Harris Hawks Optimization Method based on Convolutional Neural Network for Face Recognition Systems

Sulayman Ahmed
ENET'COM
Universite de Sfax
Sfax, Tunisia
sulaimanah2000@uokirkuk.edu.iq

Mondher Frikha
ENET'COM
Universite de Sfax
Sfax, Tunisia
mondher.frikha@enetcom.usf.tn

Taha Darwassh Hanawy Hussein
ENET'COM
Universite de Sfax
Sfax, Tunisia
taha_hussein@uokirkuk.edu.iq

Javad Rahebi
Department of Software Engineering
Istanbul Ayvansaray University
Istanbul, Turkey
cevatrahebi@ayvansaray.edu.tr

***Abstract*— This paper discusses the momentum gradient dependent on the convolutional neural organization's strong point. It is a new methodology introduced to detect evenness in the data set of faces. The proposed face recognition framework was created for various purposes. Through Gabor wavelet change, facial evenness was extracted from the face-preparing information. After that, we applied a profound learning process to carry out verification. After applying the proposed method to YALE and ORL data sets, we simulated them using MATLAB 2021a. Before this, similar trials were directly applied through Harris Hawks Optimization (HHO) for including the determination approach. The extraction process was conducted with many picture tests to execute the Gabor wavelet method, which proved more viable than other strategies applied in our examination. When we applied the HHO on the ORL dataset, the acknowledgment rate was 93.63%. It was 94.26% when the three techniques were applied to the YALE dataset. It shows that the HHO calculation improved the exactness rate to 96.44% in the case of the YALE dataset and 95.88% in the ORL dataset.**

Keywords—Face recognition, Convolutional neural network, Momentum gradient

I. Introduction

Face processing is a main problem in machine vision algorithms, which uses intelligent methods to identify human beings based on their physiological characteristics. This technology has advantages such as high accuracy and low level of individual intervention. It has benefits in many applications regarding information security, law enforcement, monitoring, traffic control, and registration in attendance systems. In recent years, in-depth learning has achieved tremendous performance in computer vision, especially in terms of face recognition. Deep learning uses a combination of abstract features and deep neural networks to generate results that are generally unachievable and indistinguishable. The structure of the deep network-based face recognition model depends on the training data quality and number. In general, face recognition systems have challenges such as changes in brightness, different facial expressions, face coverage, low image resolution, and limited examples [1][2].

For a human brain, face recognition is a high-level visual work. It is easy for a person to detect/identify faces but developing a computer system that can perform the same task is a significant challenge because it involves several complex procedures. It is a considerable challenge because some factors hinder face recognition, specifically when there are uncontrollable and inappropriate conditions for getting a facial image. Two types of facial differences exist in this context. The first exists because of physical similarity/similarities between individuals, which are generally limited. The second difference exists when the changes in a person's face are significant, and that is because of factors, like facial expressions and lighting conditions, and such factors may fail a face recognition system. Several researchers have attempted to resolve the mentioned problems; however, there is no perfect solution [2][3].

Indeed, a face is a biological characteristic to identify humans. The human face recognition system lies at the core of a face recognition system, so this process has been repeatedly studied [4][5][1]. But, the mentioned studies are ineffective when there is a small sample size, which gives modest results. This study, however, presents an approach to solving this issue.

II. Related Works

In the available literature, researchers have presented different methods to recognize faces. They include hybrid, local, and holistic processes [6][7]. Some recent examinations have uncovered that the face recognition data should be balance-based for a more accurate face recognition because it uses facial evenness for recognition [8]. In this context, facial balance is valuable for two principal face-recognition approaches. The first includes the specific numbers of face-preparation tests while the second focuses on the varieties in lighting conditions, outward appearances, and postures. This strategy utilizes facial balance to reduce the mentioned issues.

