of an image has given good results for the recovery of the watermark. This is illustrated by the simulation results given in the previous section. The watermark has been well recovered at the receiver level, which highlights the robustness of this scheme. The robustness of this scheme and the security of the watermark are the primary focus of this research, which explains that its main contribution is geared towards fractional and integer order chaotic systems combined with watermarking. The proposed scheme can be improved by taking into account the transmission channel's noise and using only a single channel. The robustness can be improved by designing an optimal protocol for real-time key transmission, and analyze the different robustness methods. We can also plan the experimental design of the proposed transmission scheme using microcontroller circuits' devices.

REFERENCES

- [1] C. Zhu, K. Sun. "Chaos Application In Digital Watermarking. Applications of Chaos and Nonlinear Dynamics in Science and Engineering" - Vol. 2, pp.187-232.
- [2] N. Khorrami, P. Ayubi, S. Behnia, J. Ayubi. "A SVD-Chaos Digital Image Watermarking Scheme Based on Multiple Chaotic System", V.V. Das and P.M. El-Kafrawy (Eds.): SPIT 2012, LNICST 117, pp. 9–18, 2014.
- [3] M. Lahdir, H. Hamiche, S. Kassim, M. Tahanout, K. Kemih, S-A Addouche, "A novel robust compression-encryption of images based on SPIHT coding and fractional-order discrete-time chaotic system", *Optics & Laser Technology*, vol. 109, pp. 534-546, 2019.
- [4] A. Ali-Pacha, N. Hadj-Said, A. M'hamed, A. Belghoraf. "Chaos Crypto-Système basé sur l'Attracteur de Hénon-Lozi". Conference: Proceedings of the 2nd Conférence Internationale sur l'Informatique et ses Applications (CIIA'09), Saida, Algeria, May 3-4, (2009).
- [5] H.Hamiche, S.Guermah, R. Saddaoui, K. Hannoun, M.Laghrouche and S. Djennoune, "Analysis and implementation of a novel robust transmission scheme for private digital communications using Arduino-Uno board", *Nonlinear. Dynam.* Vol. 81, No. 4, pp.1921–1932, 2015.
- [6] O. Megherbi, S Guermah, H Hamiche, S Djennoune, M Ghanes, "A novel transmission scheme based on impulsive synchronization of two Colpitts chaotic systems", *Systems and Control (ICSC)*, 3rd International Conference on, pp. 117-122, 2013.
- [7] S. Kassim, H. Hamiche, S. Djennoune, M. Bettayeb, "A novel secure image transmission scheme based on synchronization of fractional-order discrete-time hyper chaotic systems", *Nonlinear Dynam*, vol. 88, no. 04, pp. 2473-2489, 2017.
- [8] H. Hamiche, S. Guermah, S. Kassim, M. Lahdir, S. Djennoune, M.Bettayeb. "Secure data transmission scheme based on fractional-order discrete chaotic system". *Control, Engineering and Information Technology (CEIT)*, 2015 3rd International Conference on chaos, Tlemcen, Algérie, (2015).
- [9] L.M. Pecora and T.L. Carroll, "Synchronization in Chaotic Systems, *Physicals Review and Letters*", pp. 821–824, 1990.
- [10] Qi H, Zheng D, Zhao J. "Human visual system based adaptive digital image watermarking". *Signal Process*, vol. 88, pp.174–88, 2007.
- [11] F.Y. Shih, S.Y.T.Wu: Combinational image watermarking in the Spatial and frequency domains [J]. *Pattern Recogn*, vol. 36, no. 4, pp.969–975, 2003.
- [12] P. Khare, A.K. Verma and V. K. Srivastava: "Digital Image Watermarking Scheme in Wavelet Domain Using Chaotic Encryption", *SCES*, 2014.
- [13] A.A. Kilbas, H.M.Srivastava, J.J.Trujillo: "Theory and Application of Fractional Differential Equations". North Holland Mathematics Studies. Editor Jan van Mill, Elsevier, Amsterdam, (2006).
- [14] K. Hannoun, H. Hamiche, M. Lahdir, M. Laghrouche, S. Kassim, "A Novel DWT Domain Watermarking Scheme Based On a Discrete-Time Chaotic System", *IFAC PapersOnline*, vol. 51-33, pp. 50–55, Eindhoven, Netherlands, 2018.
- [15] H. Hamiche, O. Megherbi, R. Kara, R. Saddaoui, M. Laghrouche, S. Djennoune, "A new implementation of an impulsive synchronization of two discrete-time hyper chaotic systems using Arduino-Uno boards", *International Journal of Modeling, Identification and Control*, , vol. 28, no.2 , pp. 177-186, 2017.

An Improved PSO Based Approach For Two-Dimensional IIR Filter Design

Farid Hammou

Electrical Engineering Department
Polytechnic School of Oran - ENPO
Oran, Algeria
Email: faridhammou@yahoo.fr

Kamal Hammouche

Laboratoire Vision Artificielle et Automatique des Systemès
Mouloud Mammeri University of Tizi-Ouzou
Tizi-Ouzou, Algeria
Email: kamal_hammouche@yahoo.fr

Abstract—In this paper, an improved version of the Particle Swarm Optimization (PSO) algorithm is proposed for the design of two-dimensional digital Infinite Impulse Response (2D-IIR) filters. The proposed algorithm, called Cooperation- Hierarchization based PSO (CHPSO), introduces a new strategy based on cooperation and hierarchization concepts for the updating of the best positions of particles in order to improve the convergence of the PSO algorithm. Experimental results demonstrate that the proposed CHPSO gives better performance as compared to two others variants of PSO and to tree popular evolutionary optimization algorithms.

Index Terms—Two-dimensional IIR filter, Particle swarm optimization, Cooperation-Hierarchization concepts, Evolutionary optimization techniques.

I. INTRODUCTION

Two-dimensional (2D) filters are widely used in image processing [1], medical imaging [2], face recognition [3]. In digital image processing, the main function of a filter is to remove unwanted parts of an image, such as random noise, or to extract useful parts of the image, such as the components lying within a certain frequency range.

Digital filters are broadly classified into two main categories namely, finite impulse response (FIR) filters and infinite impulse response (IIR) filters [4]. A zero-phase 2D-IIR filter is a special case of a linear phase 2D-IIR filter in which the phase slope is zero. Noting that, the zero-phase 2D-IIR filter is not causal and cannot be used in on-line applications. However, the zero-phase 2D-IIR filter has some advantages: It needs fewer arithmetical operations and has smaller memory usage than the linear phase 2D-IIR filters and in many off-line applications, where the signal to filter is stored as files on a computer. Moreover, causality is not a requirement and zero-phase 2D-IIR filters are usually preferred to linear phase 2D-IIR filters [5].

Design of a digital IIR filter consists in determining the coefficients of a stable and realizable function transfer from the prescribed specifications. It can be formulated as an optimization problem with nonlinear and multi-modal objective function with respect to the filter coefficients [6]. The most popular design methods, which may result in an unstable filter, are based either on an appropriate transformation of 1D-IIR filters [7], spectral or frequencial transformation of a low-Passe 1D-IIR filter prototype [8], [9] or on standard optimization

optimization techniques like Gradient based algorithms [10]. Modern heuristic methods have also been employed for 2D-IIR filter design problems, such as Genetic Algorithm (GA) [11], Differential Evolution (DE) [12], Bees Algorithm (BEA) [13], Bacteria Foraging Optimization (BFO) [14], Simulated Annealing (SA) [15], Particle Swarm Optimization (PSO) [5] and Quantum PSO (QPSO) [16]. These techniques were able to find out better solutions than those previously mentioned.

Particle swarm optimization (PSO) is a population based evolutionary algorithm which mimic biological mechanisms, like bird flocking or fish schooling [17]. The key attraction of PSO is its simplicity in concept, ease to implement and its speed converges towards a high quality of solution. Although the PSO algorithm has advantage in solving global optimization problem compared to other evolutionary algorithms, it suffers from some limitations related to the premature convergence and stagnation problem. Several modifications of the basic PSO algorithm involving different strategies have been proposed in the literature and applied to 2D-IIR filter design [6], [18], [19]. More recently, we have proposed in [20] an improved version of PSO, so called CHPSO, that introduces a new strategy for updating personal best position of each particle. The idea behind this strategy is to allow a particle to use the positions experienced by its congeners as its own best position.

In the present paper, we propose to use CHPSO for 2D-IIR filter design problem.

The rest of this paper is organized as follows. The digital 2D-IIR filter design problem is outlined in section II. Section III give an overview of the proposed CHPSO algorithm for digital 2D-IIR filter design. Simulations results are presented section IV.

II. IIR FILTER DESIGN PROBLEM

A Two-dimensional digital IIR filter is described by the following input-output relationship :

$$y(n_1, n_2) = - \sum_{i=1}^{L_1} \sum_{j=1}^{L_2} a(i, j) y(n_1 - i, n_2 - j) + \sum_{i=0}^{L_1} \sum_{j=0}^{L_2} b(i, j) x(n_1 - i, n_2 - j) \quad (1)$$

where $x(n_1, n_2)$ and $y(n_1, n_2)$ are the filter's input and output, respectively. The sequences $a(i, j)$ and $b(i, j)$ ($i = 0, 1, 2, \dots, L_1$ and $j = 0, 1, 2, \dots, L_2$) are the matrices of the filter coefficients.

The transfer function for the zero-phase 2D-IIR filter with taking assumption that $b_{00} = 1$ is expressed as :

$$H(z_1, z_2) = H_0 \frac{\sum_{i=0}^L \sum_{j=0}^L b_{ij} z_1^i z_2^j}{\prod_{k=1}^K (1 + q_k z_1 + r_k z_2 + s_k z_1 z_2)} \quad (2)$$

where b_{ij} , q_k , r_k and s_k are the coefficients of the filter, L and K are the orders of the numerator and denominator, respectively, with $K \leq L$, H_0 is a constant. $z_1 = e^{j\omega_1}$, $z_2 = e^{j\omega_2}$ and $\omega_1, \omega_2 \in [-\pi, \pi]$.

The design task of the zero-phase 2D-IIR filter is to determine the values of the coefficients $a(i, j)$, $b(i, j)$, q_k , r_k and s_k of a stable and realizable transfer function $H(z_1, z_2)$, such that the magnitude function $|H(\omega_1, \omega_2)|$ approximates as close as possible to the desired amplitude response $D(\omega_1, \omega_2)$. This, can be formulated as an optimization problem of the cost function $J(\alpha)$, stated as follows:

$$\alpha^* = \operatorname{argmin} J(\alpha) \quad (3)$$

where $\alpha = [H_0, b_{ij}, q_k, r_k, s_k]$ denotes the filter coefficient vector.

The cost function $J(\alpha)$ can be expressed as an error function between the magnitude response $|H(\omega_1, \omega_2)|$ and the desired response magnitude $D(\omega_1, \omega_2)$ at $(N_1 \times N_2)$ points:

$$J(\alpha) = \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} [|H(\omega_1, \omega_2)| - D(\omega_1, \omega_2)]^p \quad (4)$$

where p is an even positive integer (usually $p = 2$ or 4), $H(\omega_1, \omega_2) = H(z_1, z_2)$ with $z_1 = e^{j\omega_1}$ and $z_2 = e^{j\omega_2}$. $\omega_1 = i(\pi/N_1)$ and $\omega_2 = j(\pi/N_2)$ in $[-\pi, \pi]$ are the digital frequencies.

Equation (4) can be expressed as:

$$J(\alpha) = \sum_{i=0}^{N_1} \sum_{j=0}^{N_2} \left[\left| H \left(\frac{i\pi}{N_1}, \frac{j\pi}{N_2} \right) \right| - D \left(\frac{i\pi}{N_1}, \frac{j\pi}{N_2} \right) \right]^\gamma \quad (5)$$

In addition, to ensure the stability of the designed 2D-IIR filter, stability conditions are integrated in the objective function. Since the denominator of the transfer function $H(z_1, z_2)$ of zero-phase 2D-IIR digital filter contains only first degree factors, we can assert the stability conditions, following [5], [11], [21]

$$|q_k + r_k| - 1 < s_k < |q_k - r_k| \quad k = 1, 2, \dots, K \quad (6)$$

Hence, the design of 2-D recursive filters is a constrained minimization problem. The aim is to Minimize $J(\alpha)$ [5] subject to the constraints imposed in [6]. This optimization problem has been tackled using GA in [11], DE in [12] and using BEA in [13]. Related work can be found in [6].

III. THE PROPOSED CHPSO ALGORITHM

CHPSO algorithm starts with a population of M particles characterized principally by their positions α_p , ($p = 1, 2, 3, \dots, M$) in Q dimension search space which correspond to the candidate solutions of an optimization problem. In the case of 2D-IIR filter design, each particle is a set of $Q = (L+1)^2 + 3K$ filter coefficients $\alpha_p = [H_0, b_{ij}, q_k, r_k, s_k]$. The quality of each particle is evaluated by the error function (fitness) $J(\alpha_k)$ Eq. [5]. At time t , each particle k moves in the search space with a velocity $v_p(t)$. The velocity and the position of the p^{th} particle are updated in each iteration according to its previous position $\alpha_p(t)$, its own best position $\beta_p(t)$ found so far and the global best position $\alpha^*(t)$ found by all particles, such that:

$$v_p(t+1) = wv_p(t) + c_1\varphi_1(\beta_p(t) - \alpha_p(t)) + c_2\varphi_2(\alpha^*(t) - \alpha_p(t)) \quad (7)$$

$$\alpha_p(t+1) = \alpha_p(t) + v_p(t+1) \quad (8)$$

Acceleration coefficients c_1 and c_2 are the two factors which control the influence of the attraction exerted by the best personal experience β_p of each particle and by the attraction of the global best position α^* , respectively. φ_1 and φ_2 are two uniformly distributed random numbers, independently generated within $[0, 1]$. The inertia weight ω controls how much the particles tend to follow their current direction $\alpha(t)$ compared to the memorized positions $\beta_p(t)$ and $\alpha^*(t)$.

Unlike PSO as well as all most of its variants, in CHPSO a new strategy for updating the personal best position $\beta_p(t+1)$ of each particle p is adopted in order to improve the convergence speed and the robustness of PSO.

Indeed, in PSO, the new personal best position $\beta_p(t+1)$ of each particle p at the next time step, $t+1$, is updated by comparing its new position $\alpha_p(t+1)$ and its previous own best position $\beta_p(t)$:

$$\beta_p(t+1) = \begin{cases} \alpha_p(t+1) & \text{if } J(\alpha_p(t+1)) < J(\beta_p(t)) \\ \beta_p(t) & \text{else} \end{cases} \quad (9)$$

The global best position, $\alpha^*(t+1)$, at time step $t+1$, is updated as follows:

$$\alpha^*(t+1) = \begin{cases} \beta_{p^*}(t+1) & \text{if } J(\beta_{p^*}(t+1)) < J(\alpha^*(t)) \\ \alpha_p(t) & \text{else} \end{cases} \quad (10)$$

where $p^* = \operatorname{argmin}_{p=1, \dots, M} [J(\beta_p(t+1))]$.

While, in CHPSO algorithm, the personal best position update of a particle is done by taking into account all old personal best positions and all new positions of all particles. The idea behind this strategy is to allow a particle to use the positions experienced by its congeners as its own best position.

This procedure can be summarized by the following three steps:

Step I: We gather into a vector $Y = [y_1, y_2, \dots, y_{2M}]^T$ of size $(2M)$ all personal best positions $\beta_p(t)$ ($p = 1, 2, 3, \dots, M$) of all particles, at the time step t , and the

new positions $\alpha_p(t+1)$ ($p = 1, 2, 3, \dots, M$) of all particles generated at the next time step $t+1$:

$$Y = [\beta_1(t), \beta_2(t), \dots, \beta_M(t), \alpha_1(t+1), \alpha_2(t+1), \dots, \alpha_M(t+1)]^T \quad (11)$$

The p^{th} element of this vector is:

$$y_p = \begin{cases} \beta_p(t) & \text{if } k \leq M \\ \alpha_{p-M}(t+1) & \text{if } p > M \end{cases}, p = 1, 2, \dots, 2M \quad (12)$$

Step II: The $2M$ elements of the vector $Y(t+1)$ are arranged in ascending order according to their fitness leading to the following vector:

$$Y = [y_{(1)}, y_{(2)}, \dots, y_{(2M)}]^T \quad (13)$$

Step III: The M new personal best positions of the particles are finally chosen as the M first ordered elements of Y :

$$\beta_p(t+1) = y_{(k)}, p = 1, 2, \dots, M \quad (14)$$

Finally, the global best position $\alpha^*(t+1)$, at time step $t+1$, is updated simply from the personal best position of the more experienced particle, i.e., the first particle in the population:

$$\alpha^*(t+1) = \begin{cases} \beta_1(t+1) & \text{if } J(\beta_1(t+1)) < J(\alpha^*(t)) \\ \alpha^*(t) & \text{else} \end{cases} \quad (15)$$

In this updating strategy of personal best positions, the first particle memorizes the first best position found by all particles, the second particle memorizes the second best position found by all particles and so on until the last particle which stores the M^{th} best position found by all particles. This means that the particles cooperate with each other by exchanging their experiences, because a particle can consider a position occupied by one of its congeners as its own best experience, and that the positions experienced by a particle can serve as personal best experiences to other particles. Moreover, the particles are hierarchized according to their personal best experiences that have been transmitted to them, since a particle receives from its congeners one position to memorize according to the rank that it occupies in the population.

As CHPSO acts on the cognitive component of PSO, we apply the concept of time-varying acceleration coefficients [22]. Moreover, unlike the basic PSO algorithm, the random numbers φ_1 and φ_2 that determine the weight between the cognitive attraction and social attraction are dependently generated as $\varphi_2 = 1 - \varphi_1$.

