Indonesian Journal of Electrical Engineering and Computer Science
Vol. 38, No. 2, May 2025, pp. 1289~1298
ISSN: 2502-4752, DOI: 10.11591/ijeecs.v38.i2.pp1289-1298

Face recognition based on landmark and support vector machine

Hassan Afifi[1], Abdallah Marhraoui Hsaini[2], Mostafa Merras[2], Aziz Bouazi[2], Idriss Chana[2]
[1]Department of Computer Science, High School of Technology, Moulay Ismail University, Meknes, Morocco
[2]IMAGE Laboratory, Department of Computer Science, High School of Technology, Moulay Ismail University, Meknes, Morocco

Article Info

Article history:

Received Jun 26, 2024
Revised Nov 7, 2024
Accepted Nov 24, 2024

Keywords:

Deep learning
Face recognition
Features extraction
Landmark
Support vector machine
Transfer learning

ABSTRACT

Nowadays, the fast development of face recognition technologies used in fields such as security and video surveillance, gives us many theories and algorithms, a view of these algorithms provides us with an idea of their performance and limitations. In this paper, we will develop a new face recognition approach using the face estimation landmark algorithm to detect faces in real-time videos. Then, we use a pre-trained neural network to extract the 128 facial features of each face detected in the database images and register each vector of 128 values with the corresponding person's name. Then, we form the linear support vector machine (SVM) classifier to recognize faces. Extensive experiments on real and generated data are presented to demonstrate the quality of the proposed method in terms of accuracy, reliability, and speed.

Corresponding Author:

Hassan Afifi
Department of Computer Science, High School of Technology, Moulay Ismail University
Meknes, Morocco
Email: hassaninf447@gmail.com

1. INTRODUCTION

The current development of computer technology and multimedia equipment such as cameras, has become indispensable everywhere and for everyone and is accompanied by the rapid development of artificial intelligence in many areas such as robotics [1], health [2], aviation [3], military [4], and industry [5]. Computer vision is also keeping pace with these developments. It has even benefited greatly from machine learning (ML) and deep learning (DL) algorithms that help computers see, understand, and process the content of images and videos in a very sophisticated way. One of the most common uses of computer vision is face recognition, which matches a person's face with their identity, this technology can be found today in many fields, and it has become the most popular of the authentication systems, such as eye print, voice print or fingerprints, because it respects people's privacy and does not require direct contact with sensors, and also benefits from the generalization of cameras and the progress of their manufacture. These and other factors prompt researchers to develop powerful applications for video surveillance and access control able to confront challenges such as occlusion, variation of poses, distance face to camera, illumination variances, facial expressions, and especially the execution time which decides the possibility of these applications to work in real-time. These challenges make the first step of face recognition technologies, which is face detection, the most difficult and most important step on the way to the success of any algorithm to achieve the desired accuracy in facial recognition applications, followed by the step of face features extraction, which aspires to extract the necessary measures from the image part framed by the detector that expresses the most important features of the face to facilitate the process of classification with measures trained from the database. These operations, which express the working steps of the face recognition algorithms are classified into three categories according to the community of researchers in this field.

The first is the traditional algorithms [6]–[9] these approaches need hand-craft features to reduce the complexity of the data so that ML algorithms such as support vector machine (SVM) and k-nearest neighbor (KNN) can train on these data, and small databases are preferred for these algorithms. The second is DL, especially the convolution neural networks (CNN) [10]–[13] with the advent of this approach the problems of face recognition become a thing of the past thanks to the huge success of this technique in face detection, feature extraction, and specifically in image classification. The main advantage of DL algorithms is that they attempt to learn high-level features from big data incrementally. The third category is the hybrid methods [14]–[18], which are special algorithms among face recognition techniques because they combine hierarchically or simultaneously different approaches, taking the performances of DL and the advantages of traditional methods to build models that outperform the own methods. Most of these techniques are unable to couple precision and time in material constraints.

– Problem statement

Traditional ML methods encounter difficulties associated with image quality, including fluctuating illumination, partial obstructions, and facial representations at varying proximities from the camera. Although DL and hybrid methodologies can surmount some of these obstacles, they analyze the entirety of the image, resulting in the extraction of extraneous features for facial recognition. This amplifies computational intricacy and diminishes the precision of these methods.

– Purpose of this study

In this work, we propose a new face recognition approach to overcome the challenges cited in the problem. Our approach based on three essential steps, the first is the detection of faces using landmarks, the second step is the extraction of face features using the pre-trained neural network and finally we used the SVM classifier to identify and recognize the face. The main advantages of our approach are: i) this technique is faster, due to the fact that we used landmarks to detect the face, this technique allows us to identify only the homogeneous pixels of the face in the images, which subsequently reduces the time required for extraction features instead of using the full image and ii) our methodology capable to estimate the positions of facial landmarks directly from sparse pixel intensities, which allows it to process faces located with varying distances from the camera. The remainder of this paper is organized as follows. In the section 2, we present the related works. The section 3 describes our approach to face recognition and absence management. We present experimental results in the section 4. The conclusion is presented in section 5.