Generally, an FR system has three main steps: preprocessing, feature extraction, and classification [9][10]. Preprocessing stabilizes and makes the input ready to extract features. It mainly extracts features from an image [11]. Thus, an ideal preprocessing procedure eliminates irrelevant information (such as background, rotation, or illumination) [10]. For this reason, many authors have included the preprocessing step in their studies [10][9][12][13], even though they used different techniques, including Difference of Gaussian (DOG) [12], Gaussian filter [14], Gabor filter [9], Histogram Equalization (HE) [17], 2-D Wavelet Transform (2D WT) [15][16], and gamma correction [12]. Here, the objective is to extract features from the image of a face but, it must be remembered that extracting features from an image is easy for a human but very challenging for a computer [18]. For feature extraction, many algorithms are used, including Local Binary Pattern (LBP) [18], Gray Level Co-occurrence Matrix GLCM [19], Gabor [10], Linear Discriminant Analysis (LDA) [22], Speeded Up Robust Features (SURF) [20], and Scale Invariant Feature Transform (SIFT) [21]. Classification is the last but the actual recognition step for matching the feature vector of an image, which is obtained through feature extraction, with the corresponding feature vectors of trained images in a database. Many techniques are used for classification [23], which vary from K-Nearest Neighbor [18], Support Vector Machine (SVM) [25], and Euclidean Distance [24] to more advanced algorithms such as Neural Networks [26][27]. Some proposed methods exist, which overcome the mentioned problems using facial symmetry. Such problems are still to be solved. In a recent work [8], the author proposed to improve accuracy through training the FR system and applying the facial data taken from symmetrical and original facial images.

I. PROPOSED METHOD

A. Histogram Oriented Gradient

It is a very popular feature extraction method [28]. HOG is a commonly-used descriptor for human detection [17]. HOG involves computing the gradient orientation and magnitude. For obtaining the HOG of a photo, the variations in both X and Y are computed before obtaining the direction and the magnitude.

1) Computing Gradients

A HOG operation is either the derivative or a center differential because it involves both x and y derivatives. After obtaining the derivatives, computing the gradient orientation and magnitude becomes possible.

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x-h)}{2h} \tag{1}$$

The following formula is used to calculate the magnitude:

$$s = \sqrt{s_x^2 + s_y^2} \tag{2}$$

And, the orientation is computed by:

$$\theta = \arctan\left(\frac{s_y}{s_x}\right) \tag{3}$$

2) Blocks and Cells

"Fig. 1" shows the image of a face. It is a 64x128 image, and when it is divided into 128 cells, some blocks will be selected, such as the first block (Block 1 with 2x2 cells). In this case, the second block (Block 2) is 50% overlapped, so each block has 2x2 cells of 8x8 size that means 16x16 with 7x15 = total 105 blocks.

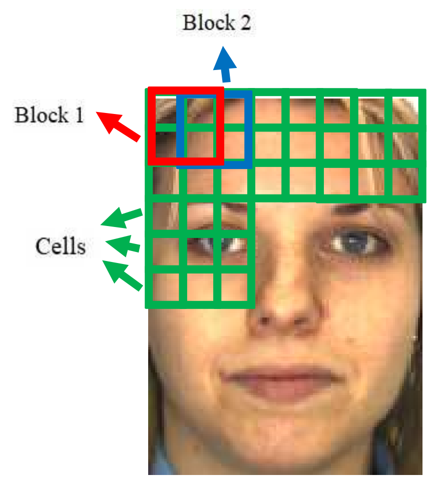


Fig. 1. Cells and blocks

3) HOG Feature Extraction Steps

For calculating the HOG for a 64x128 image, the area of the image is divided into 50% overlapping 16x16 blocks; so there will be 7x15 = total 105 blocks with 2x2 cells with 8x8 size. HOG is then quantized in nine directions/bins. In case, the HOG direction is not towards any bin, interpolation is carried out to apply Gaussian and smooth out the histogram. It is possible to concatenate all the descriptors since there are 105 blocks with nine dimensions each, which means that there are 3780 dimensions for the whole block image.

