



HAL
open science

Face Recognition In Databases of images With Hidden Markov's Models

Mbietieu Amos, Tapamo Kenfack, Eone Oscar

► **To cite this version:**

Mbietieu Amos, Tapamo Kenfack, Eone Oscar. Face Recognition In Databases of images With Hidden Markov's Models. 2022. hal-03589836

HAL Id: hal-03589836

<https://hal.science/hal-03589836>

Preprint submitted on 25 Feb 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Face Recognition In Databases of images With Hidden Markov's Models

Mbietieu Amos Mb.^{1,2}, Tapamo Kenfack H. M.^{1,2}, Eone Oscar E. V.^{1,2}

¹Ummisco, University of Yaoundé 1 - Cameroon

² Computer Science and Application Laboratory, University of Yaoundé 1 - Cameroon

*E-mail : [amos.mbietieu, hippolyte.tapamo, oscar.eone]@facsciences-uy1.cm

Abstract

In this paper, we present a new approach to Facial Recognition (FR) that uses Hidden Markov Models. The method we propose enhances the HMM by integrating a clustering step for the partitioning of each face image before building the associated model. We then propose an algorithm for converting the HMM model into a vector in order to apply the usual metrics such as the Euclidean distance and the Manhattan distance in the faces model comparison phase. We apply this algorithm to a publicly available image databases FERET, which allowed us to obtain **100.00%** True Positive Rates (TPR) and **99.98%** True Negative Rates (TNR) for our enhanced HMMs method, compared to 100.00% of TPR and 99.95% of TNR with Grid+HMMs, then 77.14% of TPR and 79.05% of TNR with Discrete Markov Models. These performances are obtained on the basis of their confusion matrix, then the sensitivity, specificity, the accuracy, th precision and the F1-Score of each of these methods.

Keywords

Database, Elbow Method, Face Recognition, Hidden Markov Models (HMMs), Image Histogram, K-means Algorithm.

I INTRODUCTION

In the ecosystem of technologies aimed at improving our daily lives, the safety of people and properties has remained a major issue. By exploiting these technologies that are available to us, we can provide mechanisms to identify a person uniquely and provide means to secure our environment. For a database of people's face images, we want to know if a new image from the outside matches one of the images encoded in the database. During these last decades characterized by technological advances, several intelligent systems have

emerged such as smart homes, automatic cars, smartphones, and many others that are thus equipped with advanced security systems such as fingerprint recognition and FR. It is therefore a question of setting up an authentication mechanism for an individual's access to these systems. For example, in our environment within the confines of University of Yaoundé 1, courses of the majority of first-year students take place in lecture halls with an average capacity of 500 seats. Due to the large size of the class, it is obvious that intruders can easily disrupt the smooth progress of a course. A face recognition mechanism would capture potential perpetrators and generate an alert.

The fact that the face has variable geometry and that the histograms also model the classification problems are the main reasons that allowed us to use k-means algorithms to find the different clusters in an image then associates a histogram for each class obtained. For the relevance of the results, we use the color images in the (R, G, B) color space; and the performance of this approach has led to very good conclusions, which offers a boost and shows the relevance of the work that we will have to do in this topic.

As the remaining of this paper, the section 2 is devoted to the related work. This review allows us to note that there is a large ecosystem of solution approaches around the problem of FR. We devoted the third section of our paper to the presentation of our approach for comparison of two images based on their HMM. In the fourth section, we presented the experimental results based on our approach compared to the existing ones, in order to put in phase the interest of inserting a clustering step in the FR process.

II RELATED WORK

Several lines of research have attempted the problem of face recognition. The oldest methods used the approaches based on the principles of statistics and linear algebra, then other approaches were based on connectionist methods, and finally stochastic approaches [1, 2, 13]. A FR system is made up of two fundamental phases: The learning phase and the recognition phase.

During the learning phase, the systems which use the approaches of statistics and linear algebra make use of a good number of algorithms such as PCA, LDA, ICA[2, 19] for the projection of the images learned in a database of the system. During the recognition phase, the appropriate similarity measures are used in order to carry out a comparison of the image models in databases. Existing approaches combine global and local methods [2, 17, 19].