2. RELATED WORK

In this section, we present a summary of some face recognition approaches. First, we propose methods based on ML which depend a lot on the quality of the image, then we examine some techniques based on DL which are less sensitive to image variations. Finally, we will see some hybrid methods that combine the two previous approaches.

2.1. Traditional approach

Several methods in the first literature have been based on ML algorithms, Chandrakala and Devi [6] propose an approach based on a sequential classification of two classifiers, KNN followed by SVM, and to extract important facial features they used a histogram of oriented gradients (HOG) [19] by adjusting its parameters, but this approach did not mention the face detection step because the authors used images containing faces only, as well as the extraction of the features of faces with e.g. beards or glasses by the HOG descriptor cannot give reliable measurements for face recognition. In the paper [7], face recognition uses the HaarCascade algorithm [20] based on Haar-like features for face detection, then LBPH to extract face features, and finally KNN to recognize them. The detector used is fast and can detect faces located far away in the image or from the camera. However, it predicts many false positives and gives unconcerned areas to the LBPH descriptor, which reduces the accuracy of this method. The manuscript [8] introduces a system that operates in real-time employing the Viola-Jones algorithm for facial detection, utilizing color space for tracking, and employing eigenfaces for recognition within uncontrolled settings, resulting in a success rate of 91.12%. However, the duration of execution is subject to fluctuations contingent upon the background and specifications of the camera, while the rate of recognition may differ based on the outcomes of both training and testing. Thammaiah and Nagavi [9] presents a method that combines HOG and Euclidean distance to recognize partially occluded faces in biometric systems. However, this approach may face challenges with heavily occluded faces or under unfavorable lighting conditions. Additionally, using Euclidean distance to compare feature vectors may not be the most effective method for all types of occlusions.

2.2. Deep learning approach

Currently, most of the techniques in the literature for face recognition are based on DL approaches, Khan *et al.* [10] propose an approach based on a CNN, which uses a region proposal network (RPN) to detect faces using intersection over union overlap of anchors that have the maximum probability of containing faces. For classification that requires a fixed image size, this problem is solved by resizing the particular part of the image containing the face, using region of interest (ROI) pooling. Despite this approach achieving very good accuracy in the test data, it can only detect 35 faces and recognize 30 out of 40 in the single images. Jose *et al.* [11] present the method for face recognition using a multicamera to detect faces with the help of the multi-task cascaded convolutional neural networks (MTCNN) algorithm, then FaceNet to recognize them. the performance of the MTCNN detector used in this technique is better than Haarcascade and HOG, but it is a bit slower than the latter, it requires more computation than them, which requires using GPU. Peng *et al.* [12] propose a face recognition approach inspired by the inception ResNet model based on two ideas, the first is to change the residual scaling factor in inception-ResNet from a hyperparameter to a trainable parameter and the second idea is to use the leaky ReLU and PReLU activation functions in an alternative way with the ReLU activation function in the inception ResNet model. Both of these changes provide training stability, but they increase the computation and make this model unable to run in real-time. Zhong *et al.* [13] propose an advanced network architecture known as GhostVLAD, specifically designed for set-based face recognition tasks. This architecture utilizes compact representations and integrates ghost clusters, with the purpose of enhancing the processing of low-quality images and ultimately achieving superior performance on intricate datasets. Although ghost clusters are crucial for improving image quality, they do not actively partake in the aggregation process within the GhostVLAD layer. Additionally, the structure incorporates an automated quality weighting mechanism for input faces, which dynamically enhances the overall performance of the network.