B. Feature selection with Harris Hawks Optimization

Group intelligence is an optimization algorithm that makes population members use both situational and group information. It attempts to solve an optimization issue. For most meta-heuristic algorithms, which use the group intelligence approach, researchers specifically model the group hunting behavior. In these algorithms, members of the population circle around the current prey or optimal point and try to search for it. And attack at the right time. Group hunting behavior is found in nature among various organisms, some of which are modeled on the behavior of arthropods, mammals, and insects. Group intelligence systems include the behaviors of several living things that work together, but their number may not be significant. An example of such a system is the falcon optimization algorithm [29], presented and modeled in

2019. In these species, groups of up to 6 birds typically participate in hunting and fly around the prey and hunt it, as shown in “Fig. 1”:

Fig. 2. Group intelligence hunting mechanism in Harris optimization algorithm

It is believed that this adaptation is for group hunting with no prey in the desert. The behavior of these creatures shows that first, a small group steps forward for hunting. Other group members also go ahead to participate in the hunt and cooperate in group hunting. All the hawks (problem-solving solutions) scatter around the prey (optimal answer) in this hunting method. Then a bird hunts it. This algorithm assumes that first the prey is identified then surrounded, and then attacked. In this algorithm, each falcon is a solution to the problem. The current optimal solution lies in the rabbit's position, and the falcons fly towards that position. These algorithms first require falcons to search the problem space before finding the prey and attacking it. Equation 1 mathematically models the falcons’ random and initial search behaviors [29]:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 . X(t)| & rand \geq 0.5 \\ (X_{rabbit}(t) - X_M(t)) - r_3(LB + r_4(UB - LB)) & rand < 0.5 \end{cases} \quad (4)$$

Here, $X(t)$ represents a hawk’s (solution’s) current position within the current iteration (t). In the new iteration, $X(t+1)$ is the position of a hawk and $X_{rabbit}(t)$ is the position that shows the optimal answer. In the problem space, $X_{rand}(t)$ has a random position, $X_M(t)$ is the center of gravity of the falcon population, and r_1, r_2, r_3 and r_4 show random numbers in the range 0-1, LB, and UB, respectively. To calculate $X_M(t)$ in the problem space, we can use Equation 2, where the number of solutions is equal to N Harris Hawks [29]:

For computational processes, HHO is applied for an irregular advancement calculation because it includes characterization and choice. It finishes when repeated efforts are made to choose the most family members and also through the highlights’ helpful arrangement. It improves or maintains the grouping execution of a face recognition system[30-36].

Behind this calculation, the essential thought is the co-evolvement of various birds in place of zeroing in on a specific type of bird, which contributes to powerful hunting capacities.

Initially, every particle doles out with necessary qualities, after which, assessment of fit qualities is conducted for each particle. At that point, the worth of the current fit becomes clear, and if it is superior to the previous one, we update it as the present worth, but keep it the same in case the worth of the old fit is better. This process continues until achieving the best possible arrangement.

C. *Convolutional neural network*

The convolution neural network (CNN) has a principle segment called the convolution layer and the process behind it can be described as: In different areas of the equivalent information source, a locally-adapted element of some random contribution (for example, 2-D pictures) can be helpful. For example, an edge identification element that has demonstrated usefulness in a segment of the picture may also be useful in other picture segments, which is valid during the general component extraction stage. In a photo, learning different highlights (edges at a point or a bend) is possible through sliding the channels throughout the photo using a consistent step size.

CNN consists of many subsampling and convolutional layers, alternatively followed by appended layers. In this context, m × m × r is the contribution to a convolutional layer. Here m is both the width and height of a photo and r shows the number of channels; for example, r=3 for an RGB image. There are k channels (or portions) in a convolutional layer and the size is n × n × q. In this case, q is possibly something similar or more modest, and n is more modest as compared to the elements of the picture.