IV. SIMULATION RESULTS AND DISCUSSION

In this section, we evaluate the performance of our presented CHPSO approach and demonstrate its effectiveness for 2D-IIR filter design. The results achieved by CHPSO are compared to those obtained by two improved versions of PSO, namely adaptive inertia Weight PSO (WPSO) [5] and Quantum-behaved PSO (QPSO) [16] and three other popular

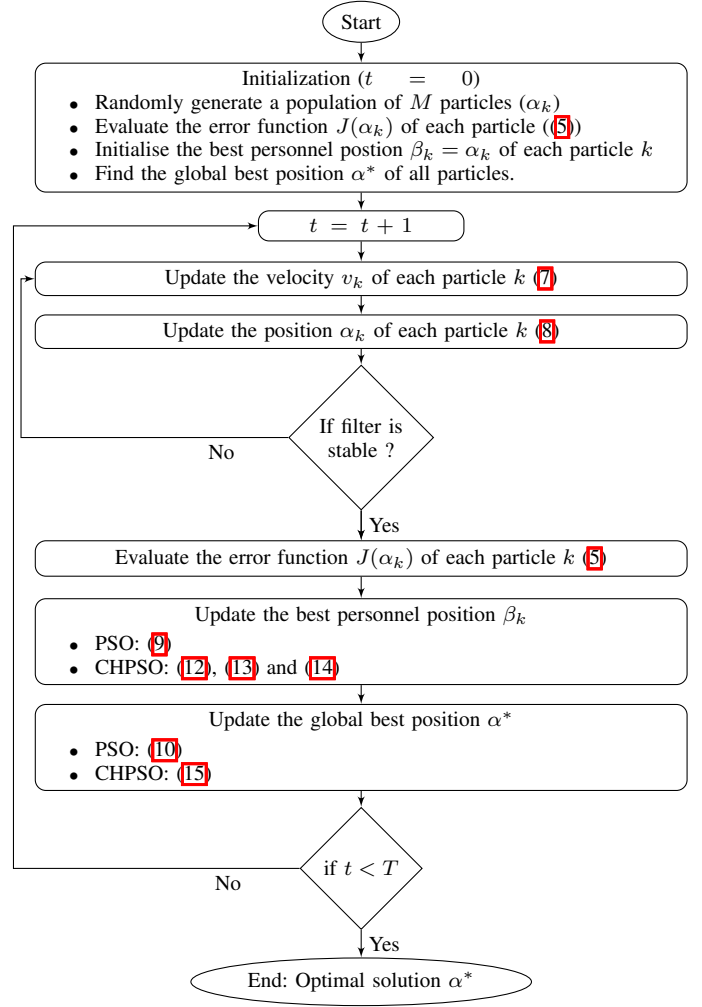


Fig. 1. Flowchart of standard PSO and the CHPSO algorithms for 2D-IIR filter design

evolutionary methods including Genetic algorithm (GA) [11], Differential evolution (DE) [12] and Bee Algorithm (BEA) [13]. The parameter of each algorithm are given in Table IV. For a fair comparison, the size of population $M = 50$ and the total number of iterations $T_{max} = 1000$ are the same for all algorithms.

TABLE I
CONTROL PARAMETERS OF THE ALGORITHMS USED IN THE SIMULATIONS

Algorithm	Parameters	Source
CHPSO	$c^{min}=0.2; c^{max}=1.2; w=0.7298$	-
WPSO	$c_1 = c_2=2.05; w^{max}=1; w^{min}=0.4$	[5]
QPSO	$\alpha^{min}=0.5; \alpha^{max}=1$	[16]
DE	$C_r=0.3; C_f=0.5$	[12]
GA	Crossover rate=1; Crossover: Two-point crossover; Mutation rate= 0.01; Mutation: Roulette; Selection probability=1/3.	[11]
BEA	Limite=40; a=0.9; b=0.1	[13]

A. Performance metrics

The value of the best fitness $J(\alpha^*)$ corresponding to the best solution is used as comparative criterion. Of course, the smaller the fitness value, the better the algorithm is. Moreover, the Error of frequency response in Pass-band and Stop-band and Maximum Pass-band magnitude ripple are also used to evaluate the quality of the designed filter.

- 1) Error of frequency response in Pass-band:

$$E_{PB}(\omega_1, \omega_2) = |H(\omega_1, j\omega_2) - D(\omega_1, j\omega_2)|_{\omega_1, \omega_2 \in [0, \omega_p]}$$

- 2) Error of frequency response in Stop-band:

$$E_{SB}(\omega_1, \omega_2) = |H(\omega_1, j\omega_2) - D(\omega_1, j\omega_2)|_{\omega_1, \omega_2 \in [\omega_s, \pi]}$$

- 3) Maximum Pass-band magnitude ripple:

$$MR_{PB}(\omega_1, \omega_2) = |H(\omega_1, j\omega_2)|_{\omega_1, \omega_2 \in [0, \omega_p]} - 1$$

where ω_p and ω_s are the Pass-band and Stop-band frequencies, respectively. Note that, the best designed filter must have the smallest error in Pass-band E_{PB} , the lowest maximum Pass-band ripple MR_{PB} to avoid the distortion of the produced output signal and the minimum error value in Stop-band E_{SB} in order to ensure the maximum attenuation of the unwanted frequencies components. Additional results are presented in order to investigate the convergence speed of each algorithm.

B. Experimental results

Let us consider a low-pass filter with a small area of passability as an example of the design problem [11], [13], [14], [16]. The user-specification for the desired circular symmetric low-Pass filter response is given as:

$$D(\omega_1, \omega_2) = \begin{cases} 1, & \sqrt{\omega_1^2 + \omega_2^2} \leq \omega_p, \\ 0.5, & \omega_p < \sqrt{\omega_1^2 + \omega_2^2} \leq \omega_s, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

For this example, the Pass-band and Stop-band frequencies are set to $\omega_p = 0.08\pi$ and $\omega_s = 0.12\pi$, respectively. Without

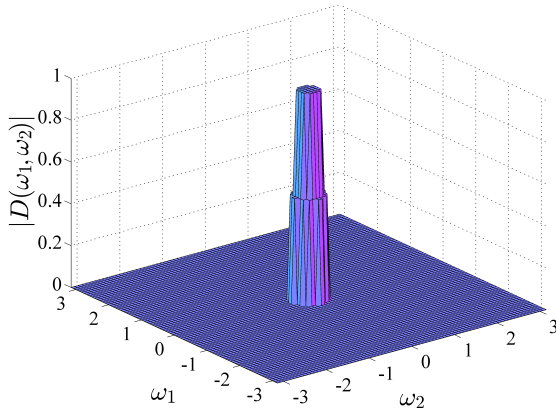


Fig. 2. Desired amplitude response of the 2D filter.

loss of generality let us assume $L = 2$, $K = 2$, the transfer function $H(z_1, z_2)$ of the 2D-IIR filter is given by:

$$H(z_1, z_2) = H_0 \frac{\sum_{i=0}^2 \sum_{j=0}^2 b_{ij} z_1^i z_2^j}{\prod_{k=1}^2 (1 + q_k z_1 + r_k z_2 + s_k z_1 z_2)} \quad (17)$$

Of course, the transfer function in (17) can be restated as:

$$H(z_1, z_2) = H_0 \frac{Num(z_1, z_2)}{Den(z_1, z_2)}, \quad (18)$$

with,

$$\begin{aligned} Num(z_1, z_2) &= b_{00} + b_{01} z_2 + b_{02} z_2^2 + b_{10} z_1 + b_{20} z_1^2 + \\ &\quad b_{11} z_1 z_2 + b_{12} z_1^2 z_2^2 + b_{21} z_1^2 z_2^1 + b_{22} z_1^2 z_2^2. \\ Den(z_1, z_2) &= (1 + q_1 z_1 + r_1 z_2 + s_1 z_1 z_2) \times \\ &\quad (1 + q_2 z_1 + r_2 z_2 + s_2 z_1 z_2). \end{aligned}$$

The solution vector $\alpha = [H_0, b_{00}, b_{01}, b_{02}, b_{10}, b_{20}, b_{11}, b_{12}, b_{21}, b_{22}, q_1, r_1, s_1, q_2, r_2, s_2]^T$ regroups the coefficients of the designed 2D-IIR filter. Choosing the values $\gamma = 2$, $N_1 = 50$ and $N_2 = 50$, 2D-IIR filter design problem is stated by the following constrained minimization problem,

Minimize :

$$J(\alpha) = \sum_{i=0}^{50} \sum_{j=0}^{50} \left[\left| H\left(\frac{i\pi}{50}, \frac{j\pi}{50}\right) - D\left(\frac{i\pi}{50}, \frac{j\pi}{50}\right) \right|^2 \right] \quad (19)$$

subject to the stability constraints (20) derived from (6).

$$\begin{aligned} -(1 + s_k) &< (q_k + r_k) < (1 + s_k), \\ -(1 - s_k) &< (q_k + r_k) < (1 - s_k), \\ (1 + s_k) &> 0, \\ (1 - s_k) &> 0. \end{aligned} \quad (20)$$

C. Analysis of magnitude responses

As the evolutionary algorithms are of stochastic type, they are run 20 times. Table II reports the best, the mean and the standard deviation of the fitness values over 20 runs achieved by each competitor algorithm. The best results among different methods are highlighted in bold. We can see that CHPSO gives the best results both in terms of accuracy and robustness. Indeed, the best and the mean fitness of CHPSO is always the smallest.

TABLE II
STATISTICAL RESULTS OF FITNESS FOR 2nd ORDER LOW-PASS 2D-IIR FILTERS

Algorithm	CHPSO	WPSO	QPSO	DE	AG	BEA
Best	3.4466	4.0297	5.0953	3.8488	5.2730	4.9376
Mean	4.0370	5.2159	7.3895	4.2596	6.0820	5.6541
Std	0.5268	0.7533	1.8283	0.2919	0.4960	0.5075

Figure 3 shows the plots of the normalized gains $H(\omega)$ computed with the coefficients corresponding to the best solution obtained by each competitor algorithm.

In Tables III the smallest (Best), mean (Mean) and standard deviation (Std) of Maximum Pass-band ripple MR_{PB} are listed for all algorithms, which clearly reveal that the Best/Mean solutions achieved by CHPSO algorithm are best for all runs.

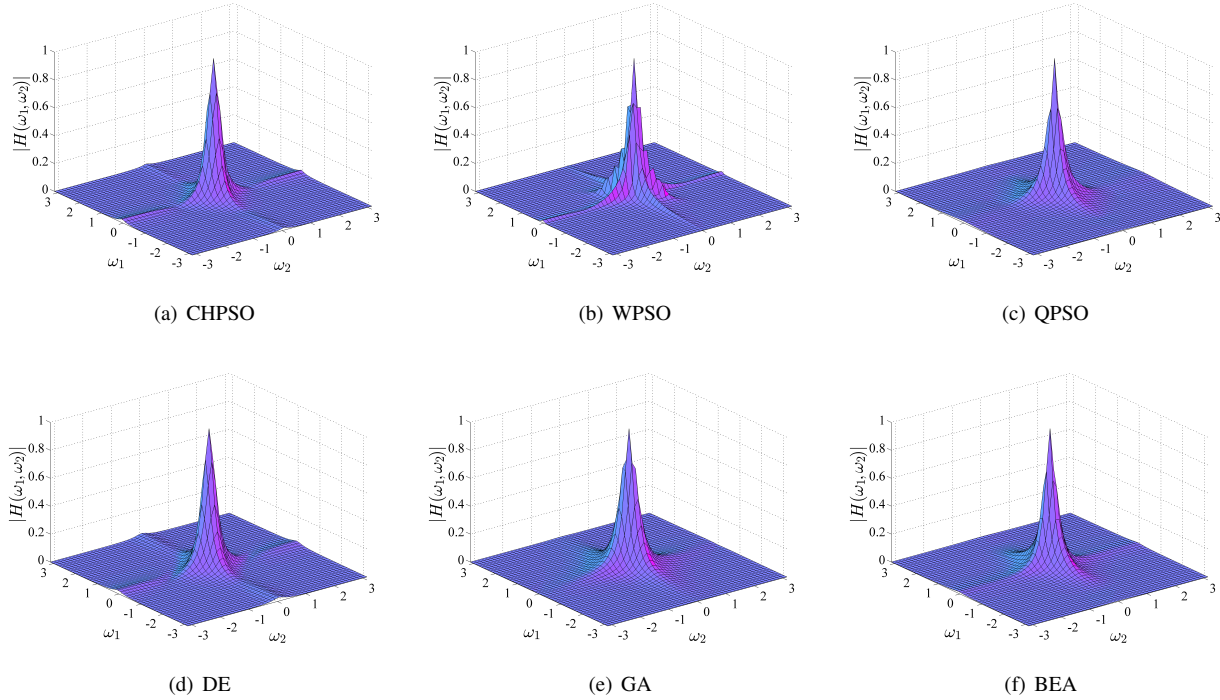


Fig. 3. Normalized gain pots of 2^{nd} order Low-Pass 2D-IIR filter.

TABLE III

STATISTICAL ANALYSIS OF MAXIMUM PASS-BAND MAGNITUDE RIPPLE OF THE DESIGNED 2D-IIR FILTERS

Algorithm	CHPSO	WPSO	QPSO	DE	AG	BEA
Max	17.453	19.999	23.843	18.604	20.306	19.276
Mean	16.280	17.920	19.834	16.317	18.799	18.221
Std	0.7162	1.0178	2.0306	1.4314	0.8795	0.7036

TABLE V

STATISTICAL ANALYSIS OF FREQUENCY RESPONSE ERROR IN STOP-BAND OF THE DESIGNED 2D-IIR FILTERS

Algorithm	CHPSO	WPSO	QPSO	DE	AG	BEA
Max	12.315	18.224	26.611	15.004	21.581	27.002
Mean	21.348	29.072	39.717	22.712	30.735	35.536
Std	06.510	07.244	10.030	03.577	04.622	07.320

Tables IV and V reports the best, the mean and the standard deviation of frequency response error in Pass-band (E_{PB}) and Stop-band (E_{SB}) respectively, of the designed 2D-IIR filters by each competitor algorithm. Basing on results of Table V, it can be easily verified that the designed 2D-IIR filter using CHPSO has the best statistical values in terms of E_{PB} and E_{SB} .

TABLE IV

STATISTICAL ANALYSIS OF FREQUENCY RESPONSE ERROR IN PASS-BAND OF THE DESIGNED 2D-IIR FILTERS

Algorithm	CHPSO	WPSO	QPSO	DE	AG	BEA
Max	7.9588	10.739	13.843	9.2906	10.306	9.5436
Mean	7.0357	8.3301	10.040	7.1712	9.0059	8.3506
Std	0.5586	0.9542	2.0094	0.9814	0.7710	0.6830

The standard deviation obtained by the proposed CHPSO is slightly larger than the values provided by the DE algorithm in term of E_{SB} and than BEA algorithm in term of MR_{PB} .

D. Convergence speed

Figure 4 shows how the mean fitness varies with the iteration number. At each iteration, the fitness obtained by the

CHPSO is much lower than those obtained by the other algorithms. Moreover, the fitness curve decreases faster in the case of CHPSO comparatively to all other algorithms. Therefore, the iteration number necessary to ensure the convergence of CHPSO is lower than those of other algorithms.

In order to quantify the computational effort required by the proposed algorithm, the average execution time consumed by each algorithm over 20 runs is recorded in Table VI.

TABLE VI

COMPUTATIONAL TIME VALUES (in seconds S) CONSUMED BY EACH ALGORITHM

Algorithm	CHPSO	WPSO	QPSO	DE	AG	BEA
AET	1888.5	1900.7	1879.6	1888.1	1787.5	22026.

The average execution time value (AET) of the computational time consumed by CHPSO and all PSO's variants over 20 runs are almost similar. This shows that the proposed updating rule of personal best position of particles does not affect the time execution.

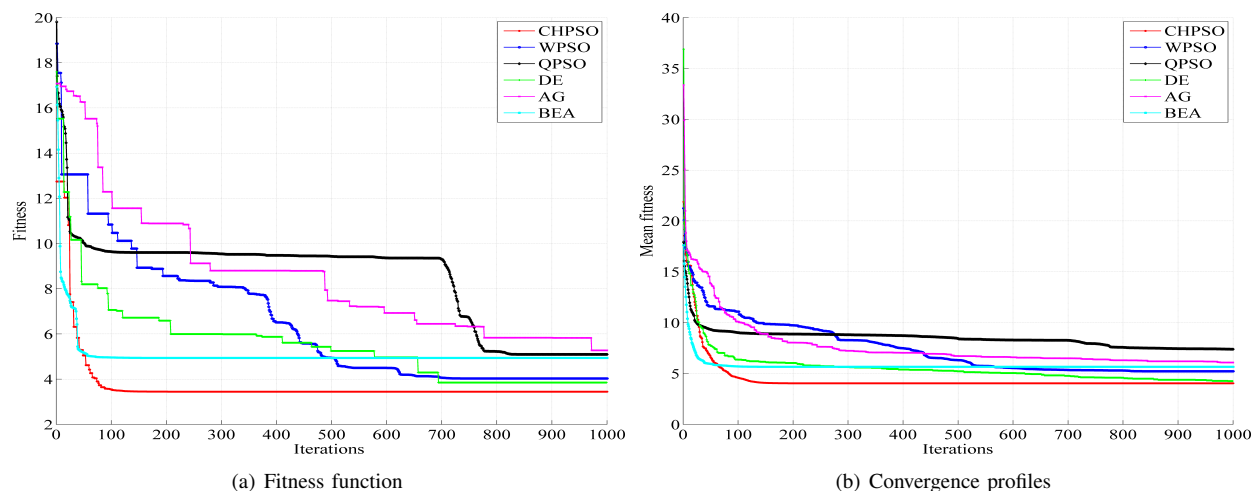


Fig. 4. Fitness and convergence profiles obtained by all competitor algorithms.

V. CONCLUSION

In this paper, an improved version of PSO algorithm is used for the 2D-IIR filter design. This improved version of PSO stands out from the other existing variants by the way of updating the personal best position of the particles. It introduces a novel concept of cooperation between particles and that of their hierarchization according to their personal best positions.

Experiments on the design of 2nd order Low-Pass 2D-IIR filters reveal the efficiency of the proposed CHPSO algorithm in terms of accuracy, robustness and convergence speed, compared to two PSO's variants and tree popular evolutionary optimization methods.