2.3. Hybrid approach

Most solutions in the existing literature use hybrid approaches, Moustafa *et al.* [14] focuses on enhancing DL features for age-invariant face recognition (AIFR) using transfer learning, genetic algorithm (GA), and KNN ML for classification. It achieves notable recognition accuracy on FGNET and MORPH datasets without requiring preprocessing steps, using direct input into the VGG-face model. The study highlights the effectiveness of Smart GA for feature selection and optimization in AIFR tasks. Importantly, the research overlooks the face detection step and processes entire image surfaces indiscriminately, resulting in non-significant values that reduce model performance. Additionally, image properties such as lighting significantly influence recognition performance in the MORPH dataset. In the paper [16], the face recognition technique uses eigenface values for face detection, principal component analysis (PCA), and CNN to recognize faces. The implementation of PCA, which reduces the measurements, will not help neural networks extract the best features from images. Shawhney *et al.* [15] compare the performance of two libraries dlib and openCV in face recognition using an approach based on the histogram of the oriented gradient for face detection, estimation landmarks for orientation, a pre-rained neural network to extract the features of the faces, and classification with the SVM. The latter is fast and produces better classification results for the 128-value vectors provided by the extractor, which allows this method to work in real-time, but the HOG detector used does not allow the detection of side faces, those located far from the camera and the images with illumination variances, which limits its applications. Abuzneid and Mahmood [17] propose a hybrid approach using an LBPH descriptor, multi-KNN, and BPNN neural network. the principle of this method is the generation of a new set called the T-Dataset from the original training dataset, which is used to train the BPNN. The new T-Dataset is obtained by using the correlation between the training images without using a common technique of image density. The correlated T-Dataset provides a high distinction layer between the training images, which helps the BPNN to converge faster and achieve better accuracy. Nguyen *et al.* [18] offers an enhanced system that uses the fusion of RGB, HSV and local binary patterns (LBP) features to increase facial recognition performance, Furthermore, it utilizes the FaceNet DL architecture along with the SVM for facial localization and identification. This approach attains a detection accuracy of 96.25% on a tailored dataset. However, the fusion adopted by this model faces challenges such as the heterogeneity of RGB, HSV, and LBP features, as well as overlearning due to high dimensionality. Elaggoune *et al.* [21] propose a novel approach to improving facial recognition using hybrid descriptors that combine various sources of information in an optimal way. Two hybrid structures have been developed: the first integrates traditional descriptors such as the Gabor filter with HOG, LPQ and PCA to extract facial features, while the second uses transfer learning with CNN (AlexNet) to capture the most relevant features. Features are then optimized by PSO, followed by data reduction via LDA for the first structure and classification with "Softmax" for the second, but the optimization using PSO is highly influenced by how parameters are configured and meta-parameters are selected, which directly impacts how efficiently features are extracted and the overall performance of the face recognition system.

3. PROPOSED METHOD
3.1. Data acquisition
Datasets play a crucial role in measuring the performance of facial recognition techniques, so we have chosen to evaluate two datasets, The first is labeled faces in the wild (LFW), and the second, is a concrete data set consisting of a collection of images of our students. these images are in RGB, BVR, and HSV color mode and have almost the same resolution.

3.2. Data preprocessing
Currently, our dataset contains 27 images, one for each student. To improve image quality and eliminate noise from the acquisition phase, we applied a median filter to all images. This treatment makes the images sharper and more precise, facilitating their use in future analyses.

3.3. Data augmentation
In order to build the most accurate and robust model possible, it is necessary to have a large set of images at our disposal. For this we have used data augmentation, which allows us to avoid over-fitting and improve generalization, using techniques such as geometric transformations, color adjustments, flipping, cropping and noise addition. These functionalities are available in the Keras library's ImageDataGenerator class, which enables us to generate new labeled images from those already available to obtain a generated database of 1,620 images.

3.4. Framework architecture
The aim of our paper is to propose a new approach to face recognition that confronts the challenges associated with image quality, such as illumination variations, partial obstructions and facial representations at different camera distances. Although some approaches overcome these difficulties, they analyze the image in its entirety, which can lead to the extraction of irrelevant features, thus increasing the computational complexity and reducing the accuracy of facial recognition systems. To this end, an approach has been adopted that uses landmarks to detect faces, then a pre-trained neural network to extract their features and classify them using a linear SVM, as illustrated in Figure 1.