II. RESULTS AND DISCUSSION

In this section, we have illustrated the results obtained after simulation using MATLAB. This method performs three face recognition processes. The first process uses HOG for feature extraction and saves it as a variable. The second process is the feature selection process, and it applies the Harris Hawks Optimization (HHO) for finding the best possible features from the HOG output. Usually, the number of features is high, but they reduce when HHO is applied. In the end, we trained the system and made it classify/identify the faces using the convolutional neural network.

The data were taken from ORL and Yale datasets in this examination. In the ORL (Olivetti Research Laboratory) facial dataset, 400 pictures of 40 unique people and ten diverse grayscale images of the unmistakable people exist. The photos mentioned were taken on different occasions, so they vary in terms of articulation (grinning/not grinning and open/shut eyes) and subtleties of the face (with/without glasses). The mentioned pictures were taken keeping in view the possibility of some shifting and the possibility that the face may turn up to 20 degrees [8].

To assess the proposed method, we used mean absolute percentage error (MAPE), R square, and mean squared error (MSE) approaches. MSE is given below:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y_i})^2 \tag{5}$$

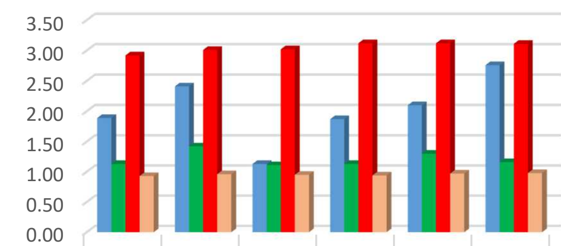
The following equation shows the MAPE:

$$\text{MAPE} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \tag{6}$$

The estimated value of R square is given below:

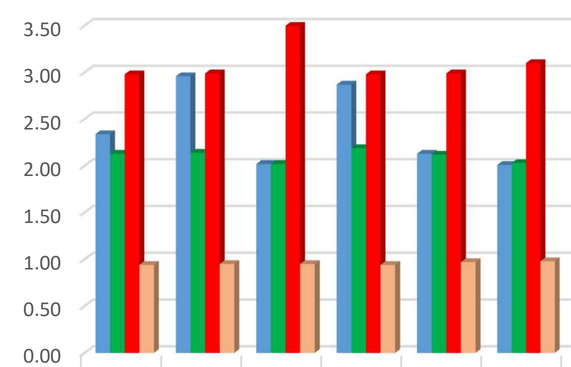
$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \tag{7}$$

This method was performed using genuine balanced photos from ORL and YALE datasets. "Figs 2 and 3" show the outcomes. The ORL dataset's testing procedure has shown two things: How preprocessing improves the exactness and how we can unite or break both techniques to extract features and create an incredible third strategy for achieving our objectives.



	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold
MSE	1.89	2.41	1.13	1.87	2.10	2.76
RMSE	1.13	1.42	1.11	1.13	1.30	1.16
MAPE	2.92	3.01	3.02	3.12	3.12	3.11
R2	0.93	0.96	0.95	0.94	0.97	0.98

Fig. 3. R^2, MAPE, MSE, and RMSE for training



	1-Fold	2-Fold	3-Fold	4-Fold	5-Fold	6-Fold
MSE	2.34	2.96	2.02	2.87	2.13	2.01
RMSE	2.13	2.14	2.02	2.19	2.12	2.03
MAPE	2.98	2.99	3.50	2.98	2.99	3.10
R2	0.94	0.95	0.95	0.94	0.97	0.98

Fig. 4. R^2, MAPE, MSE, and RMSE for testing

It is evident that the results are insufficient in terms of the acknowledgment rate and exactness, but when Gabor wavelet and learning processes were used, they prompted defilement. Along these lines, when the assortment is enormous (14-254). Thus, it is recommended to select the ideal highlights.