REFERENCES

- [1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006.
- [2] J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, "A recursive filter for noise reduction in statistical iterative tomographic imaging," pp. 6065 – 6065 – 10, 2006.
- [3] S. Gong, S. J. McKenna, and A. Psarrou, *Dynamic Vision: From Images to Face Recognition*, 1st ed. London, UK, UK: Imperial College Press, 2000.
- [4] A. V. Oppenheim, R. W. Schaffer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Prentice-hall Englewood Cliffs, 1999.
- [5] S. Das and A. Konar, "A swarm intelligence approach to the synthesis of two-dimensional IIR filters," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 8, pp. 1086–1096, 2007.
- [6] S. Kockanat and N. Karaboga, "The design approaches of two-dimensional digital filters based on metaheuristic optimization algorithms: a review of the literature," *Artificial Intelligence Review*, vol. 44, no. 2, pp. 265–287, 2015.
- [7] C. L. Chan and H. K. Kwan, "Simple design of 2-dimensional circularly symmetric recursive digital filters using 1-dimensional analogue filters," *International Journal of Electronics Theoretical and Experimental*, vol. 67, no. 4, pp. 585–590, 1989.
- [8] R. Matei and D. Matei, "Orientation-selective 2d recursive filter design based on frequency transformations," in *IEEE EUROCON 2009*, 2009, pp. 1320–1327.
- [9] L. Harn and B. Sheno, "Design of stable two-dimensional IIR filters using digital spectral transformations," *IEEE Transactions on Circuits and Systems*, vol. 33, no. 5, pp. 483–490, May 1986.
- [10] A. Antoniou, *Digital filters*. McGraw Hill, 1993.
- [11] N. E. Mastorakis, I. F. Gonos, and M. N. S. Swamy, "Design of two-dimensional recursive filters using genetic algorithms," *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, vol. 50, no. 5, pp. 634–639, May 2003.
- [12] L. Wu, Y. Wang, and X. Yuan, "Design of 2-d recursive filters using self-adaptive mutation differential evolution algorithm," *International Journal of Computational Intelligence Systems*, vol. 4, no. 4, pp. 644–654, 2011.
- [13] D. T. Pham and E. Koç, "Design of a two-dimensional recursive filter using the bees algorithm," *International Journal of Automation and Computing*, vol. 7, no. 3, pp. 399–402, 2010.
- [14] R. Panda, N. Kumar, and N. Mishra, "Design of two-dimensional recursive filters using bacteria foraging optimization," pp. 188–193, 04 2013.
- [15] S. Dhabal and P. Venkateswaran, "Two-dimensional IIR filter design using simulated annealing based particle swarm optimization," *Journal of Optimization*, vol. 2014, no. Article ID 239721, 2014.
- [16] W. Fang, J. Sun, and W. Xu, "Design of two-dimensional recursive filters by using quantum-behaved particle swarm optimization," in *Intelligent Information Hiding and Multimedia Signal Processing, 2006. IHH-MSP'06. International Conference on*, 2006, pp. 240–243.
- [17] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN95 - International Conference on Neural Networks*. IEEE, 1995, pp. 1942–1948.
- [18] J. Sun, C. Lai, and X. Wu, *Particle Swarm Optimisation: Classical and Quantum Perspectives*, ser. Chapman & Hall/CRC Numerical Analysis and Scientific Computing Series. CRC Press, 2016.
- [19] C. Lv, S. Yan, G. Cheng, L. Xu, and X. Tian, "Design of two-dimensional iir digital filters by using a novel hybrid optimization algorithm," *Multidimensional Syst. Signal Process.*, vol. 28, no. 4, pp. 1267–1281, Oct. 2017.
- [20] F. Hammou and K. Hammouche, "An improved particle swarm optimization algorithm towards iir system identification," in *020 1st International Conference on Communications, Control Systems and Signal Processing (CCSSP)*, 2020, pp. 107–112.
- [21] T. Miyata, N. Aikawa, Y. Sugita, and T. Yoshikawa, "A design method for separable-denominator 2d IIR filters using a stability criterion based on the system matrix," in *2008 15th IEEE International Conference on Electronics, Circuits and Systems*, Aug 2008, pp. 826–829.
- [22] A. Ratnaweera, S. K. Halgamuge, and H. C. Watson, "Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients," *IEEE Transactions on Evolutionary Computation*, vol. 8, no. 3, pp. 240–255, 2004.

Video encryption-decryption based on synchronisation of chaotic systems for secure communication

Ahcene Hamoudi¹, Nadia Djeghali¹ and Maamar Bettayeb^{2,3}

1.Mouloud Mammeri University, Tizi-Ouzou, Algeria

Laboratoire de Conception et Conduite des Systèmes de Production

ahcene.hamoudi@ummto.dz, nadia.djeghali@ummto.dz

2.University of Sharjah, UAE

3.(CEIES) King Abdulaziz University, KSA

maamar@sharjah.ac.ae

Abstract—This paper presents a secure video transmission scheme based on synchronisation of discrete-time chaotic systems. The transmitter is made up of two discrete-time chaotic systems: the logistic and modified-Henon maps. The logistic map is used to encrypt the original video signal, increasing the transmission scheme robustness. The modified-Henon system's dynamics are modified to include the encrypted message obtained. A dead-beat observer is used as the receiver. The video message can be recovered at the receiver end once synchronisation between the transmitter and receiver has been established. In order to prove the effectiveness of the proposed method, several tests were presented and different analysis are given to illustrate the security of the proposed scheme.

Index Terms—Video encryption, Chaotic systems, Synchronisation, Cryptography, Secure communication, Video processing, Dead-beat observer.

I. INTRODUCTION

Security and confidentiality are becoming a real challenge and a major research topic as information and communication technologies advance. One of the options being considered is encryption, which is a good way to share information in a secure manner by encrypting data on the emitter and decrypting it at the receiver [1]. Digital communication techniques based on chaotic systems have been the subject of intensive research since Pecora and Carroll pioneering work in the field of chaos synchronisation in 1991 [2].

Researchers are increasingly interested in the application of chaos theory and data encryption [3]. Chaotic systems, in fact, have unique and advantageous properties, such as a high sensitivity to initial conditions and system parameters, a seemingly random evolution, a high level of security, and ease of implementation. In light of these benefits, this work focuses

on a secure video encryption system based on chaotic system synchronisation [4], [5].

Synchronisation of chaos is a phenomenon that may occur when two chaotic systems are coupled. Because of the exponential divergence of the nearby trajectories of chaotic systems, having two chaotic systems evolving in synchrony might appear surprising [6]. However, synchronisation of coupled or driven chaotic oscillators is a phenomenon well established experimentally and reasonably well-understood theoretically. The synchronisation of chaos is a rich phenomenon and a multi-disciplinary subject with a broad range of applications [7].

Since the work of Nijmeijer and Mareels, the synchronisation can be considered as a state estimation problem where the slave system is designed based on a state observer for the master system [8].

Chaos theory has been extensively studied as a new tool for designing digital ciphers, particularly to develop image encryption algorithms [9], [10]. Many discrete chaotic maps are simple, and the well-known chaos theory allows for practical approaches to image encryption [11]. The success and failure of chaos-based encryption schemes have yielded some valuable lessons and experiences that can be applied to future research on chaos-based multimedia encryption [12]. At this time, chaos theory appears to be promising for video encryption.

In this paper, a data transmission scheme based on discrete-time chaotic systems is proposed. The transmitter, is made up of two discrete-time chaotic systems, namely the Logistic and modified-Henon chaotic systems. The original video signal is encrypted using the Logistic map, which increases the transmission scheme robustness. The dynamics of the

modified-Henon system are used to include the encrypted message obtained. The receiver is a step by step delayed observer (dead-beat observer). Once the transmitter and receiver are synchronised, the video (message) can be recovered at the receiver side.

The advantage of separating the encryption and synchronisation tasks in order to design an amplitude independent scheme based on nonlinear chaotic systems was addressed in [13]. The proposed communication scheme is robust to channel noise and other disturbances, as demonstrated by the results in [13]. The exact synchronisation [14] method is used in this paper. The true state is recovered from a finite ring of delayed inputs and delayed outputs. This technique has many advantages, and it has demonstrated great efficiency in chaos communication applications.

This work is organised as follows. In Section 2, we present the proposed secure transmission scheme. In Section 3, simulation examples are given to show the good reconstruction of both state and video in a secure communication system. Finally, the paper ends with conclusion.

II. THE PROPOSED SECURE TRANSMISSION SCHEME

In this section, we present the proposed secure transmission scheme based on discrete-time chaotic system. The synoptic bloc diagram of the proposed video encryption/decryption scheme is displayed in Fig. 1. The complete procedure of the encryption-decryption algorithm is presented below.

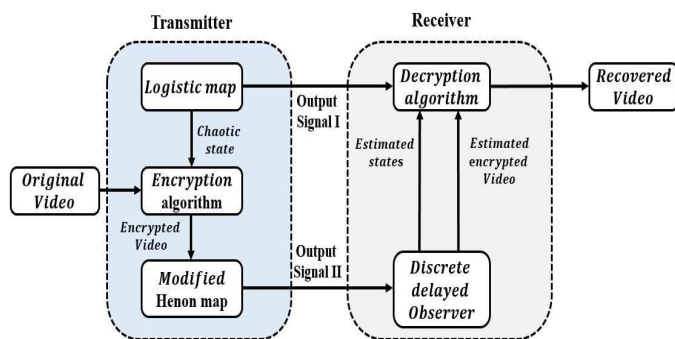


Fig. 1. Scheme of the proposed chaotic video encryption-decryption.

A. Original Video

Video files are made up of series of image frames, by a 2D array of pixels. Not only between adjacent pixels in each frame, but also between successive frames, video files have a high degree of correlation.

B. The Transmitter

The transmitter is composed of two different discrete-time systems: the Logistic map and the modified-Henon map.

The logistic map is a polynomial mapping that is frequently cited as an archetypal example of complex, chaotic behaviour can emerge from very simple nonlinear dynamical equation.

The map was popularised in 1976 by biologist Robert May [15], in part as a discrete-time demographic model analogous by Pierre Franois Verhulst logistic equation [16]. The logistic map is written mathematically as follows:

$$z(k+1) = rz(k)(1-z(k)) \quad (1)$$

The parameter $r = 3.62$ and the initial condition $z(0) = 0.5$. Under the condition $3.56994 < r \leq 4$, the researcher proves that the system is in a chaotic state. However, for almost all initial values of $z(0)$, r beyond 4 causes the value of $z(k)$ to leave the interval $[0, 1]$ and to diverge. $z(k)$ will behave differently depending on the value of r , as shown in Fig. 2

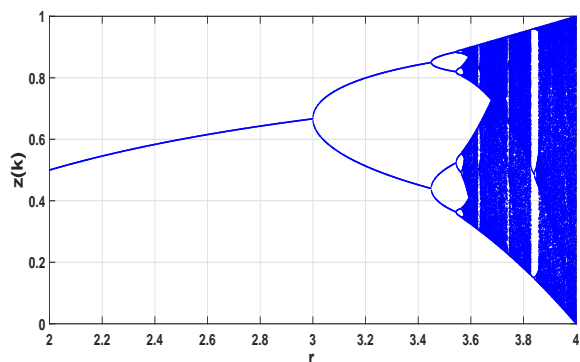


Fig. 2. Bifurcation diagram of the logistic map.

The modified-Henon system is selected. The state model of this system is presented as follows [17]:

$$\begin{cases} x_1(k+1) = a - x_2^2(k) - bx_3(k) \\ x_2(k+1) = x_1(k) \\ x_3(k+1) = x_2(k) \\ y(k) = x_2(k) \end{cases} \quad (2)$$

Where $x = [x_1 \ x_2 \ x_3]^T \in \mathbb{R}^3$ is the state vector and $y(k)$ is the output.

The chaotic behaviour of system (2) is exhibited for $a = 1.6$ and $b = 0.1$ and the initial conditions chosen inside the basin of attraction are: $x_1(0) = 0.1$, $x_2(0) = -0.1$, $x_3(0) = 0.1$.

The phase portrait of the states $x_1(k)$ vs $x_3(k)$ is presented in Fig. 3 and the states response in Fig. 4. The simulation results illustrate the chaotic behaviour of System (2).

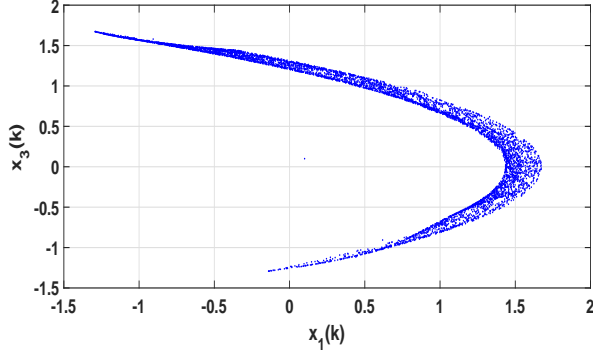


Fig. 3. Phase portrait of the states $x_1(k)$ vs $x_3(k)$ of the modified-Henon map.

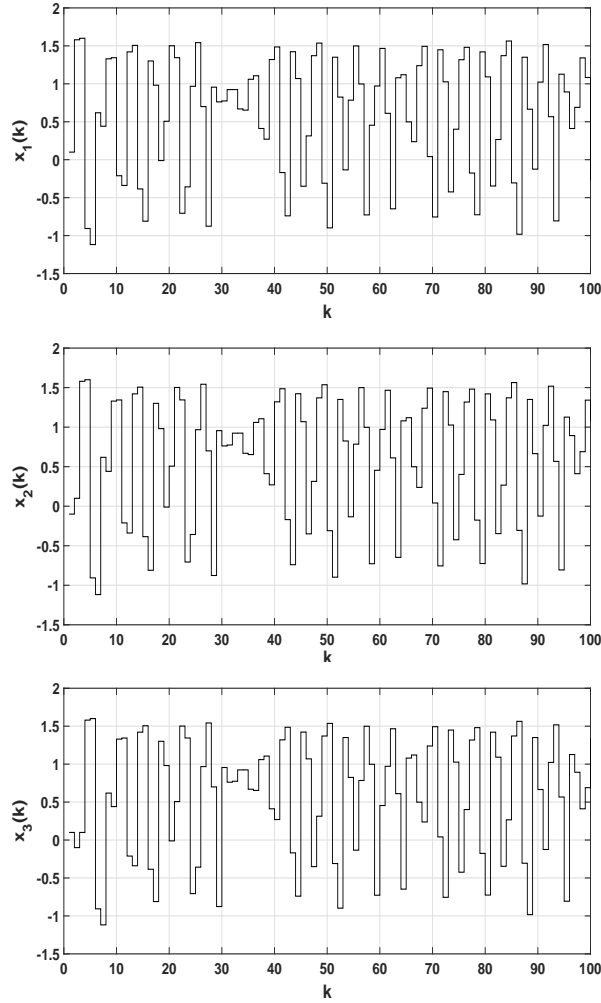


Fig. 4. State responses of the modified-Henon system.

C. The Encryption Algorithm

The original video (message) to send is encrypted in several steps to increase transmission security. The various treatments applied to the original message before sending it on the public channel are described below.

Step 1: We represent the original video frame by matrix $A_{M \times N \times 3}$ for grayscale, (M and N are the rows and columns of the frame). Then, we arrange, by order, the frame pixels from left to right and from top to bottom so as to form first a decimal set $A = \{A^1; A^2; \dots; A^f\}$, with f is the total number of frames.

Step 2: We iterate a Logistic map for $k = 1 : S$, with $S = M \times N \times 3$. We get, then, chaotic sequences $z(k)$. Then we apply chaotic confusion [18], where the process permutes a video frame with a chaotic map. The process is as follows:

$$B(k) = \text{Confusion}(A, z(k)).$$

The latter is subsequently transformed on a binary set, $C(k) = \text{de2bi}(B(k))$.

Step 3: Key generation [19]

Algorithm 1: Key $D(k)$

Initialisation:

$$P \leftarrow 3.628$$

$$K(1) \leftarrow 0.632$$

for $i = 1 \rightarrow S - 1$ **do**

$$\quad \lfloor K(i+1) = \cos(P * \arccos(K(i)))$$

$$\quad K_1 = \text{abs}(\text{round}(K(i) * 255)) \quad K_2 = \text{de2bi}(K_1)$$

$$\quad K_3 = \text{circshift}(K_2, 1)$$

$$\quad D(k) = K_1 \oplus K_3$$

Step 4: We apply a XOR function between the original video frame $C(k)$ and the encryption key $D(k)$. The ciphered set (encrypted video) is obtained as follows:

$$m_c(k) = C(k) \oplus D(k)$$

Step 5: We introduce the ciphered set $m_c(k)$ in the third dynamic component of the modified-Henon system to preserve the chaotic behaviour of [2].

Then, we have:

$$\begin{cases} x_1(k+1) = a - x_2^2(k) - bx_3(k) \\ x_2(k+1) = x_1(k) \\ x_3(k+1) = x_2(k) + m_c(k) \\ y(k) = x_2(k) \end{cases} \quad (3)$$

D. The Receiver

In this subsection, based on the works [14], [20], a step-by-step delayed observer is designed and used to resolve the synchronisation problem, and to allow the reconstruction of the states and the unknown input (message).

The step-by-step delayed observer is given below.
From (3), we have:

$$\hat{x}_2(k) = y(k) \quad (4)$$

Estimation of state $\hat{x}_1(k)$

From System (3) we have:

$$\hat{x}_2(k+1) = \hat{x}_1(k)$$

Then, we apply one step delay, we can deduce that:

$$\hat{x}_1(k-1) = y(k) \quad (5)$$

Estimation of state $\hat{x}_3(k)$

From System (3) we have:

$$\hat{x}_3(k) = \frac{a - \hat{x}_1(k+1) - \hat{x}_2^2(k)}{b}$$

Then, we apply two step delay, we can deduce that:

$$\hat{x}_3(k-2) = \frac{a - y(k) - y^2(k-2)}{b} \quad (6)$$

The unknown input $\hat{m}_c(k)$ of the system can be exactly reconstructed as follows:

$$\hat{m}_c(k) = \hat{x}_3(k+1) - \hat{x}_2(k)$$

By applying three steps delay, we deduce that:

$$\hat{m}_c(k-3) = \frac{a - y(k) - y^2(k-2)}{b} - y(k-3) \quad (7)$$

Subsequently, the dead-beat observer equations are given by:

$$\begin{cases} \hat{x}_1(k-1) = y(k) \\ \hat{x}_3(k-2) = \frac{a - y(k) - y^2(k-2)}{b} \\ \hat{m}_c(k-3) = \frac{a - y(k) - y^2(k-2)}{b} - y(k-3) \end{cases} \quad (8)$$

E. The Decryption Algorithm:

Step 1: We apply a XOR function between the estimated unknown input $\hat{m}_c(k)$ and the key $D(k)$. The decrypted set is obtained as follows:

$$Z(k) = \hat{m}_c(k) \oplus D(k)$$

The latter is subsequently transformed on a decimal set, $Y(k) = bi2de(Z(k))$.