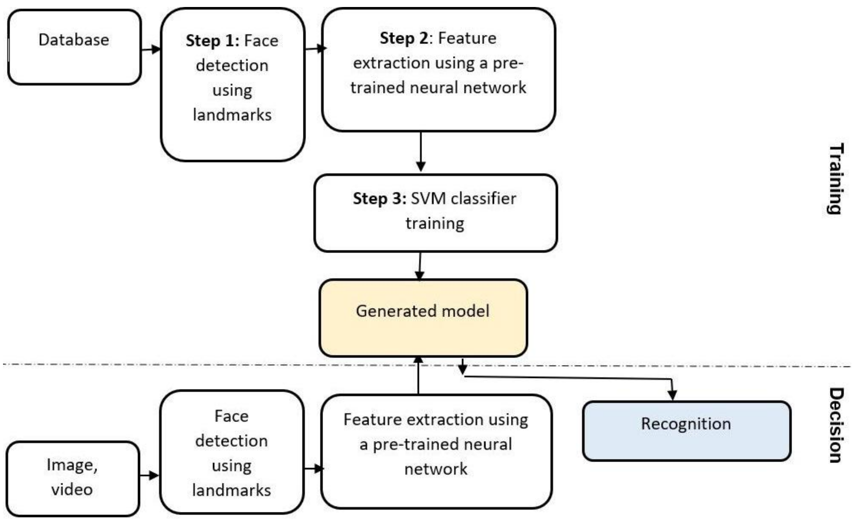

Figure 1. Flowchart of the proposed approach

- Step 1: face detection using landmark: the process of facial detection involves identifying the presence of a face in an image. This operation plays a crucial role in the process of facial recognition, as it needs to be performed with both efficiency and accuracy. Given the multitude of available face detection algorithms, our primary objective is to evaluate three specific ones. To this end, we're going to run a demonstration in Figure 2 that will attempt to show the differences between the HaarCascade classifier developed by Paul Viola and Michael Jones. HOG+SVM detector, and the MMOD. Once the results have been obtained, we can see that the CNN detector is perfect, it detects all the faces, even the lateral ones, but it takes longer and requires more GPUs. On the other hand, the HOG and HaarCascade detectors are fast but less accurate than the CNN detector, as they can't detect faces located far away or laterally. In this case, we can use an algorithm to solve this problem that was proposed in 2014 by [22] called the face landmarks estimation algorithm. This algorithm finds 68 important points in the face; these points are called landmarks and are located above the chin-the outer edge of the eye-the inner edge of the eye-the eyebrow-around the mouth-around the nose. These points are found by training a ML algorithm. Then, we will center the face by rotating these landmarks of the eyes and the mouth as much as possible; and for the coordinates of the faces provided to the descriptor to be correct, we must locate the faces by the landmarks taking into account the resizing necessary for the pre-trained neural network of the images captured by the webcam, this will help us a lot in the next step by making it more precise and efficient. Face detection using landmarks shows that this detector is perfect in terms of accuracy in Figure 3 and time in Table 1, as it can detect all the faces and only the faces in the image, even the laterals, and faces located far from the camera, in a reasonable time compared with other detectors, enabling us to exploit it for real-time face recognition applications.

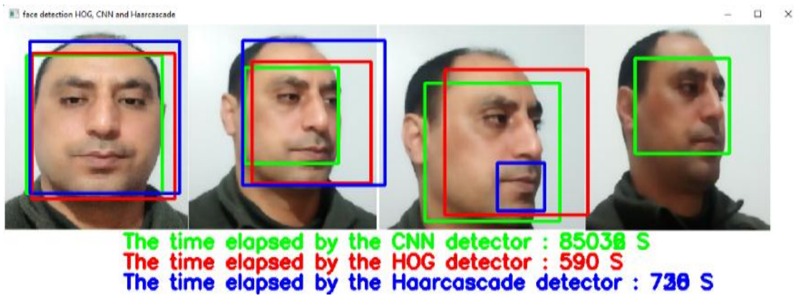

Figure 2. Face detection using various detectors

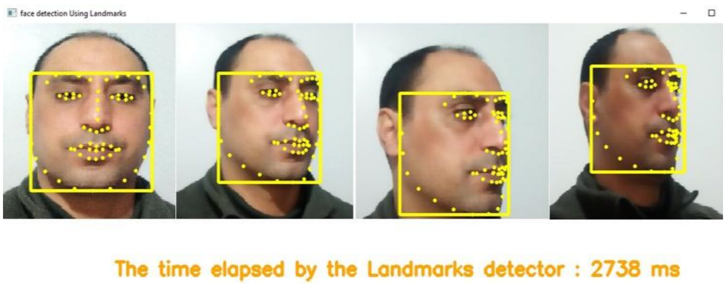

Figure 3. Landmark detector results

Table 1. Time elapsed by different detectors

Detector	Time elapsed (ms)
CNN detector	85,038
HOG detector	590
HaarCascade detector	720
Landmarks detector	2,738

- Step 2: features extraction: the easiest way to recognize the face is to compare the resulting image with all the images in the database (for example, we can measure the size of each ear or the space between our eyes) but the comparison process can be very time-consuming. A solution to this problem was proposed in 2015 by Google researchers, which is to build and train the folding neural network on many face images, to extract a ray of 128 dimensions for each face, called the embedding ray, which reduces the image with its large data to a few digits, so that instead of comparing the image completely with the images in the database, the comparison is the modulation ray of the entered face image with the modulation rays of each face image in the database. In our project, we will rely on a pre-trained neural network model to do this task. This model was inspired by the ResNet-34 model by Davis E. King who removed some layers and reconstructed a neural network composed of 29 convolution layers. It expresses inputs of size 150×150×3 and represents the face images as vectors of 128 dimensions. The model has already been trained on millions of training images and can therefore create reliable encodings for faces that the model has never seen before. Images of the same person should yield approximately the same encoded vector of 128 values. The process looks a bit like as shown as in Figure 4.