III. CONCLUSION

The current study has proposed a face-recognition process, which uses the momentum gradient of an image for feature extraction. The study has used the HHO method for choosing efficient features and applied the convolutional neural network for classification and to train the system. This technique applies the facial balance both in the picture spaces or elements. The element space is an approach to gain the advantages of using facial evenness and improving the acknowledgment rate. Using the YALE and ORL datasets, the trials of the proposed technique were conducted for face recognition, and the exploratory outcomes show the effectiveness of this methodology.

ACKNOWLEDGMENT

The authors appreciate the Istanbul Ayvansaray University and Sfax University for their support.

REFERENCES

[1] S. Ahmed, M. Frikha, T. D. H. Hussein, and J. Rahebi, "Optimum Feature Selection with Particle Swarm Optimization to Face Recognition System Using Gabor Wavelet Transform and Deep Learning," *Biomed Res. Int.*, vol. 2021, 2021.

[2] S. Allagwail, O. S. Gedik, and J. Rahebi, "Face recognition with symmetrical face training samples based on local binary patterns and the Gabor filter," *Symmetry (Basel).*, vol. 11, no. 2, p. 157, 2019.

[3] U. Jayaraman, P. Gupta, S. Gupta, G. Arora, and K. Tiwari, "Recent development in face recognition," *Neurocomputing*, vol. 408, pp. 231–245, 2020.

[4] J.-Y. Gan and Y.-W. Zhang, "A new approach for face recognition based on singular value features and neural networks," *Acta Electron. Sin.*, vol. 32, no. 1, pp. 170–173, 2004.

[5] J. Gan, Y. Zhang, and S. Mao, "Adaptive Principal Components Extraction Algorithm and Its Application In the Feature Extraction of Human Face," *Acta Electron. Sin.*, vol. 30, no. 7, pp. 1013–1016, 2002.

[6] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006.

[7] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Face recognition from a single image per person: A survey," *Pattern Recognit.*, vol. 39, no. 9, pp. 1725–1745, 2006.

[8] Z. Liu, J. Pu, Q. Wu, and X. Zhao, "Using the original and symmetrical face training samples to perform collaborative representation for face recognition," *Optik (Stuttg).*, vol. 127, no. 4, pp. 1900–1904, 2016.

[9] L. Liu, P. Fieguth, Y. Guo, X. Wang, and M. Pietikäinen, "Local binary features for texture classification: Taxonomy and experimental study," *Pattern Recognit.*, vol. 62, pp. 135–160, 2017.

[10] Y. Sun and J. Yu, "Facial expression recognition by fusing Gabor and local binary pattern features," in *International Conference on Multimedia Modeling*, 2017, pp. 209–220.

[11] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. image Process.*, vol. 19, no. 6, pp. 1635–1650, 2010.

[12] S. Anila and N. Devarajan, "Preprocessing technique for face recognition applications under varying illumination conditions," *Glob. J. Comput. Sci. Technol.*, 2012.

[13] M. M. Kasar, D. Bhattacharyya, and T. H. Kim, "Face recognition using neural network: a review," *Int. J. Secur. Its Appl.*, vol. 10, no. 3, pp. 81–100, 2016.

[14] X. Zou, J. Kittler, and K. Messer, "Illumination invariant face recognition: A survey," in *2007 first IEEE international conference on biometrics: theory, applications, and systems*, 2007, pp. 1–8.

[15] Y. Z. Goh, A. B. J. Teoh, and M. K. O. Goh, "Wavelet-based illumination invariant preprocessing in face recognition," in *2008 Congress on Image and Signal Processing*, 2008, vol. 3, pp. 421–425.

[16] H. R. E. Doost and J. Rahebi, "An efficient method for texture classification with local binary pattern based on wavelet transformation," *Int. J. Eng. Sci. Technol.*, vol. 4, no. 12, pp. 4881–4885, 2012.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, 2005, vol. 1, pp. 886–893.

[18] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," in *2016 international conference on wireless communications, signal processing and networking (WiSPNET)*, 2016, pp. 2264–2268.