Step 2: Then we apply chaotic inverse confusion, in the following process:

$$M = InverseConfusion(Y(k), z(k)).$$

Finally, the message (video) is recovered.

III. SIMULATION RESULTS

The proposed video encryption scheme is used to encrypt video data. The input videos considered are test videos consisting of 46 and 60 frames for the 1st and 2nd video, respectively.

The simulation results of the synchronisation of system (3) and its observer (8) are presented in Fig. 5. We see that the states $x_1(k), x_2(k), x_3(k)$ are well reconstructed.

Some frames of the original, encrypted and decrypted video are shown in Fig. 6. Their histograms are shown in Fig. 7. We found that the histogram of the ciphered video has approximately a uniform distribution and is completely different from that of the original video, which is in turn exact with the recovered video.

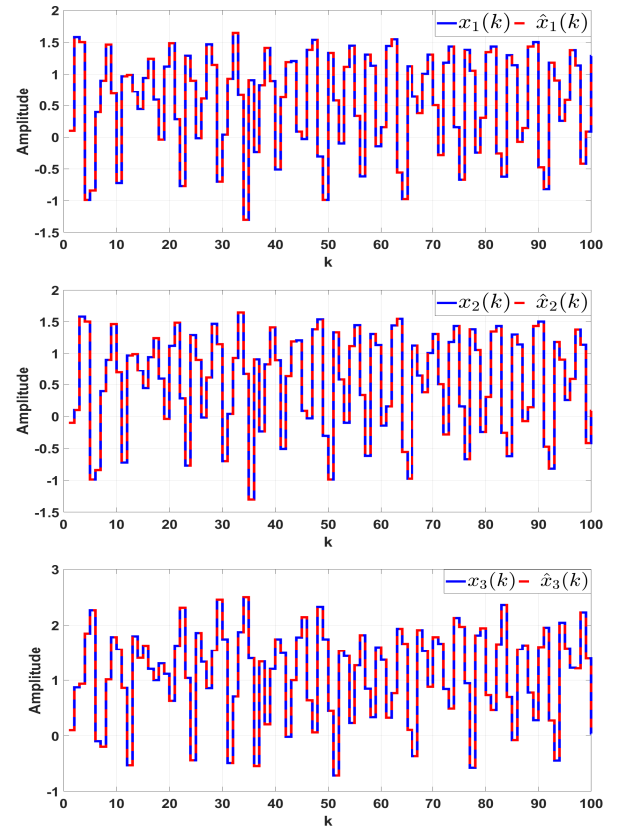


Fig. 5. State variables and their estimates.

For the second video, the original, encrypted and decrypted video frame are shown in Fig. 9. The histograms are shown in Fig. 10. We see that the histogram of the ciphered video has approximately a uniform distribution and is completely different from that of the original second video.

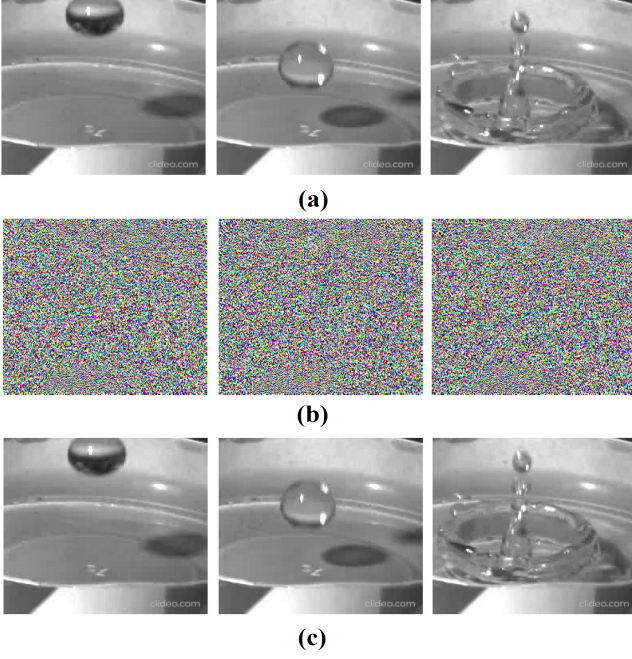


Fig. 6. The original video frame (a), the encrypted video frame (b) and the decrypted video frame (c) of first video.

Each pixel in an ordinary frame video is usually highly correlated with its adjacent pixels in vertical, horizontal, or diagonal directions, as determined by a correlation coefficient value close to 1 [21].

The correlation coefficient was used to evaluate the correlations between adjacent pixels of the original/encrypted video frame in each direction in order to evaluate the encryption quality of the proposed algorithm.

The correlation coefficient is defined as:

$$cov(x, y) = \frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2\right) \left(\frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2\right)}} \quad (9)$$

Where $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$. $(x_i; y_i)$ is the i^{th} pair of adjacent pixels in the same direction and N is the total number of pixel pairs.

The correlation distribution of two horizontally, vertically, and diagonally neighboring pixels in the original video and the encrypted video are shown in Fig. 8 for the first video and Fig. 9 for the second video. The results of the video correlation coefficients, for the first and second video, are presented in Table I and Table II respectively. As a result, the proposed scheme is highly secure to statistical attacks.

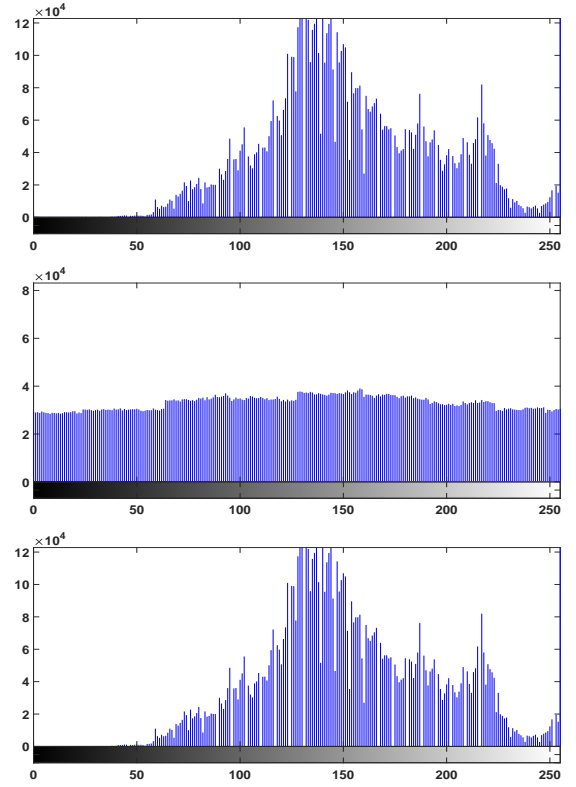


Fig. 7. The histograms of the original video (first line), the encrypted video (second line), and decrypted video (third line) of first video.

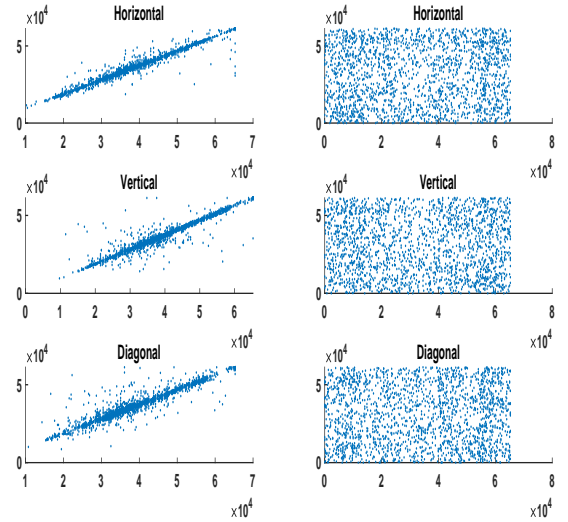


Fig. 8. The correlation among the pixels of the original and encrypted first video frames along the horizontal direction (first row), the vertical direction (second row), and the diagonal direction (third row).

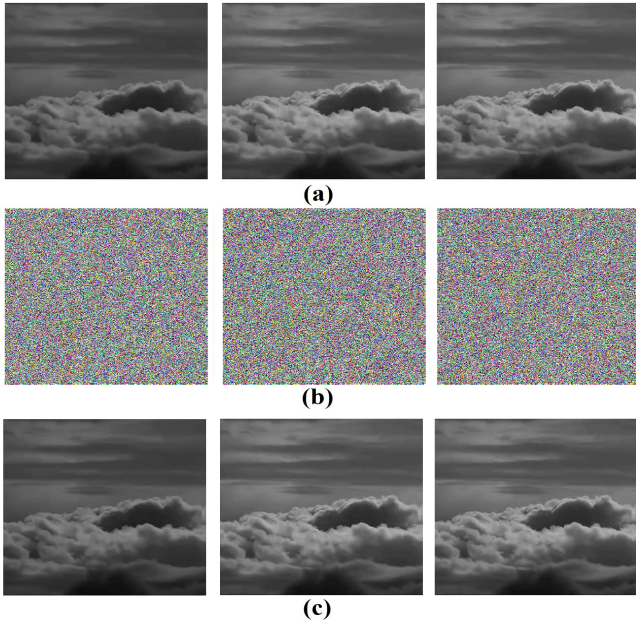


Fig. 9. The original video frame (a), the encrypted video frame (b) and the decrypted video frame (c) of second video.

TABLE I
CORRELATION COEFFICIENTS IN THE ORIGINAL FIRST VIDEO AND ENCRYPTED FIRST VIDEO.

	Original video	Encrypted video
Horizontal	0.9483	-1.7091×10^{-4}
Vertical	0.9576	-5.2380×10^{-4}
Diagonal	0.9367	5.4987×10^{-5}

TABLE II
CORRELATION COEFFICIENTS IN THE ORIGINAL SECOND VIDEO AND ENCRYPTED SECOND VIDEO.

	Original video	Encrypted video
Horizontal	0.9916	-3.7415×10^{-4}
Vertical	0.9986	-2.2917×10^{-4}
Diagonal	0.9899	2.6430×10^{-4}

The number of pixels change rate (NPCR) is one of the most common metrics used to assess the security of encryption algorithms and ciphers against different attacks [22]. A high NPCR score is commonly interpreted as a high level of resistance to different attacks.

Table III represents the NPCR percentage, between original and encrypted video of the proposed approach.

Tables IV and V show the NPCR of the original and recovered first video with various parameters and key, and we find that the NPCR is close to 100%, which demonstrates the proposed transmission scheme robustness.

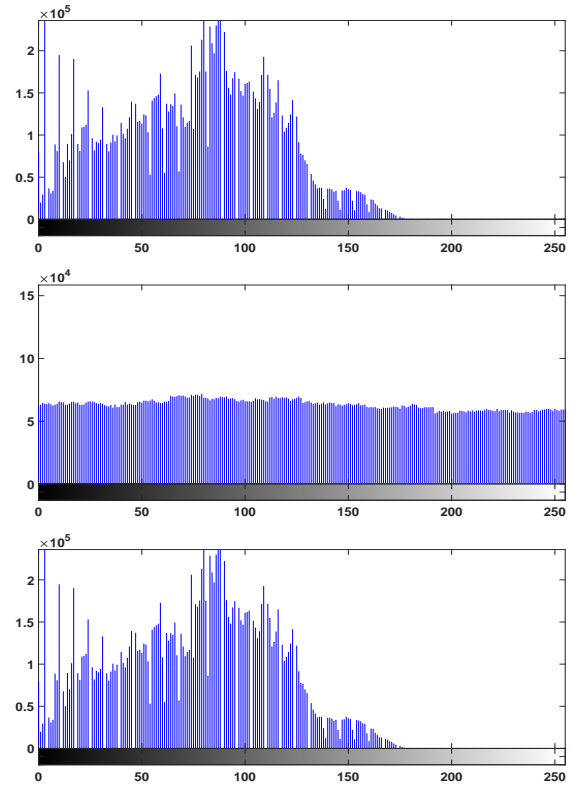


Fig. 10. The histograms of the original video (first line), the encrypted video (second line), and decrypted video (third line) of second video.

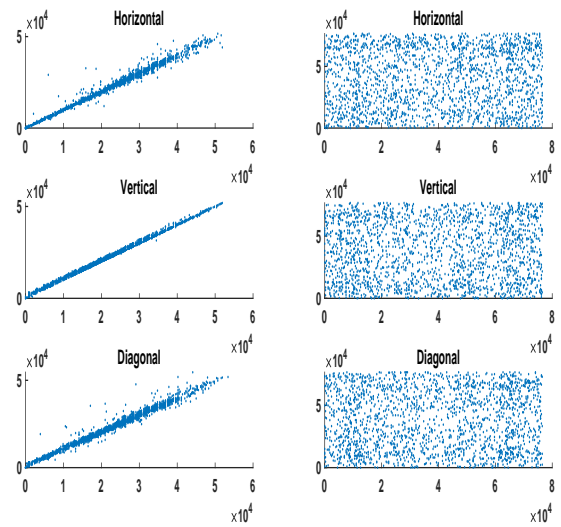


Fig. 11. The correlation among the pixels of the original and encrypted second video frames along the horizontal direction (first row), the vertical direction (second row), and the diagonal direction (third row).

TABLE III
NPCR FOR THE VIDEO SIMULATION BETWEEN ORIGINAL AND ENCRYPTED VIDEO.

Test video	NPCR(%)
1 st video	99.638
2 nd video	99.624

TABLE IV
NPCR OF ORIGINAL VIDEO AND RECOVERED VIDEO WITH DIFFERENT PARAMETERS.

Correct key	NPCR(%)
Incorrect parameter $a + 10^{-4}$	98.14
Incorrect parameter $b + 10^{-4}$	97.04
Incorrect parameters $a + 10^{-4}; b + 10^{-4}$	98.51

TABLE V
NPCR OF ORIGINAL VIDEO AND RECOVERED VIDEO WITH DIFFERENT KEY.

Correct parameters ($a; b$)	NPCR(%)
Incorrect key $K(1) + 10^{-16}$	99.52

IV. CONCLUSION

Video encryption plays an important role in today's multimedia world. Although many encryption schemes have been proposed to provide security for digital videos, some of them are too weak to resist various attacks designed by cryptanalysts. Basically, many efforts have been devoted to study the security issue, but for multimedia, the security is still not strong from a cryptographic point of view. To design truly secure video encryption schemes, chaotic cryptology must be employed.

So in this paper we presented a novel secure video transmission scheme based on a discrete time chaotic system. The video frame was permuted (confusion) at the transmitter and then encrypted using an encryption function. The keys used were generated by a chaotic system. The dead-beat observer was used at the receiver to determine the exact synchronisation between the two levels. The proposed scheme for video transmission has yielded good results. The simulation results in the previous section demonstrate this.

To improve the security of the transmission scheme, a fractional-order chaotic system should be used.

REFERENCES

[1] W Stallings. Web-based security protocols. In *Encyclopedia of Cryptography and Security*, pages 1383–1384. Springer US, 2011.

[2] L M Pecora and T L Carroll. Synchronization in chaotic systems. *Physical review letters*, 64(8):821, 1990.

[3] S Vaidyanathan and C Volos. *Advances and applications in chaotic systems*, volume 636. Springer, 2016.

[4] Q Huang, L Wang, and G Li. Research and application of video encryption technology based on chaotic synchronization theory. In *2018 10th International Conference on Measuring Technology and Mechatronics Automation (ICMTMA)*, pages 444–447. IEEE, 2018.

[5] H Elkamchouchi, W M Salama, and Y Abouelseoud. New video encryption schemes based on chaotic maps. *IET Image Processing*, 14(2):397–406, 2020.

[6] L Liu, S Zhao, Z Yu, and H Dai. A big data inspired chaotic solution for fuzzy feedback linearization model in cyber-physical systems. *Ad Hoc Networks*, 35:97–104, 2015.

[7] D Eroglu, J SW Lamb, and T Pereira. Synchronisation of chaos and its applications. *Contemporary Physics*, 58(3):207–243, 2017.

[8] H Nijmeijer and I M Y Mareels. An observer looks at synchronization. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 44(10):882–890, 1997.

[9] Y Wang, KW Wong, X Liao, and G Chen. A new chaos-based fast image encryption algorithm. *Applied soft computing*, 11(1):514–522, 2011.

[10] X Wang, L Teng, and X Qin. A novel colour image encryption algorithm based on chaos. *Signal Processing*, 92(4):1101–1108, 2012.

[11] M Khan and F Masood. A novel chaotic image encryption technique based on multiple discrete dynamical maps. *Multimedia Tools and Applications*, 78(18):26203–26222, 2019.

[12] S Li, G Chen, X Zheng, et al. Chaos-based encryption for digital image and video. *Multimedia Encryption and Authentication Techniques and Applications*, page 129, 2006.

[13] G Zheng, D Boutat, T Floquet, and J-P Barbot. Secure communication based on multi-input multi-output chaotic system with large message amplitude. *Chaos, Solitons & Fractals*, 41(3):1510–1517, 2009.

[14] A De Angeli, R Genesio, and A Tesi. Dead-beat chaos synchronization in discrete-time systems. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 42(1):54–56, 1995.

[15] R M May. Biological populations obeying difference equations: stable points, stable cycles, and chaos. *Journal of Theoretical Biology*, 51(2):511–524, 1975.

[16] Y Luo, J Yu, W Lai, and L Liu. A novel chaotic image encryption algorithm based on improved baker map and logistic map. *Multimedia Tools and Applications*, 78(15):22023–22043, 2019.

[17] SJ Sheela, KV Suresh, and D Tandur. Performance evaluation of modified henon map in image encryption. In *International Conference on Information Systems Security*, pages 225–240. Springer, 2016.