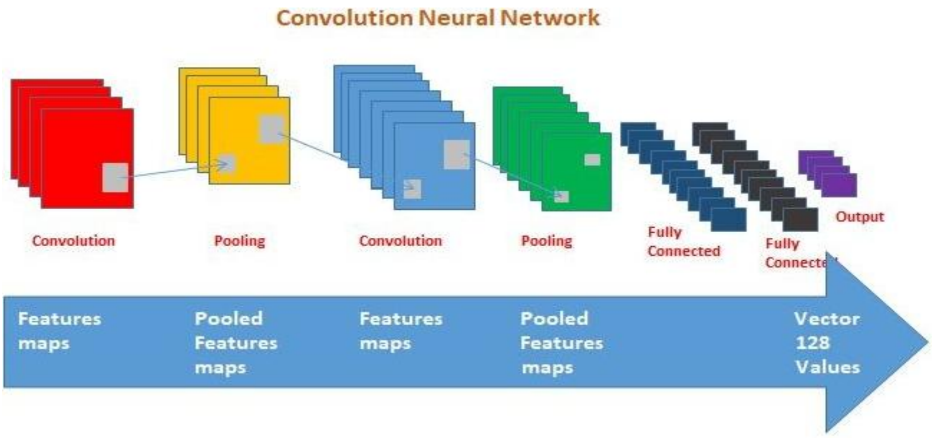

Figure 4. Pre-trained neural network for features extraction

- Step 3: classification: finally, to obtain the name of the person owning the face, we choose to use the linear SVM in Figure 5.

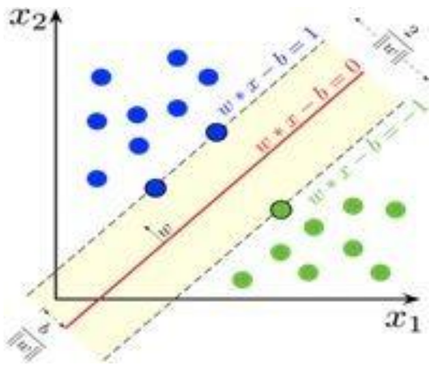

Figure 5. SVM classifier

The principle underlying the operation of SVMs is to identify a hyperplane in a multidimensional space that effectively separates the different classes. The selection of this optimal hyperplane is determined by the support vectors, which represent the data points closest to the hyperplane of each class. The main objective of SVM is to maximize the margin, defined as the distance between the hyperplane and the support vectors. Choosing the hyperplane with the largest margin ensures that the SVM can generalize well, enabling it to achieve superior performance when confronted with new data. The choice of the classifier is based on two reasons, the first is the efficiency of the SVM in unstructured data sets compared to the decision tree (DT), and the second is the efficiency of high dimension management: SVMs can work well even with a large number of features (dimensions), as they are only influenced by a subset of support vectors. This makes them suitable for high-dimensional data. On the other hand, decision trees and KNNs can suffer from the curse of dimensionality when the number of features is high, which can lead to a drop in performance. Using the available features, 128-dimensional vectors of face images in the database and we train the classifier, after which we can predict the faces detected in the video, to improve the predictive accuracy of the SVM classifier, we used GridSearchCV to optimize its hyperparameters. SVM classifier parameters: Kernel='linear'; C=0.1.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. Dataset description

The LFW dataset is widely used as a standard for evaluating the reliability and precision of facial recognition algorithms. It contains 13,233 labeled images depicting the faces of 5,749 individuals, demonstrating a wide range of variations in facial expressions, lighting situations, poses, and backgrounds. Researchers have unrestricted access to this dataset for academic and research purposes.

4.2. Experiments

In our study, we conduct a series of in-depth experiments to evaluate the performance of our proposed method. These experiments are carried out using two distinct datasets: the LFW face database [23] and a customized database specifically designed for real-time absence management, and we compare the effectiveness of our method with other methods using algorithms such as HOG, spatial local binary patterns (SLBP), LBP, and KNN on the LFW database, the results are presented in Table 2. With the material used laptop : i) processor: Intel(R) Core(TM) i5-4300U CPU @ 1.90GHz (4 CPUs), 2.5GHz; ii) memory: 12GB RAM; iii) webcam HD resolutions up to 1280×720 (720p) and up to 30 fps; and iv) software: PyCharm

Accuracy is a metric employed to delineate the proximity of measurements to a particular value. Additionally, it is characterized as the quotient of accurately matched samples divided by the total samples within the dataset of samples.

$$\text{Accuracy} = \frac{\text{Number of correctly recognized faces}}{\text{Total number of f aces}} * 100$$

Table 2. Accuracy comparison between different face recognition approaches

N°	Method	Dataset	Accuracy (%)
1	HOG+SVM [24]	LFW	91.83
2	HOG+KNN+SVM [6]	LFW	95.2
3	SLBP+HOG+SVM [25]	LFW	95.7
4	LBP+KNN [26]	LFW	95.71
5	Our method	LFW	99.25
6	Our method	Our dataset	97.98