[19] P. Yang and G. Yang, "Feature extraction using dual-tree complex wavelet transform and gray-level co-occurrence matrix," *Neurocomputing*, vol. 197, pp. 212–220, 2016.

[20] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features," in *European conference on computer vision*, 2006, pp. 404–417.

[21] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[22] A. Eleyan and H. Demirel, *Pca and lda based neural networks for human face recognition*, vol. 558. INTECH Open Access Publisher, 2007.

[23] S. J. Dixon and R. G. Brereton, "Comparison of performance of five common classifiers represented as boundary methods: Euclidean distance to centroids, linear discriminant analysis, quadratic discriminant analysis, learning vector quantization and support vector machines, as dependent on," *Chemom. Intell. Lab. Syst.*, vol. 95, no. 1, pp. 1–17, 2009.

[24] G. Mehta and S. Vatta, "An Introduction to a Face Recognition System using PCA, FLDA, and Artificial Neural Networks," *IJARCSSE*, vol. 3, no. 5, 2013.

[25] S. Meshgini, A. Aghagolzadeh, and H. Seyedarabi, "Face recognition using Gabor-based direct linear discriminant analysis and support vector machine," *Comput. Electr. Eng.*, vol. 39, no. 3, pp. 727–745, 2013.

[26] O. Nikisins and M. Greitans, "Local binary patterns and neural network-based technique for robust face detection and localization," in *2012 BIOSIG-Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1–6.

[27] C. Ravat and S. A. Solanki, "Survey on different methods to improve the accuracy of the facial expression recognition using artificial neural networks," in *Proceedings of the National Conference on Advanced Research Trends in Information and Computing Technologies*, 2018, vol. 4, no. 2.

[28] C. Tomasi, "Histograms of oriented gradients," *Comput. Vis. Sampl.*, pp. 1–6, 2012.

[29] A. A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, and H. Chen, "Harris hawks optimization: Algorithm and applications," *Futur. Gener. Comput. Syst.*, vol. 97, pp. 849–872, 2019.

[30] S. R. A. Ahmed and E. Sonuç, "Deepfake detection using rationale-augmented convolutional neural network," Applied Nanoscience, Sep. 2021.

[31] M. T. Mahmood, S. R. A. Ahmed, and M. R. A. Ahmed, "Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism," IOP Conference Series: Materials Science and Engineering, vol. 928, no. 2, p. 022027, Nov. 2020.

[32] S. Khalid Abdulateef, S. R. Ahmed AHMED, and M. Dawood Salman, "A Novel Food Image Segmentation Based on Homogeneity Test of K-Means Clustering," IOP Conference Series: Materials Science and Engineering, vol. 928, no. 3, p. 032059, Nov. 2020.

[33] M. R. AHMED, S. R. AHMED, A. D. DURU, O. N. UÇAN, and O. BAYAT, "An Expert System to Predict Eye Disorder Using Deep Convolutional Neural Network," Academic Platform Journal of Engineering and Science, vol. 9, no. 1, pp. 47–52, Jan. 2021.

[34] M. Waleed, A. S. Abdullah, and S. R. Ahmed, "Classification of Vegetative pests for cucumber plants using artificial neural networks," 2020 3rd International Conference on Engineering Technology and its Applications (IICETA), Sep. 2020.

[35] H. Mechria, M. Gouider, and K. Hassine, "Breast Cancer Detection using Deep Convolutional Neural Network," Proceedings of the 11th International Conference on Agents and Artificial Intelligence, 2019.

[36] S. R. A. Ahmed, O. N. UÇAN, A. D. DURU and O. BAYAT, "BREAST CANCER DETECTION AND IMAGE EVALUATION USING AUGMENTED DEEP CONVOLUTIONAL NEURAL NETWORKS", AURUM Mühendislik Sistemleri ve Mimarlik Dergisi, vol. 2, no. 2, pp. 121-129.