[18] A Qayyum, J Ahmad, W Boulila, S Rubaiee, F Masood, F Khan, W J Buchanan, et al. Chaos-based confusion and diffusion of image pixels using dynamic substitution. *IEEE Access*, 8:140876–140895, 2020.

[19] M H Khalid. *Multimedia security using chaotic maps: principles and methodologies*, volume 884. Springer Nature, 2020.

[20] S Djenoune, M Bettayeb, and U M Al-Saggaf. Synchronization of fractional-order discrete-time chaotic systems by an exact delayed state reconstructor: Application to secure communication. *International Journal of Applied Mathematics and Computer Science*, 29(1), 2019.

[21] M Kumar and A Chahal. Effect of encryption technique and size of image on correlation coefficient in encrypted image. *International Journal of Computer Applications*, 97(12), 2014.

[22] F Özkaynak. Role of npcr and uaci tests in security problems of chaos based image encryption algorithms and possible solution proposals. In *2017 International conference on computer science and engineering (UBMK)*, pages 621–624. IEEE, 2017.

An efficient WSN Localization Scheme using the Improved Chaos-GBO algorithm

Mohammed Amin Khelifa*, Badis Lekouaghet[§], Abdelkrim Boukabou[§]

**Department of Electronics, University of Tizi-Ouzou, BP 17, Tizi-Ouzou 15000, Algeria*

mohammedamin.khelifa@ummto.dz

[§]*LER, Department of Electronics, University of Jijel, BP 98 O.A., Jijel 18000, Algeria*

b.lekouaghet@univ-jijel.dz

aboukabou@univ-jijel.dz

Abstract—A wireless sensor network (WSN) is characterized by a set of connected sensing nodes with the principal purpose of managing the environmental conditions of a specific area of interest. Localizing the precise position of each WSN sensor node is very important to facilitate data routing over the different network nodes. In this paper, an improved meta-heuristic optimization algorithm based on the combination of the gradient-based optimizer (GBO) and a chaos search operator is proposed to solve the WSN node localization problem. The simulation results show that this proposed algorithm has a considerable potential to localize the WSN nodes with great precision. Furthermore, two important performance metrics, which are localization accuracy and computational time, are addressed to demonstrate the superiority of the proposed algorithm compared with two state-of-the-art optimization algorithms.

Index Terms—WSN, Localization, Meta-heuristic, GBO, Chaos.

I. INTRODUCTION

WSNs are being the subject of many real-world applications to monitor and manage different smart environments. Indeed, rapid developments in wireless communication systems and embedded devices have greatly improved the transmission of information from one point to another, and also the development of new network configurations using innovative solutions to link network nodes together. This trend has resulted in a number of developments including large-scale networks as well as in marketing communication [1]. Because of their wireless and dynamic nature, WSNs are facing the challenge of localizing the sensor nodes precisely since their position may not be in a predetermined location, and thus, will ultimately lead to a labor intensive task [2]. The simplest method to localize the different sensor nodes is to equip each one with a global positioning system (GPS) [3]. Surely this choice provided satisfactory results in some specific real-world situations; however, it has been shown that it was not very effective from a practical point of view in almost all indoor applications. Also, equipping every sensing node with GPS will lead to a design and dimensioning problem with energy consumption constraints and all that at high additive manufacturing costs. Many alternatives have been proposed to deal with this issue by combining GPS with localization-based algorithms, especially the geometric algorithms and the meta-heuristic algorithms.

The WSN node localization problem can be divided into two complementary stages. The first stage concerns the estimation of the distances between sensor nodes and their anchor nodes which is performed using analytical metrics, including the received signal strength indicator (RSSI), the angle of arrival (AoA), the time of arrival (ToA), and the time difference of arrival (TDoA). In the second stage, the localization process of the sensor node is estimated using the obtained local information in the previous stage where the estimation problem is solved by minimizing an objective function that corresponds to the localization error.

Meta-heuristic optimization algorithms have been extensively applied in the last two decades to solve various multidimensional optimization problems due to their fast convergence rates and higher accuracy [4]. In particular, some popular nature-inspired meta-heuristic optimization algorithms have been successfully employed to solve the WSN node localization problem such as genetic algorithm (GA) [5], particle swarm optimization (PSO) algorithm [6]– [9], the grey wolf optimization (GWO) algorithm [10], and the chicken swarm optimization (CSO) algorithm [11].

Of a large number of approaches to solve the optimization problems that have been proposed in the literature, the gradient-based optimizer (GBO) algorithm is a good alternative to search for the optimal solutions in terms of the convergence rate, performance, and space search ability [12]. Moreover, many recent works have proposed more advanced optimization techniques by combining the meta-heuristic optimization algorithms with different mathematics theories, especially the chaos theory. This is due to the fact chaotic systems are deterministic with unpredictable behavior, and thus, can be used as random-like sequences to support modern applications of parallel computing [13], [14]. For this purpose, this paper proposes an improved meta-heuristic optimization algorithm by combining the GBO algorithm with a chaotic sequence as a search operator to solve the WSN node localization problem. The chaotic sequence is generated using the well-known logistic equation to guarantee the diversity of the initial population, speed up the convergence rate, and enhance global search capability. The obtained results are compared with both CSO and PSO to demonstrate the effectiveness of this proposed improved algorithm. Obtained results confirmed that

the proposed chaos-GBO algorithm is quite competitive.

The rest of this paper is organized as follows. Section II presents the chaos-GBO optimization algorithm. The WSN sensor node localization formulation is presented in III. Simulation results and comparative studies are discussed in Section IV. Finally, the conclusion is drawn in Section V.

II. CHAOS-GBO OPTIMIZATION

A. Overview of GBO

GBO has been proposed recently by Ahmadianfar et al. [12] by combining the well-known Newton gradient method with the population-based concept. This combination gives GBO a more efficient optimization feature to converge faster towards the optimal solution, with the capability to escape from any local optimal solution. Besides the initialization process, GBO uses two operators, namely the Gradient Search Rule (GSR) and the Local Escaping Operator (LEO).

1) *Initialization process*: The initial population members in GBO are randomly generated as follows

$$X_{n,d} = [X_{n,1}, X_{n,2}, \dots, X_{n,D}] \quad (1)$$

$$X_n = X_{\min} + rand(0, 1) \times (X_{\max} - X_{\min}) \quad (2)$$

where $n = 1, 2, \dots, N$, $d = 1, 2, \dots, D$. The variables X_{\max} and X_{\min} are the border limits of the decision variable X and $rand(0, 1)$ is a randomly generated number in the range $[0, 1]$.

2) *The gradient search rule (GSR) operator*: GSR is used to accelerate the convergence rate and explore new search spaces using a random process with the aids of the Taylor series. The new position X_{n+1} is given by

$$X_{n+1} = X_n - \frac{2\Delta x \times f(X_n)}{f(X_n + \Delta x) - f(X_n - \Delta x)} \quad (3)$$

where $f(X_n + \Delta x)$ and $f(X_n - \Delta x)$ are the truncated derivatives of the function $f(x)$ whereas Δx is the change in position at each iteration.

Eq. (3) is changed to accommodate the population-based search concept as follows:

$$\begin{aligned} X_{n+1} &= X_n - randn \times \frac{2\Delta x \times X_n}{x_{worst} - x_{best} + \varepsilon} \\ &= X_n - GSR \end{aligned} \quad (4)$$

where $randn$ is a normally distributed random number, x_{worst} and x_{best} are the worst and best candidates' solutions through the process of optimization, respectively. ε is a small number arbitrarily chosen in the range of $[0, 0.1]$.

Further, the GSR is modified by introducing the parameter ρ_1 as follows

$$GSR = randn \times \rho_1 \times \frac{2\Delta x \times X_n}{x_{worst} - x_{best} + \varepsilon} \quad (5)$$

where parameter ρ_1 is used for the exploration process such that

$$\rho_1 = \alpha \times (2 \times rand - 1) \quad (6)$$

where

$$\alpha = \left| \beta \times \sin \left(\frac{3\pi}{2} + \sin \left(\beta \times \frac{3\pi}{2} \right) \right) \right| \quad (7)$$

$$\beta = 0.2 + \left(1 - \left(\frac{m}{M} \right)^3 \right)^2 \quad (8)$$

whereas m is the number of iterations, and M is the total number of iterations.

Note that Δx changes at each iteration between the best candidate solution x_{best} and a randomly selected position $x_{r_1}^m$ as follows:

$$\Delta X = rand(1 : N) \times |step| \quad (9)$$

$$step = \frac{(x_{best} - x_{r_1}^m) + \delta}{2} \quad (10)$$

$$\delta = 2 \times rand \times \left| \frac{x_{r_1}^m + x_{r_2}^m + x_{r_3}^m + x_{r_4}^m}{4} - x_n^m \right| \quad (11)$$

where $rand(1 : N)$ is a random number with N dimensions, $step$ represents a step size determined by x_{best} and $x_{r_1}^m$, and r_1, r_2, r_3, r_4 are different integers randomly chosen from the interval $[1, N]$ with $r_1 \neq r_2 \neq r_3 \neq r_4 \neq n$.

For better exploitation of the nearby area of X_n , a term called Direction of Movement (DM) is added, which is expressed by

$$DM = rand \times \rho_2 \times (x_{best} - x_n) \quad (12)$$

$$\rho_2 = \alpha \times (2 \times rand - 1) \quad (13)$$

Accordingly, the vector $X1_n^m$ is generated by updating x_n^m using both the GSR and DM as follows

$$\begin{aligned} X1_n^m &= x_n^m - GSR + DM \\ &= x_n^m - randn \times \rho_1 \times \frac{2\Delta x \times x_n^m}{x_{worst} - x_{best} + \varepsilon} \\ &\quad + rand \times \rho_2 \times (x_{best} - x_n) \end{aligned} \quad (14)$$

The new solution at the next iteration x_n^{m+1} is defined by

$$x_n^{m+1} = r_a \times (r_b \times X1_n^m + (1 - r_b) \times X2_n^m) + (1 - r_a) \times X3_n^m \quad (15)$$

where r_a, r_b are two random numbers, the vector $X1_n^m$ is given by Eq. (14), and the two vectors $X2_n^m$ and $X3_n^m$ are generated as follows:

$$\begin{aligned} X2_n^m &= x_{best} - randn \times \rho_1 \times \frac{2\Delta x \times x_n^m}{yp_n^m - yq_n^m + \varepsilon} \\ &\quad + rand \times \rho_2 \times (x_{r_1}^m - x_{r_2}^m) \end{aligned} \quad (16)$$

$$X3_n^m = X_n^m - \rho_1 \times (X2_n^m - X1_n^m) \quad (17)$$

whereas yp_n^m and yq_n^m are two positions created in regard two variables z_{n+1} and x_n as follows

$$yp_n = rand \times \left(\frac{[z_{n+1} + x_n]}{2} + rand \times \Delta x \right) \quad (18)$$

$$yq_n = rand \times \left(\frac{[z_{n+1} + x_n]}{2} - rand \times \Delta x \right) \quad (19)$$

3) *The local escaping operator (LEO)*: This operator aims to boost the ability of GBO to solve complex problems. Accordingly, a solution with superior performance, X_{LEO}^m , is generated using the best position, X_{best} , two solutions, $X1_n^m$ and $X2_n^m$, two random solutions from the population, X_{r1}^m and X_{r2}^m , and a new randomly generated solution, x_k^m . The solution X_{LEO}^m is updated based on the following layout:

```

if rand < pr
  if rand < 0.5
     $X_{LEO}^m = X_n^{m+1} + f_1 \times (u_1 \times x_{best} - u_2 \times x_k^m)$ 
     $+ f_2 \times \rho_1 \times (u_3 \times (X2_n^m - X1_n^m) + u_2 \times (x_{r1}^m - x_{r2}^m))/2$ 
     $X_n^{m+1} = X_{LEO}^m$ 
  else
     $X_{LEO}^m = X_{best} + f_1 \times (u_1 \times x_{best} - u_2 \times x_k^m)$ 
     $+ f_2 \times \rho_1 \times (u_3 \times (X2_n^m - X1_n^m) + u_2 \times (x_{r1}^m - x_{r2}^m))/2$ 
     $X_n^{m+1} = X_{LEO}^m$ 
  End
End

```

where f_1, f_2 are uniform distributed random numbers in the range $[-1, 1]$, pr is a probability value, and u_1, u_2, u_3 are three random numbers introduced to increase the diversity of the population and escape from local optimal solutions, expressed by

$$u_1 = \begin{cases} 2 \times rand & \text{if } \mu_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (20)$$

$$u_2 = \begin{cases} rand & \text{if } \mu_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (21)$$

$$u_3 = \begin{cases} rand & \text{if } \mu_1 < 0.5 \\ 1 & \text{otherwise} \end{cases} \quad (22)$$

Or simply by

$$u_1 = L_1 \times 2 \times rand + (1 - L_1) \quad (23)$$

$$u_2 = L_1 \times rand + (1 - L_1) \quad (24)$$

$$u_3 = L_1 \times rand + (1 - L_1) \quad (25)$$

where μ_1 is a random parameter in the range $[0, 1]$ and L_1 is a binary parameter such that $L_1 = 1$ for $\mu_1 < 0.5$ and $L_1 = 0$ otherwise.

The solution x_k^m in Eq. (6) is generated using a random parameter $\mu_2 \in [0, 1]$ as follows

$$x_k^m = \begin{cases} x_{rand} & \text{if } \mu_2 < 0.5 \\ x_p^m & \text{otherwise} \end{cases} \quad (26)$$

where $x_{rand} = X_{\min} + rand \times (X_{\max} - X_{\min})$ is a new solution and x_p^m is a randomly selected solution of the population ($p \in [1, 2, \dots, N]$).

Eq. (26) can be simplified as:

$$x_k^m = L_2 \times x_p^m + (1 - L_2) \times x_{rand} \quad (27)$$

where L_2 is a binary parameter such that $L_2 = 1$ for $\mu_2 < 0.5$ and $L_2 = 0$ otherwise.

B. Proposed Chaos-GBO

In order to improve the performance of GBO and ensure that the obtained solution converges rapidly towards the optimal solution, the direction of movement (DM) part of GSR is modified by introducing the chaotic logistic equation. Accordingly, the DM is changed as follows

$$\overline{DM} = rand \times \bar{\rho}_2 \times (x_{best} - x_n) \quad (28)$$

$$\bar{\rho}_2 = \alpha \times y_k \quad (29)$$

where the parameter α is given by Eq. (7), and y_k is generated by a chaotic map. Let this chaotic map be represented by the logistic equation with the following recursive form

$$y_{k+1} = 4y_k(1 - y_k) \quad (30)$$

The variation of the modified parameter $\bar{\rho}_2$ using the logistic equation is shown in Fig. 1. Obviously, it can be seen that the chaotic equation guarantees a good compromise between the exploitation and exploration phases.

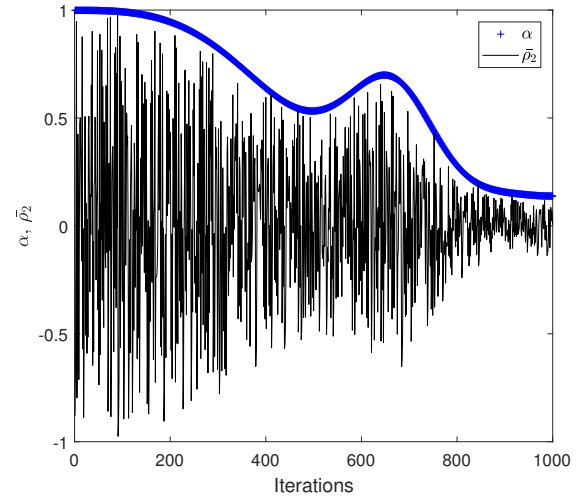


Fig. 1. Modified parameter $\bar{\rho}_2$ using the logistic equation.

Taking into account the modification given by Eq. (28), the new solution at the next iteration x_n^{m+1} is expressed by Eq. (15) where the vectors $X1_n^m$, $X2_n^m$ and $X3_n^m$ are generated as follows:

$$\begin{aligned} X1_n^m &= x_n^m - GSR + \overline{DM} \\ &= x_n^m - randn \times \rho_1 \times \frac{2\Delta x \times x_n^m}{x_{worst} - x_{best} + \varepsilon} \\ &\quad + rand \times \bar{\rho}_2 \times (x_{best} - x_n) \end{aligned} \quad (31)$$

$$\begin{aligned} X2_n^m &= x_{best} - randn \times \rho_1 \times \frac{2\Delta x \times x_n^m}{yp_n^m - yq_n^m + \varepsilon} \\ &\quad + rand \times \bar{\rho}_2 \times (x_{r1}^m - x_{r2}^m) \end{aligned} \quad (32)$$

$$X3_n^m = X_n^m - \rho_1 \times (X2_n^m - X1_n^m)$$

III. WSN NODE LOCALIZATION PROBLEM

A. Problem formulation

The WSN node localization problem aims to estimate the location of n target nodes using prior information about the location of m anchors and the distance between the target nodes and the neighboring anchors.

Let (x_i, y_i) be the coordinates of the anchor, given by

$$x_i = [x_1, x_2, \dots, x_m], y_i = [y_1, y_2, \dots, y_m], \quad (33)$$

Let (x_j, y_j) be the coordinates of the target nodes to be estimated such that

$$x_j = [x_1, x_2, \dots, x_n], y_j = [y_1, y_2, \dots, y_n], \quad (34)$$

In the first stage, each anchor node will estimate the distance from its position to the neighboring target nodes. The effect of measurement noise is simulated as an additive white Gaussian noise (AWGN). Thus, the estimated distance from a sensor to an anchor i is given by

$$\hat{d}_i = d_i + n_i \quad (35)$$

where d_i is the actual distance and n_i is the AWGN uniformly distributed according to the estimated distance \hat{d}_i , which can be described as: $[d_i - d_i \times P_n/100, d_i + d_i \times P_n/100]$.

The distance d_i is expressed as

$$d_i = \sqrt{(x - x_i)^2 + (y - y_i)^2} \quad (36)$$

Here, (x, y) is the location of target node and (x_i, y_i) is the location of the neighboring anchor.