4.3. Discussion

Previous methods for face recognition have mainly focused on traditional feature descriptors such as HOG, LBP, SLBP, Haar cascade, and KNN, often in close-range face recognition contexts. However, while these techniques have demonstrated commendable effectiveness under controlled circumstances, they have considerable limitations when analyzing faces at considerable distances or under varying lighting and partial occlusion conditions. For example, some HOG or LBP derived methodologies may fail in situations where faces are partially obscured or located at long distances from the camera in Figure 6. Previous studies have not explicitly considered the implications for the effectiveness of face recognition systems of detecting faces at varying distances and accurately recognizing them in complex environments.

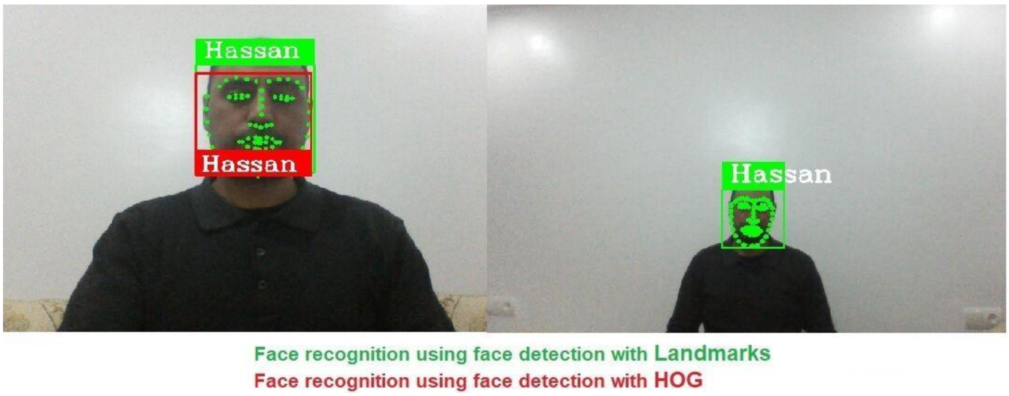

Figure 6. Face recognition comparison between HOG and landmarks

Our study suggests that the use of landmarks to detect faces, a pre-trained neural network to extract features from faces and the use of SVM as a classifier lead to higher accuracy than methods based on HOG, LBP, SLBP or KNN as shown in Table 2. In particular, the approach combining landmarks, a pre-trained neural network and SVM outperforms the other models in terms of classification accuracy (99.25%), compared with models such as KNN (97.78%) and DT (97.03%) as shown in Table 3 and Figure 7. This finding is in line with results observed in previous studies, where the integration of DL techniques or previously trained CNNs offered better performance. However, our approach is based on an interesting compromise without requiring computational resources as high as those required for more complex networks. This enables our method to operate efficiently in real time, unlike other approaches based on heavier deep models, as shown by the performance comparison in Tables 1 and 3.

Table 3. Face recognition accuracy comparison between different classifiers

N°	Detector	Features extractor	Classifier	dataset	Accuracy (%)
1	HOG	Pre-trained neural network	SVM	LFW	88.67
2	Landmarks	Pre-trained neural network	DT	LFW	97.03
3	Landmarks	Pre-trained neural network	KNN	LFW	97.78
4	Landmarks	Pre-trained neural network	SVM	LFW	99.25
5	Landmarks	Pre-trained neural network	SVM	Our dataset	97.98