B. The objective function

Since the localization of WSN sensor nodes is considered as an optimization problem of continuously differentiable functions with boundary constraints, therefore, the objective is to minimize an objective function to guarantee that the obtained distances reach the best solution within the applied algorithm. For a precise localization of the used sensor nodes, the objective function is expressed by the following localization error function

$$f(x, y) = \frac{1}{M} \sum_{i=1}^M (\sqrt{(x - x_i)^2 + (y - y_i)^2} - \hat{d}_i)^2 \quad (37)$$

where (x, y) is the location of the target node, (x_i, y_i) is the location of the neighboring anchor, \hat{d}_i is the estimated distance, and $M \geq 3$ is the number of anchors within the transmission radius of target nodes (x, y) .

Accordingly, each WSN sensor node is localized using the localization error function at only the condition that this node has at least three neighboring anchors.

In addition, the absolute error (AE) for each estimated WSN sensor node is given by

$$AE_x = |x - x_i| \quad (38)$$

$$AE_y = |y - y_i| \quad (39)$$

IV. SIMULATION RESULTS AND DISCUSSION

The simulations are mainly performed in a $100m \times 100m$ wireless sensor environment, where only the measured distances between anchor and target nodes are used to estimate the location of the sensor nodes through multilateration, i.e., the time difference between signals arriving at multiple base stations [15]. Accordingly, there are N anchor nodes and M target nodes in the WSN system. The following parameters are considered for the localization of the target nodes using the proposed chaos-GBO optimization technique: number of particles (population size) $nP = 30$, number of iterations $MaxIt = 100$, and the initial deployment of target nodes is randomly generated. Simulated WSN localization result of 40 target nodes with 8 anchor nodes is illustrated in Fig. 2. It is clear that the estimated nodes match the target nodes with a large degree of accuracy. This can be confirmed by the achieved RMSE value (RMSE= 0.059918 meter).

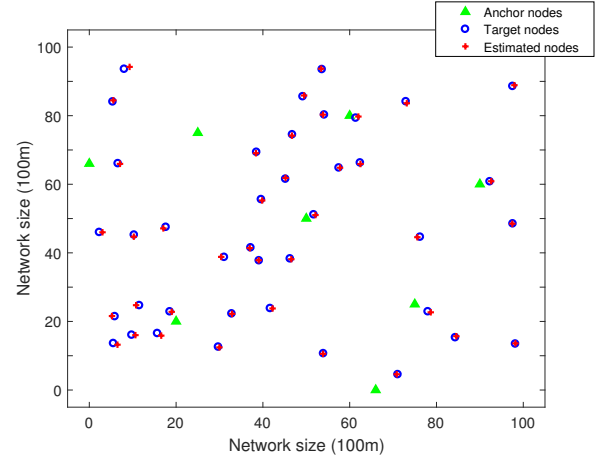


Fig. 2. Localization of target nodes using Chaos GBO.

Further simulations have been carried out to show the effectiveness of the proposed chaos-GBO algorithm by studying the effect of the number of anchor nodes (N), the number of iterations ($MaxIt$), the size of population (nP), and the AWGN. Accordingly, four cases are investigated throughout this section as follows.

Case 1: The effect of the number of anchor nodes N is shown in Fig. 3, where the target nodes, the number of iterations and the number of population are fixed at $M = 20$, $MaxIt = 100$, and $nP = 30$, respectively.

Case 2: The effect of the maximum number of iterations $MaxIt$ is depicted in Fig. 4, for which the number of anchor nodes, the target nodes, and the population size are fixed at $N = 8$, $M = 40$ and $nP = 30$, respectively.

Case 3: The effect of the number of population is given in Fig. 5. Here, the size of population nP is varied from 10, 20, 30 to 40, and the numbers of anchor nodes, target nodes and max iterations are fixed to $N = 8$, $M = 40$, and $MaxIt = 100$, respectively.

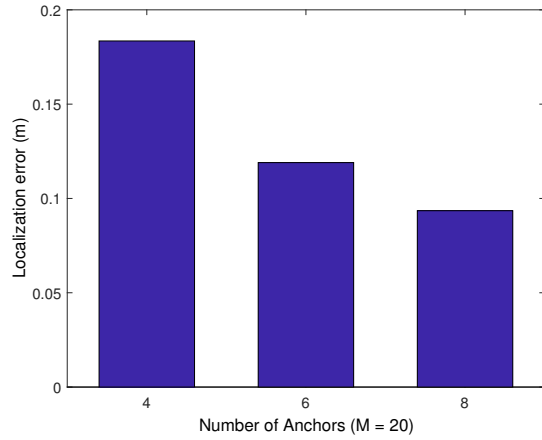


Fig. 3. Localization error vs anchor nodes number.

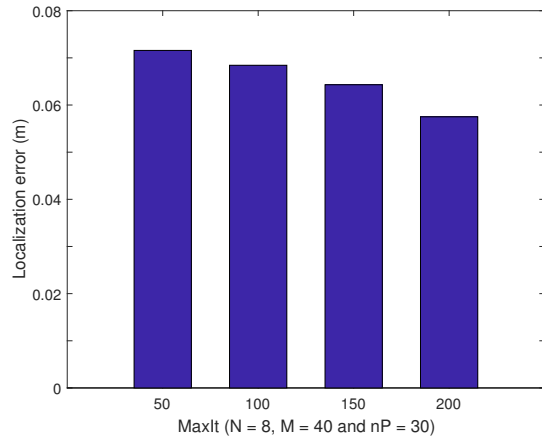


Fig. 4. Localization error vs maximum of iterations.

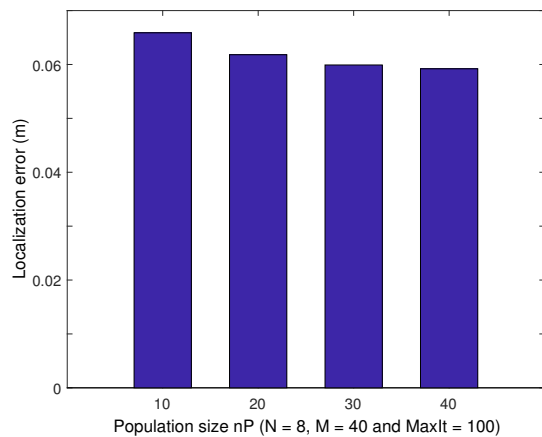


Fig. 5. Localization error vs size of population.

Case 4: In this last case, the effect of the AWGN is illustrated in Fig. 6 for different values of AWGN, i.e., $P_n = 5$, $P_n = 10$, $P_n = 15$, and $P_n = 20$. The anchor nodes, target nodes, population size and max iteration are fixed at $N = 8$, $M = 40$, $nP = 30$, and $MaxIt = 100$, respectively.

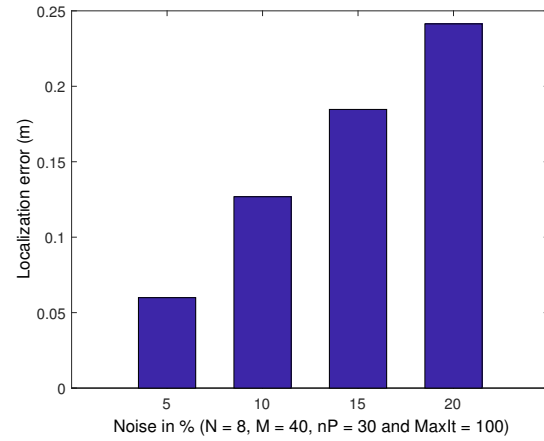


Fig. 6. Localization error vs rate of noise.

Note that the AWGN value was fixed to $P_n = 5$ in the first three cases.

Based on the obtained results, it is clearly shown that the performance of the localization error depends considerably on the four parameters mentioned before. Thus, the localization error of target nodes is inversely proportional to the number of anchor nodes, the number of iterations, and the population size. Also, it is obvious that the most sensitive parameter for the WSN localization process is the AWGN. Finally, the obtained results using chaos-GBO are compared with CSO and PSO algorithms as given in Tabs. I, II, and III in terms of the number of anchor nodes, the AWGN, and the computation time, respectively. Obviously, the proposed chaos-GBO algorithm is quite competitive for all performed results.

TABLE I
ANALYSIS OF ANCHOR NODES USING CHAOS-GBO, CSO AND PSO. RESULTS OBTAINED FOR $M = 40$, $P_n = 2$, $MaxIt = 150$.

Algorithm	number of anchor nodes N	RMSE (m)
chaos-GBO	10	0.0249
	12	0.0229
CSO [11]	10	0.11
	12	0.08
PSO	10	0.38
	12	0.315

V. CONCLUSION

This paper addressed the WSN localization problem using an improved meta-heuristic algorithm. For this purpose, a combined chaos gradient-based optimizer method (chaos-GBO) has been developed to estimate the position of the unknown

TABLE II
NOISE COMPARISON USING CHAOS-GBO, CSO, AND PSO.
RESULTS OBTAINED FOR $N = 10$, $M = 40$, $P_n = 2$, $MaxIt = 150$.

Algorithm	Noise %	RMSE (m)
chaos-GBO	2	0.0249
	3	0.0322
	4	0.0517
	5	0.0612
CSO [11]	2	0.11
	3	0.15
	4	0.23
	5	0.31
PSO	2	0.315
	3	0.414
	4	0.550
	5	NA

TABLE III
SUMMARY OF COMPUTATION TIME USING CHAOS-GBO, CSO,
AND PSO. RESULTS OBTAINED FOR $N = 10$, $M = 40$, $P_n = 2$,
 $MaxIt = 150$.

Algorithm	Computation time	RMSE (m)
chaos-GBO	126.1358	0.0249
CSO [11]	255.0341	0.11
PSO	874.24	0.38

sensor nodes by only considering the neighboring anchors. The obtained results illustrate the effectiveness of this proposed algorithm to localize the WSN nodes accurately. Further, the effectiveness of this proposed algorithms has been demonstrated when compared with two optimization algorithms in terms of the number of anchor nodes, the AWGN, and the computation time. In fact, the computational time taken by the proposed chaos-GBO is less than other algorithms, which means the obtained solution converges rapidly to the optimal solution when the DM that is part of GSR is modified by introducing the chaotic logistic equation.

REFERENCES

- [1] R. Dobrescu and F. Ionescu, Large Scale Networks. CRC Press, 2016.
- [2] A. Boukerche, H. A. B. F. Oliveira, E. F. Nakamura, and A. A. F. Loureiro, "Localization Systems for Wireless Sensor Networks." *Algorithms and Protocols for Wireless Sensor Networks*, pp. 307-340, 2008.
- [3] W. Ren, "A Rapid Acquisition Algorithm of WSN-aided GPS Location." *Second International Symposium on Intelligent Information Technology and Security Informatics*, pp. 42-46, Jan. 2009.
- [4] P. M., Vasant, Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance. IGI Global, 2012.
- [5] B. Peng and L. Li, "An improved localization algorithm based on genetic algorithm in wireless sensor networks." *Cognitive Neurodynamics*, vol. 9, no. 2, pp. 249-256, 2015.
- [6] A. Gopakumar and L. Jacob, "Localization in wireless sensor networks using particle swarm optimization." *IET Conference on Wireless, Mobile and Multimedia Networks*, pp. 227-230, 2008.
- [7] A. Kumar, A. Khosla, J. S. Saini, and S. Singh, "Computational Intelligence based algorithm for node localization in Wireless Sensor Networks." *6th IEEE International Conference Intelligent Systems*, pp. 431-438, 2012.
- [8] R. Stoleru and J. Stankovic, "Probability grid: a location estimation scheme for wireless sensor networks." *First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks (IEEE SECON)*, pp. 430-438, 2004.
- [9] P.-J. Chuang and C.-P. Wu, "An Effective PSO-Based Node Localization Scheme for Wireless Sensor Networks." *Ninth International Conference on Parallel and Distributed Computing, Applications and Technologies*, pp. 187-194, 2008.
- [10] R. Rajakumar, J. Amudhavel, P. Dhavachelvan, and T. Vengattaraman, "GWO-LPWSN: Grey Wolf Optimization Algorithm for Node Localization Problem in Wireless Sensor Networks." *Journal of Computer Networks and Communications*, vol. 2017, pp. 1-10, 2017.
- [11] M. Al Shayokh and S. Y. Shin, "Bio Inspired Distributed WSN Localization Based on Chicken Swarm Optimization." *Wireless Personal Communications*, vol. 97, no. 4, pp. 5691-5706, 2017.
- [12] I. Ahmadianfar, O. Bozorg-Haddad, and X. Chu, "Gradient-based optimizer: A new metaheuristic optimization algorithm." *Information Sciences*, vol. 540, pp. 131-159, 2020.
- [13] A. Senouci, H. Bouhedjeur, K. Tourche, and A. Boukabou, "FPGA based hardware and device-independent implementation of chaotic generators." *AEU - International Journal of Electronics and Communications*, vol. 82, pp. 211-220, 2017.
- [14] M. A. Khelifa and A. Boukabou, "Design of an intelligent prediction-based neural network controller for multi-scroll chaotic systems." *Applied Intelligence*, vol. 45, pp. 793-807, 2016.
- [15] Y. Zhou, L. Jun Li, and L. Louise, "Multilateration localization in the presence of anchor location uncertainties." *IEEE Global Communications Conference (GLOBECOM)*, pp. 309-314, 2012.

CONTRIBUTION TO AUTOMATIC TOMATOES SORTING ACCORDING TO THEIR MATURITY USING AN EVOLUTIONARY CLASSIFICATION METHOD

Salim Iratni¹, Ouiza Boukendour², Souhila Boumati³, Moussa Diaf⁴

^{1,2,3,4}Laboratoire Vision Artificielle et Automatique des Systèmes (LVAAS);
Mouloud Mammeri University (UMMTO); Tizi-Ouzou – Algeria

¹salim.iratni@ummto.dz

²ouiza.boukendour@ummto.dz

³gh_souhila@yahoo.fr

⁴moussa.diaf@ummto.dz

Abstract—In pattern recognition, the process consists, first of all, in extracting relevant features from the images of the objects in order to perform a classification method for grouping the objects in their corresponding classes and then, assign the unknown object to one of these classes. However, some objects evolve over time, and their shapes, colors or textures change. In this case, classification becomes complex and the usual methods lose their performances and can become even inefficient. To take this evolution into account, other automatic classification approaches, called dynamic classification, have been developed. In the present work, we are interested by an important application which consists of estimating the maturity level of tomatoes which are distinguished by the change of color over time. This application can take effect in South Algeria which is known for an extensive production of tomatoes of different qualities. To do this, we applied the k nearest fuzzy neighbors method and the $L^*a^*b^*$ space representation. The obtained results were very satisfactory.

Keywords— *Evolutionary classification, color spaces, DFKNN, tomatoes maturity,*

I. INTRODUCTION

The problem of automatic recognition of natural objects is often more complex than in the case of manufactured objects, especially when their attributes (color, texture, shape) change with time and space. The classification needed to this type of objects must be dynamic to better adapt to evolutionary classes. Let us recall that objects are dynamic if they are in perpetual evolution. This evolution can include four cases. The first one is when the objects are static and their corresponding classes are also static. This problem is widely addressed in the literature as for example the automatic sorting of manufactured products. The second case is when the objects are dynamic and their classes are static, in other words, the objects move but their characteristics remain unchanged. In the third case, the objects are static and their classes dynamic or evolutionary, which means that the objects are immobile but their characteristics change over time. The fourth case corresponds to the situation where the objects, as well as their classes, are dynamic, *ie* the objects are moving

and their characteristics change over time. In this paper, we are interested by the third case. To perform pattern recognition on this type of evolutionary data, two approaches have been proposed in the literature [1]. The first one consists in using dynamic classification algorithms that implement the notions of moving, eliminating, merging and splitting classes, while the other concerns periodic evolutions and consists in dividing the period of the evolution into intervals, which set us back to consider the classes as static. Let us note that the evolution of a dynamic class can be translated by a displacement of a part or the totality of its elements. Thus, it can result in its merging with other classes or its splitting into several new classes. In the case of merging, one or more initially disjointed classes can join and share data. These data called ambiguous data carry mutual information. In order to remove this ambiguity, the dynamic classification method merges the concerned classes in order to form a single class. Concerning splitting classes, the opposite phenomenon of merging can occur, *i.e.* a class can split into two or more classes. On the other hand, the evolution of classes can also lead to their deletion if they are obsolete, of small sizes or made up of noise. In order to take into account these evolutions, different dynamic recognition methods have been developed. Among these methods, we can mention the connectionist methods, the kernel methods, the Adaptive Fuzzy C-Means method (AFCM) and the Dynamic Fuzzy K-Neighbors method (DFKNN). In connectionist methods with adaptive architecture, the number of neurons and the number of connections vary. This architecture can be constructive or destructive. It is constructive when neurons and connections are added and it is destructive when the useless neurons corresponding to non-representative classes are pruned [2]. Among the proposed constructive neural architecture methods, we can mention CDL (Cluster Detection and Labeling) [3][4] where the similarity notion of observations with respect to the initial predefined prototypes are used. Thus, it is according to this CDL value that new classes and then new neurons are created. In the category of kernel methods, the most used is certainly the SVM method [5]

and its dynamic version proposed in [6]. As for the Fuzzy C-Means Adaptive method (FCMA) [7], its principle is the same as that of the static one [8] but with the on-line integration of new data and the update of the characteristics of the classes, in an incremental way according to the evolution of the entropy. This entropy is calculated for all classes, which can lead to the creation of new classes. The DFKNN is another dynamic classification method that L. Harter proposed [1] based on FKNN [9] in which he added similarity and validity measures to merge, split or delete some classes [10][11][12]. This is the method we have adopted in the present work.