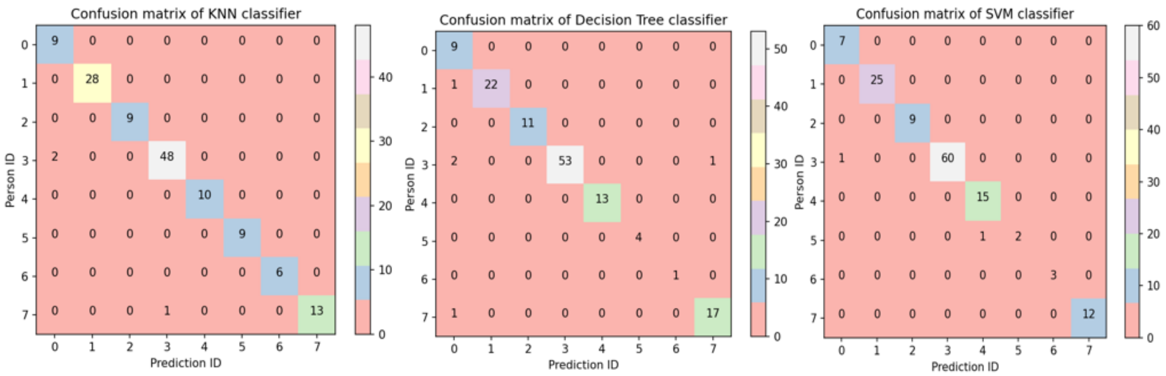

Figure 7. Confusion matrix comparison between SVM, KNN, and DT classifiers

5. CONCLUSION

Our study demonstrates that the methodology amalgamating the extraction of reference points, the utilization of a previously trained CNN for the extraction of features, and the classification via SVM exhibits outstanding performance concerning accuracy and resilience. Our findings furnish compelling substantiation that this technique is particularly well-adapted for real-time applications that necessitate precise facial recognition, even under arduous circumstances, such as variations in pose, distances, and partial occlusions. This approach is excellent for spotting faces compared to other approaches using the HOG which is based on gradient orientation on a dense grid for each pixel or HaarCascade which is based on Haar features. In our environment, this application gives us satisfactory results in terms of accuracy and robustness, and it opens the following future works: i) it is necessary to experiment with it in more real situations to overcome problems such as light and facial expression and ii) another improvement, where the real-life video based facial recognition is used as a resource for collecting the data sets. These data could be used as a source data, necessary to apply an intelligent strategy in a particular engineering problem.

REFERENCES

[1] X. Bin Peng, M. Andrychowicz, W. Zaremba, and P. Abbeel, "Sim-to-real transfer of robotic control with dynamics randomization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*, May 2018, pp. 3803–3810, doi: 10.1109/ICRA.2018.8460528.

[2] P. Esmaeilzadeh, "Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, Dec. 2020, doi: 10.1186/s12911-020-01191-1.

[3] K. N. Amirkolaii, A. Baboli, M. K. Shahzad, and R. Tonadre, "Demand forecasting for irregular demands in business aircraft spare parts supply chains by using artificial intelligence (AI)," *IFAC-PapersOnLine*, vol. 50, no. 1, pp. 15221–15226, Jul. 2017, doi: 10.1016/j.ifacol.2017.08.2371.

[4] A. P. Pope *et al.*, "Hierarchical Rreinforcement learning for air-to-air combat," in *2021 International Conference on Unmanned Aircraft Systems (ICUAS)*, Jun. 2021, pp. 275–284, doi: 10.1109/ICUAS51884.2021.9476700.

[5] T. Tankova and L. S. da Silva, "Robotics and additive manufacturing in the construction industry," *Current Robotics Reports*, vol. 1, no. 1, pp. 13–18, Mar. 2020, doi: 10.1007/s43154-020-00003-8.

[6] M. Chandrakala and P. D. Devi, "Two-stage classifier for face recognition using HOG features," *Materials Today: Proceedings*, vol. 47, pp. 5771–5775, 2021, doi: 10.1016/j.matpr.2021.04.114.

[7] N. Gupta, P. Sharma, V. Deep, and V. K. Shukla, "Automated attendance system using OpenCV," in *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)*, Jun. 2020, pp. 1226–1230, doi: 10.1109/ICRITO48877.2020.9197936.

[8] A. H. T. Al-Ghrairi, N. H. F. Al-Anbari, and E. Z. Sameen, "Realtime human face tracking and recognition system on uncontrolled environment," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, pp. 4246–4255, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4246-4255.

[9] C. H. Thammaiah and T. C. Nagavi, "An approach to partial occlusion using deep metric learning," *International Journal of Informatics and Communication Technology (IJ-ICT)*, vol. 10, no. 3, 2021, doi: 10.11591/ijict.v10i3.pp204-211.

[10] M. Z. Khan, S. Harous, S. U. Hassan, M. U. G. Khan, R. Iqbal, and S. Mumtaz, "Deep unified model for face recognition based on convolution neural network and edge computing," *IEEE Access*, vol. 7, pp. 72622–72633, 2019, doi: 10.1109/ACCESS.2019.2918275.

[11] E. Jose, M. Greeshma, T. P. M. Haridas, and M. H. Supriya, "Face recognition based surveillance system using FaceNet and MTCNN on Jetson TX2," in *2019 5th International Conference on Advanced Computing and Communication Systems, ICACCS 2019*, 2019, pp. 608–613, doi: 10.1109/ICACCS.2019.8728466.

[12] S. Peng, H. Huang, W. Chen, L. Zhang, and W. Fang, "More trainable inception-ResNet for face recognition," *Neurocomputing*, vol. 411, pp. 9–19, Oct. 2020, doi: 10.1016/j.neucom.2020.05.022.

[13] Y. Zhong, R. Arandjelović, and A. Zisserman, "GhostVLAD for set-based face recognition," in *Computer Vision--ACCV 2018: 14th Asian Conference on Computer Vision*, Perth, Australia, 2019, pp. 35–50.

[14] A. A. Moustafa, A. Elnakib, and N. F. F. Areed, "Optimization of deep learning features for age-invariant face recognition," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 2, pp. 1833–1841, Apr. 2020, doi: 10.11591/ijece.v10i2.pp1833-1841.

[15] S. Sawhney, K. Kacker, S. Jain, S. N. Singh, and R. Garg, "Real-time smart attendance system using face recognition techniques," in *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, Jan. 2019, pp. 522–525, doi: 10.1109/CONFLUENCE.2019.8776934.

[16] N. Boyko, O. Basystiuk, and N. Shakhovska, "Performance evaluation and comparison of software for face recognition, based on dlib and opencv library," in *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, Aug. 2018, pp. 478–482, doi: 10.1109/DSMP.2018.8478556.

[17] M. A. Abuzneid and A. Mahmood, "Enhanced human face recognition using LBPH descriptor, multi-KNN, and back-propagation neural network," *IEEE Access*, vol. 6, pp. 20641–20651, 2018, doi: 10.1109/ACCESS.2018.2825310.

[18] Q. N. Nguyen, N. C. Debnath, and V. D. Nguyen, "Face recognition based on deep learning and HSV color space," 2023, pp. 171–177.

[19] 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)*, vol. 1, pp. 886–893, doi: 10.1109/CVPR.2005.177.

[20] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, doi: 10.1109/CVPR.2001.990517.

[21] H. Elaggoune, M. Belahcene, and S. Bourennane, "Hybrid descriptor and optimized CNN with transfer learning for face recognition," *Multimedia Tools and Applications*, vol. 81, no. 7, pp. 9403–9427, Mar. 2022, doi: 10.1007/s11042-021-11849-1.

[22] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2014, pp. 1867–1874, doi: 10.1109/CVPR.2014.241.

[23] G. B. Huang, M. Ramesh, T. Berg and Erik Learned-Miller, "Labeled faces in the wild: a database for studying face recognition in unconstrained environments," 2007, [Online]. Available: http://www.tamaraberg.com/papers/lfw.pdf.

[24] B. T. Hung, "Face recognition using hybrid HOG-CNN approach," pp. 715–723, 2021.

[25] M. Annalakshmi, S. M. M. Roomi, and A. S. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Computing*, vol. 22, no. 1, pp. 11–20, Jan. 2019, doi: 10.1007/s10586-017-1585-x.

[26] I. L. K. Beli and C. Guo, "Enhancing face identification using local binary patterns and K-nearest neighbors," *Journal of Imaging*, vol. 3, no. 3, Sep. 2017, doi: 10.3390/jimaging3030037.

BIOGRAPHIES OF AUTHORS

Hassan Afifi is a Ph.D. student in computer science at High School of Technology, (EST) Meknes, Moulay Ismail University Morocco. He received the master's degree in education technology in 2020 and Maîtrise in software engineering in 2004. His research interests include deep learning and machine learning and hybrid technologies for face detection, face recognition, and computer vision. He can be contacted at email: hassaninf447@gmail.com.

Abdallah Marhraoui Hsain electronics engineer obtained in 1990 from the University of Technology of Oran (USTO). Qualified professor at the higher school of technology, electrical engineering department in Meknes (ESTM). Member of the Electrical Engineering, Vision and Artificial Intelligence (I.E.V.I.A) and Computer Science, Applied Mathematics Laboratory (I.M.A.G.E). Research areas: electrical engineering, robotics and computer vision. He can be contacted at email: a.marhraouihsaini@umi.ac.ma.

Mostafa Merra received the Ph.D. degree from Sidi Mohamed Ben Abdellah University, Fez, in 2016. He is currently a professor in computer science at the High School of Technology of Meknes, Moulay Ismail University, Meknes, Morocco. He is also a member of the LISAC, IMAGE Laboratory. His research interests include camera self-calibration, 3D reconstruction, genetic algorithms, and real-time rendering. He can be contacted at email: merras.mostafa@gmail.com.

Aziz Bouazi obtained his Ph.D. from University of Le Havre (France) specializing in electrical engineering and computer vision. He is currently a research professor in the Electrical Engineering Department of E.S.T Meknes. His research works mainly focus on artificial vision and intelligence as well as on the control of electrical machines, and he is at the same time responsible for the IEVIA team of the Image Laboratory at My Ismail University. He can be contacted at email: a.bouazi@gmx.fr.

Idriss Chana received the Ph.D. degree from Mohamed V University of Rabat, Morocco in 2013. He is currently an associate professor at Moulay Ismail University of Meknes, Morocco. His research interests include information and communication technologies and artificial intelligence. A large part of his research projects is related to error correcting codes and Turbo codes. Idriss Chana has published more than 40 papers in major journals and conferences in information theory and artificial intelligence. He can be contacted at email: idrisschana@gmail.com.