II. DYNAMIC FUZZY K NEAR NEIGHBOUR METHOD

The DFKNN method proceeds in several phases. From initial data, the DFKNN method allows, first of all, to carry out the classification of each of the new forms. In a second phase, as proposed in [1], the method updates the parameters to follow the evolution of the classes. When a change is detected and confirmed by several indicators, the classes are adopted, but only the most representative ones are retained. The different steps of this method are detailed as follows. In the learning phase, the user gives the initial classes C_i , $i=1,2,\dots,K$. Thus, the gravity centers and the initial standard deviation of each one of these classes C_i are calculated. During class evolution detection, the class C_i that receives a new object is the only one that needs to be updated, in other words, the values of the standard deviation ($\sigma_i^{j\text{Actu}}$) and the gravity center ($CG_i^{j\text{Actu}}$) of C_i are incrementally recomputed for each attribute j using the following expressions (1) and (2) :

$$\sigma_i^{j\text{Actu}} = \sqrt{\frac{N_i - 1}{N_i} \times (\sigma_i^{j\text{Actu}-1})^2 + \frac{(x^j - CG_i^{j\text{Actu}-1})^2}{N_i + 1}} \quad (1)$$

$$CG_i^{j\text{Actu}} = \frac{CG_i^{j\text{Actu}-1} \times N_i}{N_i + 1} + \frac{x^j}{N_i + 1} \quad (2)$$

where N_i is the number of forms present in C_i .

To track temporal changes in the system based on the calculated values of ($\sigma_i^{j\text{Init}}$), ($CG_i^{j\text{Actu}}$) and ($\sigma_i^{j\text{Actu}}$), two indicators ind_{1j} and ind_{2j} are used.

The indicator ind_{1j} which represents the evolution of class compactness is calculated from the deviation between ($\sigma_i^{j\text{Actu}}$) and ($\sigma_i^{j\text{Init}}$) for each attribute j using the following expression:

$$ind_{1j} = \frac{\sigma_i^{j\text{Actu}} \times 100}{\sigma_i^{j\text{Init}}} - 100 \quad (3)$$

This indicator is expressed in percent. If, at least, one of the attributes j receives a value of ind_{1j} higher than a given threshold $th1$, it means that the characteristics of the class C_i have changed. This threshold $th1$ is usually set at a small value of the order of 5. On the contrary, when the class changes abruptly, a larger value of this threshold is required.

As for the ind_{2j} indicator, it represents the deviation of one point from the average dispersion of the class. It is the distance between the j th attribute and the current center of gravity ($CG_i^{j\text{Actu}}$) as a function of the current standard deviation ($\sigma_i^{j\text{Actu}}$). This indicator is calculated by the following expression:

$$ind_{2j} = \frac{|x^j - CG_i^{j\text{Actu}}| \times 100}{\sigma_i^{j\text{Actu}}} - 100 \quad (4)$$

Like the previous one, this indicator is given in percent. If at least one of the attributes j obtains a value of ind_{2j} greater than the threshold $th1$ ($\max(ind_{2j}) \geq th1$), this indicates that the value of membership of this point in the class C_i is low. However, a single point away from the class does not necessarily mean a change in characteristics. Indeed, this point may be noise. Therefore, it is necessary to define a threshold $NbMin$ representing the number of successive times that ind_{2j} must exceed $th1$ to confirm a change. If $NbMin$ is set to a large value, the evolution of the class can be detected after a large delay. This number $NbMin$ must be defined according to a compromise between the noise present in the shapes and the maximum delay of detection of evolution desired. The evolution of the class is confirmed when $NbMin$ successive values of the two indicators ind_{1j} and ind_{2j} are greater than $th1$.

In the adaptation phase, the DFKNN method is used to adjust the parameters of a class that has evolved when enough feature changes have been detected and confirmed in the detection phase. Thus, the new center of gravity ($CG_i^{j\text{Actu}}$) and the initial standard deviation of ($\sigma_i^{j\text{Init}}$) of the class are calculated according to the previous equations (1) and (2).

In the class validation phase, the DFKNN method performs the operations of deleting outliers and non-representative classes, merging classes that have strong similarity and splitting classes that can no longer be considered as a single class.

The deletion of non-representative classes is performed if an insufficient number $n1$ of shapes is contained in the class ($n1 > k$) or if no shape has been classified in the class since a sufficient number $n2$ of shapes is classified in other classes.

Class merging is performed when these classes evolving towards the same direction overlap sufficiently at a given time.

To decide on this merger, a similarity measure δ_{iz} between the two classes C_i and C_z proposed in [13] is computed for all pairs of classes after the classification of each new shape. This measure is given by the expression (5). If this measure exceeds a threshold th between two classes then they are merged.

$$\delta_{iz} = 1 - \frac{\sum_{x \in C_i \text{ and } x \in C_z} |\pi_i(x) - \pi_z(x)|}{\sum_{x \in C_i} \pi_i(x) + \sum_{x \in C_z} \pi_z(x)} \quad (5)$$

where $\pi_i(X)$ and $\pi_z(X)$ are the membership values of x in C_i and C_z , respectively. The closer δ_{iz} is to 1, the more similar the two classes are. The maximum value represents two

completely overlapped classes so there is no need to wait until the obtained similarity value is equal to 1 to merge these two classes.

Regarding the detection of splitting, the operation is more complex. To detect the split of a class, we can apply the well-known Fuzzy-C-Means (FCM) algorithm and the validity measure of Xie and Beni [14] defined by the following expression (6):

$$XB_{(c)} = \frac{J_m(U,V)}{n} \frac{1}{\min_{i,j=1,c;j \neq i} \|v_i - v_j\|^2} \quad (6)$$

$$\text{where } J_m(U,V) = \sum_{k=1}^n \sum_{i=1}^c \pi_i(x_k)^m \|x_k - v_i\|^2$$

In this expression, J_m represents the objective function to be minimized, the Euclidean norm, and $\pi_i(x_k)$, the membership value of the k th point of the i th class that has center v_i . All membership values are contained in U and the centers of the classes are contained in V . The coefficient m is generally set at 2. It allows obtaining a more or less fuzzy partitioning. The number of classes found by FCM must be greater than 1 to apply this validity measure. The lower the measure, the more optimal the number of classes. Once the optimal number of classes is estimated by FCM, and if each of these classes contains at least k objects, the split is performed. If the number of shapes is not sufficient when the number of classes is estimated by FCM, then it is necessary to wait until this number of objects is classified to check again the splitting of the class.

Therefore, before applying this method to the estimation of tomato maturity, we first re-evaluated it by applying it on artificially generated data to study the three cases of evolution namely class displacement, merging and splitting because it is difficult to obtain a real database containing classes with features evolving with time. These artificially generated classes are Gaussian and perfectly reflect the different aspects of evolution.

A. Moving classes

The DFKNN is evaluated on displacement of a class. This displacement was generated in the following way. At $t=0$, 150 observations forming the initial class are used as the training set. The mean and standard deviation values of the class are for attribute 1, $\mu^1=3$ and $\sigma^1=1$, and for attribute 2, $\mu^2=3$ and $\sigma^2=1$. At $t=1$ through 50, 50 new shapes appear with the same standard deviations and means as those in the initial class. In this case, there is no evolution or shift. At $t=51$ until 200, a change appears in the values of the class mean with respect to each attribute j , $j=1,2$. This change is a gradual shift in the class mean with respect to each attribute according to the equations below.

$$\left. \begin{aligned} \mu^{1'}(t) &= \mu^1 + 2 + \frac{4 \times (t-50)}{150} \\ \mu^{2'}(t) &= \mu^2 + 2 + \frac{2 \times (t-50)}{150} \end{aligned} \right\} 51 \leq t \leq 200 \quad (7)$$

After this move, we obtain a new or final class.

- At $t=201$ to 300, 100 new observations appear. They have the same standard deviations and means as those of the final class. These observations are classified in the final class without a sufficiently important new evolution being detected. This is done with the fixed parameter values of $k=5$; $th_1=5$; $NbMin=10$; $th_{Fusion}=0,2$; $n1=10$ and $n2=20$. Figure 1 shows an example of this shift.

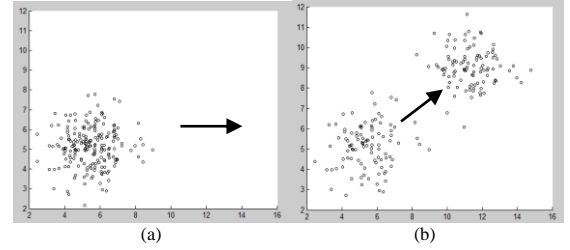


Fig.1 Moving a class. (a), initial class. (b), moving some objects of the class.

B. Merging classes

In this example, we take two classes moving to a common area to merge. The merging of two classes can represent the case of a temporal evolution of a system towards a new system. The database of the two merging classes was created as follows. At $t=0$, 125 observations are used to learn the membership functions of class C1 ($\mu^1=5$, $\mu^2=7$) and 125

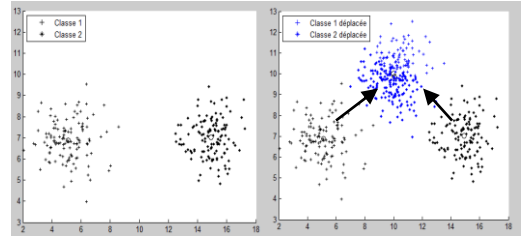


Fig.2 Displacement of two classes to the same area.

others are used for class C2 ($\mu^1=15$, $\mu^2=7$). The standard deviation of these classes is equal to 1 for each attribute. At $t=1$ until 150, the new shapes belonging to each class go to a common area for both classes (Fig.2).

These moves are generated for both classes based on each attribute j , $j=1,2$, using the following expressions:

$$\left. \begin{aligned} \text{C1: } \mu_1^{1'}(t) &= \mu_1^1 + 2 + \frac{3 \times t}{150} \\ \mu_1^{2'}(t) &= \mu_1^2 + 2 + \frac{t}{150} \\ \text{C2: } \mu_2^{1'}(t) &= \mu_2^1 - 2 - \frac{3 \times t}{150} \\ \mu_2^{2'}(t) &= \mu_2^2 + 2 + \frac{t}{150} \end{aligned} \right\} 1 \leq t \leq 150 \quad (8)$$

the computation of the similarity value by expression (5) gives a result higher than the set threshold, $th_{Fusion}=0,2$. This means that the two classes overlap enough to merge.

Figure 3 shows the result of merging the classes and classifying the set of observations.

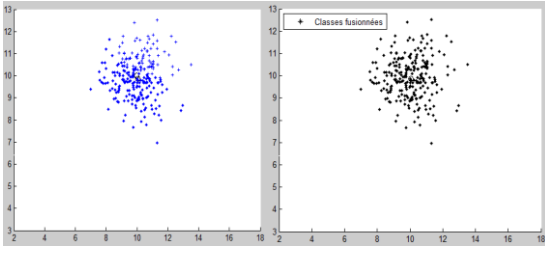


Fig.3 Classification result obtained by DFKNN after fusion and classification of the set of shapes . Here $k=5$; $thl=5$; $NbMin=10$; $th_{Fusion}=0,2$; $n1=10$ and $n2=20$.

C. Sorting Classes

The DFKNN method has also been tested in the case of the splitting of a dynamic class on a database is created evolving with time as. Thus, at $t=0$, 100 observations are created as initial class and used for training. At $t=1$ to 150, the new points evolve in two different directions as shown in Figure 4.

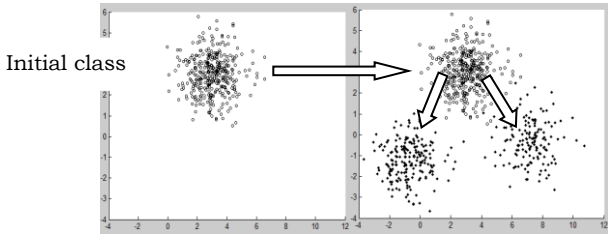


Fig.4 Evolution of the points in two different directions.

This evolution is generated for $C1$ according to each attribute j , $j=1,2$, by the expressions (9).

At $t=151$ until 250, new observations are classified with the same standard deviation and mean as the last observations of the evolution.

$$\begin{aligned} \mu_1^1 &= \mu_1^1 - 2 - \frac{4 \times t}{150} \\ \mu_1^2 &= \mu_1^2 - 2 - \frac{5 \times t}{150} \end{aligned} \quad 1 \leq t \leq 150, \quad (9)$$

$$\begin{aligned} \mu_1^1 &= \mu_1^1 + 2 + \frac{5 \times t}{150} \\ \mu_1^2 &= \mu_1^2 - 2 - \frac{3 \times t}{150} \end{aligned} \quad 1 \leq t \leq 150,$$

Several adaptations of the class took place for this case. Following the previous adaptation of the class at $t=8$, a split into two classes was revealed by the application of the FCM method which detected the 2 classes. However, with this FCM method, too few observations whose number is lower than k are present in the 2 classes obtained at this time. It is therefore necessary to wait for k new points to appear in order to recheck if the split is still necessary.

At $t=18$, the class is again studied by FCM. The validity measures corresponding to several numbers of classes are

given in Table 1. The minimum value obtained in this table for $c=2$ indicates that the best number of classes found is two. Figure 5 shows.

TABLE I
Validity measures for the class obtained at $t=18$.

Number of classes	Validity measure
2	0.03
3	0.18
4	0.20
5	0.40

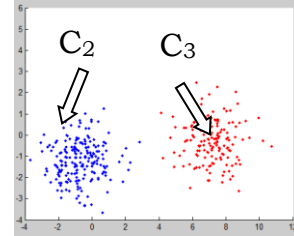


Fig.5 Classification result obtained by FCM for the split case with $k=9$, $thl=5$, $NbMin=15$, $th_{Fusion}=0,2$, $n1=10$ and $n2=20$.

The split of one class into two classes was detected. After this, the rest of the classification can take place.

The classification results obtained by the DFKNN method, in particular in the different cases of evolution of Gaussian classes generated, show that this method is effective for the detection of the displacement of classes as well as their adaptations and validations. Nevertheless, in the case of class splitting and on several tests performed, we find that the results obtained are less good, but acceptable. Although minimal, the estimation errors of the number of classes obtained after splitting using the FCM algorithm show the difficulty of detecting a split and its validation. However, in general, for the three cases of evolution, i.e., displacement, merging and splitting, we have achieved a classification rate of 97%, which can be considered satisfactory.

III. APPLICATION TO THE MATURITY LEVEL ESTIMATION OF TOMATOES

In this work, we are interested in the maturity of the tomato because this fruit is of considerable importance as it is widely consumed worldwide and in our country and continues to be the subject of much scientific research. It is in fact a question of estimating the times of its conservation to ensure its distribution under the best conditions and, especially, to decide its destination according to its maturity. The objective is therefore to automate the procedure of its sorting which constitutes a wearing operation physiologically and mentally for the human being when it is himself who carries out this work and which, moreover, even experienced, can be prone to errors related to disorders of the sight and the tiredness. Therefore, the automation of this process is of great interest for agriculture and the tomato canning industry.






The estimation of the maturity tomato is based essentially on the evolution of its coloration which takes five colors going

gradually from the whitish green to the dark red while passing by the yellow, the pink, the turning, the orange and the clear red. This evolution has been codified by the USDA (United States Department of Agriculture) according to Table 2 [15].

A. The used method

To characterize the color of tomatoes, we can use all the systems of colorimetric representation but the color space $L^*a^*b^*$ CIE 1976 still named CIELAB, is the color space that best suits our application. Indeed, this space is often used for

TABLE II
Maturity levels of tomatoes.

Image	Maturity stage	Description
	Green	The surface of the tomato is completely green.
	Tournant	10% to 30% of the tomato surface is not green.
	Orange	30% to 60% of the tomato surface is not green.
	Light red	60% to 90% of the tomato surface is not green.
	Dark red	More than 90% of the tomato surface is not green.

the characterization of surface colors. Recall that in this system the parameter L^* derives from the luminance and values 0 and 100 corresponding respectively to the absolute black and perfect white while b^* represents the value on the blue-yellow axis and a^* , the value on the green-red axis. After verification, we have indeed found that the use of the only parameter a^* is largely sufficient, which is reasonable given that the color of the tomato evolves gradually from green to red. This verification was done by plotting the histograms of both a^* and b^* values of tomatoes throughout the ripening period. The results show well that the value of b^* does not show great changes while the value of a^* increases with the rate of ripening, which allows us to conclude that the value of a^* is highly correlated with the level of ripeness of the tomato and can be used for the estimation of the latter. This result was first proven in [16] as shown in Table 3 which maps the value of a^* to the maturity rate. This table serves as a reference in our application.

B. Application

For this application, we obtained tomatoes at different color levels from the region of Tizi-Ouzou to establish a database containing 500 images that we gathered into 5 classes containing 100 images each at different maturity levels. These images were taken 3 to 5 days apart using a CCD camera with a resolution of 10 mega pixels and placed

approximately 100 mm above the tomato with similar lighting conditions.

Note that the tomatoes are placed so that the image captured is of the base facing up and placed on a black background to mitigate the black shadow. To these images, we have made some transformations to obtain images devoid of any unnecessary information and have just the object of interest. Indeed, we resized the images in 250×250 pixels, converted the RGB system in $L^*a^*b^*$ system (Fig. 6 b), filtered the images using a median filter and eliminated the background of the images using morphological operations. We then extracted

TABLE III
Values of a^* Maturity levels of the tomato [13].

Rate of maturity	Value of a^*	Color of tomato	Shelf life
10% to 20% of maturity	$a^* < -5.8$	Green to yellow	21 to 28 days
30% to 40% of maturity	$-5.8 \leq a^* < 2.1$	Tournant	15 to 20 days s
50% to 60% of maturity	$2.1 \leq a^* < 9.2$	Orange	7 to 14 days
70% to 80% of maturity	$2.1 \leq a^* < 21.5$	Light red	5 to 6 days
Total maturity	$21.5 \leq a^*$	Dark red	2 to 4 days

the contours of the images by contour tracking by connexity on the binarized images obtained by applying the method of Otsu [17] [18] after a conversion of these images into gray levels (Fig. 6 e).

Once all these operations are completed, we proceed to the calculation of the value of a^* of all the pixels belonging only to the object and which are inside the contour of the tomato, traced in the previous step. We calculate, then, the average of all the values of a^* of all the pixels of the image. This average is the color characteristic of the whole corresponding tomato.

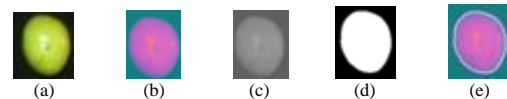


Fig.6 Pre-processing operations
(a) original image (b) $L^*a^*b^*$ image
(c) grayscale image (d) binary image, (e) contour plot.

C. Results and discussion

Once calculated, the value of a^* of each image is compared to the values determined in Table 4 to assign it to the corresponding class among the five classes already predefined. The results obtained for ten samples are presented in the table below. From the results obtained on the previous table, except for the few classification errors, which are indicated in the shaded cells, we can say that the results obtained are good. Indeed, for 100 test images we have detected six misinterpretations, which corresponds to a classification rate of 88%. However, we have listed some advantages and disadvantages of the method.

The advantage of this procedure is that it is a very good alternative compared to the work done by human vision for the judgment of the maturity level of the tomato, but also for

the estimation of the shelf life of the latter, especially when it is necessary to transport them to different distances, that is to say, to distinguish the tomatoes that will be transported further from those that will be destined to small distances. However, in addition to the constraints related to the lighting conditions and the resolution of the camera, the method has two disadvantages, namely that it can only process one tomato at a time and it cannot differentiate the tomato from another fruit. The advantage of this procedure is that it is a very good alternative compared to the work done by human vision for the judgment of the maturity level of the tomato, but also for the estimation of the shelf life of the latter, especially when it is necessary to transport them to different distances, that is to say, to distinguish the tomatoes that will be transported further from those that will be destined to small distances. However, in addition to the constraints related to the lighting conditions and the resolution of the camera, the method has two disadvantages, namely that it can only process one tomato at a time and it cannot differentiate the tomato from another fruit.

TABLE IV
Values of a^* maturity levels of the tomato [13].

Maturity level seen by man	Example	Value of a^*	Results
10% to 20%	1	-13.99	10% to 20%
	2	-19.21	10% to 20%
	3	-13.63	10% to 20%
30% to 40%	1	-2.55	30% to 40%
	2	-4.65	30% to 40%
	3	-5.25	30% to 40%
50% to 60%	1	3.13	50% to 60%
	2	2.89	50% to 60%
	3	3.43	50% to 60%
	10	5.89	50% to 60%
70% to 80%	1	10.87	70% to 80%
	2	9.95	70% to 80%
	3	10.56	70% to 80%
Total maturity	1	32.98	Total maturity
	2	28.48	Total maturity
	3	30.11	Total maturity
	4	42.36	Total maturity
	5	49.55	Total maturity
	6	52.04	Total maturity
	7	48.17	Total maturity
	8	44.00	Total maturity
	9	50.69	Total maturity
	10	54.53	Total maturity

IV. CONCLUSION

The application of the dynamic fuzzy k-nearest neighbor classification method or DFKNN, which takes into account the evolutionary aspect of the classes and which makes it possible to treat the various evolutions that dynamic classes can meet such as displacement, splitting and merging, as well as the results obtained show the interest and the contribution of a dynamic classification method, in terms of results and classification time. Nevertheless, the choice of parameter values directly influences these results. Whenever the evolution of the classes is periodic, the classification becomes

static by dividing the period of the evolution into well-defined intervals. This approach was approached and applied for the estimation of the level of maturity of the tomato carried out on RGB images and converted in the system of representation $L^*a^*b^*$, the results thus obtained were satisfactory, with an optimal time of calculations. Nevertheless, the method has some limitations mainly related to the quality of the tomato images and to the different pre-processing operations which must be efficient in order to eliminate the maximum of noise to guarantee a good classification. Although the DFKNN method has been applied on artificially generated classes, not having been able to obtain real classes with evolving characteristics, this method can be easily applied on real cases such as industrial diagnostic systems in order to take advantage of its benefits.

REFERENCES

- [1] L. Hartert, "Reconnaissance des formes dans un environnement dynamique: appliquée au diagnostic et au suivi des systèmes évolutifs", Doctoral thesis, university of Reims Champagne -Ardenne 2010.
- [2] V. Gunes, "Reconnaissance de formes évolutives par combinaison, coopération et sélection de classifieurs". Doctoral thesis, university of laRochelle 2001.
- [3] C. Lurette, "Développement d'une technique neuronale auto-adaptative pour la classification dynamique de données évolutives: application à la supervision d'une presse hydraulique". Doctoral thesis, university of Lille 2003.
- [4] H. A. Boubacar, "Classification Dynamique de Données non-stationnaires : Apprentissage séquentiel des Classes évolutives", Doctoral thesis, university of Lille, France, 2006.
- [5] V. Vapnik, "The Nature of Statistical Learning Theory". Springer-Verlag New York, 1995.
- [6] K. Boukharouba, and S. Lecoeuche, "Online Clustering of Non-stationary Data Using Incremental and Decremental SVM", International Conference on Artificial Neural Networks. Springer, Berlin, Heidelberg, 2008.
- [7] S. Marsili-Libelli, "Adaptive Fuzzy Monitoring and Fault Detection", International Journal of COMADEM, 1998.
- [8] J.C. Bezdek, "Pattern Recognition with fuzzy objective function algorithm". Plenum Press, 1981.
- [9] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy K-nearest neighbor algorithm," in IEEE Transactions on Systems, Man, and Cybernetics, 1985.
- [10] L. Hartert, and M. Sayed-Mouchaweh, "Semisupervised Dynamic Fuzzy K-Nearest Neighbors." Learning in Non-Stationary Environments. Springer, New York, 2012.
- [11] X. F. Zhong, S. Z. Guo, L. Gao, H. Shan, and J. H. Zheng, "An improved k-NN classification with dynamic k." Proceedings of the 9th International Conference on Machine Learning and Computing, 2017.
- [12] S. Taneja, C. Gupta, S. Aggarwal, and V. Jindal, "MFZ-KNN—A modified fuzzy based K nearest neighbor algorithm." International Conference on Cognitive Computing and Information Processing (CCIP). IEEE, 2015.
- [13] H. Frigui, and R. Krishnapuram, "Clustering by competitive agglomeration." Pattern recognition, 1997.
- [14] X.L. Xie, and G. Beni, "A validity measure for fuzzy clustering". IEEE transactions on Pattern Analysis and Machine Intelligence, 1991.
- [15] USDA/AMS, "United states standards for grades of fresh tomatoes", Washington, 1991.

- [16] Y. Gejima, H. Zhang, and M. Nagata, " *Judgment on level of maturity for tomato quality using L*a*b* color image processing*", Int Conf on Advanced Intelligence Mechatronics. IEEE, 2003.
- [17] N. Otsu, " *A threshold selection method from gray-level histograms*." IEEE transactions on systems, man, and cybernetics 9.1, 1979.
- [18] P.K. Sahoo, S.A.K.C Soltani, and A.K. Wong." *A Survey of thresholding techniques*", Computer vision graphics and Image processing, 1988.

Feature Extraction and Clustering of Abnormal Blood Cells

^{1*}Abdellatif BOUZID-DAHO, ²Patrick SIARRY

¹ *Laboratoire Vision Artificielle et Automatique de Systèmes (LVAAS)
Faculty of Electrical and Computer Engineering
Mouloud Mammeri University, Tizi-Ouzou, Algeria*

Paris-Est Creteil University, Fr

^{1,2} *Laboratoire Images, Signaux et Systèmes Intelligents (LISSI),
Paris-Est Creteil University, France*

abdellatif.bouزيد-daho@ummto.dz

Abstract—the progressive development of microscopic medical image segmentation is an important step for modern medicine. However, the existing methods are tricky to be applied on real problems. The objective of this paper is to create an algorithmic system, study on image clustering algorithm based on k-means method, the application context aimed at the blood cells, the experimental results obtained showed the effectiveness of this approach to classify, extract the morphological characteristics (shape, size, color, texture or movement) for the set of medical images tested, from its information one can make a classification of the objects by cluster.

Keywords—segmentation; algorithmic; clustering; k-means; blood cell.

Anisotropic Diffusion Filtering Optimized by Simplex Method Applied to MRI images with a Priori on the Rayleigh Probability Density Function of the Noise

Samir YESLI, Moussa DIAF

*Laboratoire Vision Artificielle et Automatique des Systèmes (LVAAS)
Mouloud Mammeri University (UMMTO), Tizi-Ouzou – Algeria.*

samir.yesli@ummto.dz
moussa.diaf@ummto.dz

Abstract—Magnetic resonance images (MRI) are tainted by Gaussian noise in the complex plane (K space). However, after the inverse Fourier transform, this noise follows a Rician probability distribution function (PDF). Estimating the variance of the noise is an important step in the denoising process. In order to determine the variance of the noise, we can exploit one of the following two possibilities: the first concerns the regions where the MR signal exists; this one follows Rice PDF. The second, concerns the regions where the MR signal is absent, in this case, the noise follows a Rayleigh PDF. Denoising methods based on partial differential equations (PDE), in particular diffusion equations have been widely used for their ability to remove noise while preserving essential structures such as edges. The latter are usually deduced from variational methods which consist in minimizing a functional and its partial derivatives. The combination of noise variance estimation and diffusion PDEs induces additional parameters that can play on the efficiency of the denoising process. In this perspective, we have chosen to introduce an iterative optimization method called simplex method in order to estimate the right parameters applied to the diffusion PDEs. The proposed method is a strategy based on a scheme that combines the simplex algorithm and the estimation of the noise variance according to the Rayleigh PDF, thanks to the maximum likelihood function, in order to fit the parameters as well as possible. The application of the proposed method on real MRI images of multiple sclerosis and tumors, for qualitative evaluation, shows the interest of the proposed approach to improve the quality of MRI images, especially for post-processing methods.

Keywords—Image enhancement, MRI, anisotropic diffusion PDE, simplex algorithm, Maximum likelihood, Rayleigh distribution.

Intelligent system for detecting faults in the industrial area

¹Abderrahim BENMOHAMED, ²Adil BOUGUERRA

¹*MCM university of Souk Ahras,*

²*Abbes Laghrour university of Khenchela*

benmohamed.abderrahim@gmail.com

adil.bouguerra@smartest.dz

Abstract— In this work, we propose a system capable of performing two tasks. recognition and prediction of human action using the surveillance cameras in the industrial environment to detecting employees who violate the safety regulations by not wearing the safety cloths, this system is based on deep learning methods (convolutional neural networks, long-short term memory).

For the human action recognition and prediction we proposed a new action representation, based on the human skeleton (body parts key-points), we used these points to create a vector containing the most important features of a descriptor (invariant to rotation, scale and position in the frame), and storing the spatial information of the video, we also used a position map to reduce its size and get very simple representation. Finally, we used LSTM network to preserve the temporal information, by training on these features using the dataset we created. The other part of this system is responsible of detecting employees who violate the safety regulation. So, in order to achieve this, we used the state of the art algorithm (You Only Look Once) for object detection and we adapted it to our dataset that contains four classes (Wearing-helmet, Without-helmet, Wearing-boot and Without-boot).

Index Terms— Action recognition, Action prediction, Object detection, Deep learning, Long-short term memory,, human Skeleton.

The safety of people in a intelligent building by classifying speech with neuron networks and hidden markov chains

Sidali BOURENANE, Sid Ahmed HENNI, Benabdellah YAGOUBI

Dept. Genie Electrique, Laboratory LSS

Faculty science and technology

University Abdelhamid Ibn Badis

Mostaganemm, Algeria

sidouzema@gmail.com

Abstract—people want to live in the intelligent building because it offers them comfort. But the latter is unable to ensure the safety of these people, for example a person who lives alone, and who has discomfort, or a heart attack, or who is seriously ill, is not taken care of by this intelligent building. The aim of this article is the realization of a system which ensures the safety of people in an intelligent building by implementing an automatic speech recognition system. This system will recognize these sentences that indicate that a person needs help, or who is ill, or who is in danger, which allows him to ensure the safety of that person, by informing another who takes charge of these dangerous situations. For the realization of this automatic speech recognition system, we used the MFCC method (Mel frequency ceptral coefficient) for the representation of the speech signal, and a hybrid system NN-HMM (Neural Networks - Hidden Markov's models), for voice recognition. An experiment will be implemented from an input database; the vocabulary used in this corpus is composed of 10 sentences in the English language. The recognition will be of global type, the speech signal will be parameterized using the MFCC coefficients to form the input vector of the recognition system. The NN-HMM hybrid system will be used for the learning phase and the recognition phase.

Keywords—The safety; MFCC; Intelligent Building; Speech Processing; Neural Networks; Hidden Markov's models; Classification

Stochastic and discrete-event model for simulation and performance evaluation of car-sharing networks

Ali HAMROUN, Karim LABADI, Fethi OUALOUCHE, Mourad LAZRI, Soltane AMEUR

Laboratoire LAMPA

University of Tizi Ouzou, Tizi Ouzou, Algeria

ECAM-EPMI / Quartz-Lab / 13 Boulevard de l'Hautil, 95092 Cergy-Pontoise (France)

`ali_ham82@hotmail.com`

`m_lazri@yahoo.fr`

`k.labadi@ecam-epmi.fr`

Abstract- In recent years, car-sharing systems have gained worldwide attention as a new solution to mobility problems such as pollution reduction and congestion. The operation of these transport systems needs a rigorous control and a suitable management strategy to have a high quality of service. In this paper, we develop an original discrete-event approach for modeling and analyzing the performance of car-sharing systems using stochastic Petri nets, taking into account their dynamic behavior under constraints such as battery charging and car maintenance.

Keywords- *Petri nets, Car-sharing, stochastic.*

Based robust Sliding Mode Control of two wheels EV

Fatiha ZAGHRAT, Brahim GASBAOUI, Boussema BOUCHIBA, Ismail Khalil BOUSSERHANE,
Imane CHAOUFI

*Dept. of electrical engineering
SRGE Laboratory
Bechar, Algeria*

zaghratfatiha@gmail.com

Abstract—This article shows the design modeling and simulation of an electric vehicle speed control with two rear wheels each of these wheels contains a DC motor (motor-wheel) which bases on a general sliding mode control speed scheme. The objective is to achieve the efficiency of the suggested SMC on the traction system. First, a PI controller is developed for EV. To evaluate the SMC robustness and the dynamic performance of the suggested 2WD for an EV, two EV speed control methods are presented and compared: PI controller and Sliding mode controller (SMC). The EV has been evaluated using the MATLAB / Simulink using three road constraints: the straight road, the sloping road, and the curving road to the right and left. The results of the simulation analysis and comparison of the two regulators, obviously demonstrate that SMC guarantees good efficiency and better response with zero overshoot, no steadystate error with rising time is 0.13sec compared to PI control (4.26%, 0.21sec). Moreover, the SMC guarantees the stability to drive during the curve compared to the classical controller. Instead of employing PI control, we propose sliding mode control to ensure stability and increase the control loop's strangeness. The results demonstrate the dynamic response to compare the influence of load torque perturbations in the instances of two kinds of control. In the instance of sliding mode control, results show high load rejection, excellent response, and better tracking results than the classic controller.

Keywords—electric vehicle (EV), DC motor, electronic differential (ED), sliding mode control (SMC), PI controller

Enhancing Firefly Algorithm by Modifying Evolution Rules and Introducing Position based Crossover

Mourad KHETATBA, Rachid BOUDOUR

LASE Laboratory, Badji Mokhtar University, Annaba- Algeria

khetatbam@yahoo.fr
racboudour@yahoo.fr

Abstract— Nature-inspired meta-heuristic algorithms, especially those based on swarm intelligence, have already attracted more attention. We focus on the Firefly algorithm which is inspired by fireflies behavior and more used in global optimization problems. Its main features are a strong exploration, a few parameters settings and effortless operation. However, the firefly algorithm presents some drawbacks. This algorithm might be stuck in stagnation (the solutions do not enhance anymore), providing low accuracy and easily falling into the local optima in the global optimization search. The balance between exploration and exploitation cannot be well maintained. In this paper, in order to overcome these weaknesses, we will propose a new version of firefly algorithm called EFAPOX. It has been demonstrated that the FA performance depends strongly on the control parameters. Therefore, chaotic maps are used to generate fireflies and to tune attraction coefficient while a new strategy is applied for getting the random movement factor, α . For more diversity and avoiding local optima, a crossover operator is used and, after iterations' threshold, a regeneration of some new solutions is done for replacing the worst in the population of fireflies. We based experiments on some standard testing benchmark functions. Compared to FA state-of-the art, the results of EFAPOX show the improvements are relevant in terms of high optimal solution quality and fast convergence.

Keywords— *Firefly algorithm, Meta-heuristic, Swarm Intelligence, Optimization.*

A		H		S	
Ammar M	32	Hariri W	1,13	Seridi-Bouchelaghem H	13
Attaf Y	57	Hachouf F	38	Siarry P	119
Akrour L	65,71	Harrar K	43		
Ayouni B	65	Haddad M	50	T	
Ameur S	65, 71, 123	Haouli I.E	13	Tangara B	82
		Hannoun K	86	Touba M .M	43
		Hamiche H	86		
		Hammou F	93	Y	
		Hammouche K	93	Younsi M	25
		Hamoudi A	99	Yesli S	120
		Henni S.A	122	Yagoubi B	122
		Hamroun A	123		
B		I		Z	
Bourefis L	1	Iratni S	25,112	Zaghoudi R	43
Benmohamed A	121	Idjeri B	77	Ziani S	43
Bouguerra A	121			Zaghrat F	124
Boudiaf A	43	K			
Bensafi N	57	Kharfouchi S	38		
Bougherira H	82	Karaoui F	82		
Benblidia N	82	Khettib N	82		
Benselama Z	82	Khetatba M	125,		
Bettayeb M	99	Khelifa M.A	106		
Boudour R	125				
Boukabou A	106	L			
Bensidhoum M.T	77	Laidi A	32		
Boukendour O	112	Lahdir M	57,65,71, 86		
Boumati S	112	Laghrouche M	77,86		
Bouزيد-Daho A	119	Lekouaghet B	106		
Bourenane S	122	Labadi K	123		
Bouchiba B	124	Lazri M	123		
Bousserhane I.K	124				
C		M			
Chekir A	8,20	Maarouf A .A	38		
Chaoufi I	124	Mamoudou Moumouni A	82		
		Meliani Z	82		
		Megherbi O	86		
D		N			
Diaf M	25,112 120	Nait Ali A	71		
Djehali N	99	Naceur D	82		
F		O			
Faouci S	50	Ouahioune M	65,71		
Fournier R	71	Oualouche F	123		
G					
Guessoum A	82				
Gaseb D	50				
Gasbaoui B	124				