The problem of FR took another turn in 2014 with the DeepFace and DeepID [4] with the advent of deep neural networks. This achieved the SOTA accuracy on the famous LFW benchmark [13, 20], approaching human performance on the unconstrained condition for the first time (DeepFace: 97.35% vs. Human: 97.53%), by training a 9-layer model on 4 million facial images. The concern with this CNN model is that training

requires a very large amount of data as present in this recent survey [20]. In [13, 20], the model is trained on a database of 350 million images, with 17,000 classes. Thus, these models require high power from GPUs.

The techniques used in the previous methods do not significantly exploit the relationships between the local characteristics of the images. Another more effective technique that exploits the drawbacks of the other previous methods is the HMM-based method [1]. It characterizes the face as a random process that varies depending on several parameters. Samaria et al. [1] illustrated the usefulness of HMM-based FR techniques by emphasizing on their method where the face model is subdivided into five overlapping regions, including the nose, mouth, forehead, chin, and eyes. This approach introduced by F. Samaria [1, 17] requires the face to be straight, i.e. under ideal conditions. However, FR systems can be exposed to non favorable environmental conditions [17, 20].

III COMPARISON OF IMAGES BASED ON HIDDEN MARKOV MODEL

The modeling of histograms generally deals with situations where a good number of classes are to be distinguished. In the training phase of each face image, the **Baum-Welch** algorithm which is used, takes as parameters the initial model, the set of observation sequences associated to this image as well as the maximum number of iterations in order to have a new model which “masters” better features of the statistical properties of the organization of the different pixels the image [11, 16]. After that, we compute the normalized similarity rate Ω_{IJ} between two images I and J [3, 11, 16, 17].

3.1 Approach based on grid partitioning

In this work, we started looking for the right parameters that would allow us to obtain good comparison results. This was possible thanks to the determination of the minimum size for the image grid [11, 16]. It consists in fragmenting the image into small size. We tested our model with the FERET dataset (the description can be found in [20]), as it is one of the most used datasets in this field of FR [20]. We then look for the optimal size of the grid and for this experiment, we have simplified the task of finding the size of the generalized $n \times m$ grid to $n \times n$ grids. Thus, we observe that when our images are partitioned with a grid size of $n = 5$ as in [11], almost no image wrapped in the base is recognized during the recognition phase. We have therefore chosen to use in our system the value $n = 9$, which implies that the size of a thumbnail of the grid is on average 25×25 pixels. Having the model associated with our face image, we propose in section (3.3) an algorithm **HMM_to_Vector** which converts an HMM into a vector, then we use the usual metrics 1 and 2 to establish the comparisons.

$$d_{L_1}(x, y) = \sum_{i=1}^N |x_i - y_i| \tag{1}$$

$$d_{L2}(x, y) = \sqrt{\sum_{i=1}^N |x_i - y_i|^2} \quad (2)$$

(1) is the Manhattan distance, and (2) is the Euclidean distance.

3.2 Enhanced HMM: Our approach based on clustering

In a very recent work that we carried out for the resolution of this problem of facial recognition known in the literature, we proposed a recognition model using the metric proposed by Sahraeian et al. [3]. Since this measure does not manage to give a good comparison between two identical images because it quantifies the percentage of colors levels in a bounded interval, we first propose here algorithm 1 which will allow us to go from the HMM of the image to a vector. From this vector obtained, we therefore apply the appropriate metrics, namely the Euclidean distance, the Manhattan distance, the Minkowski distance.

The idea of clustering came to us following the observation that the data we manipulate, which are images, are not labeled; there is no way to distinguish the pixels of an image that looks similar when we have the image by looking at its texture. So there must be a way to classify these pixels in order to distinguish the texture of the image. In the literature, one distinguishes hierarchical and non-hierarchical clustering, so **K-means** is an algorithm for partitioning a set of data into clusters or groups [5, 15, 21]. At least two factors influence the implementation of this algorithm to produce good results at the exit: The number of clusters to be formed at the exit and the problem of initialization of the centroids of each cluster.

(1)Choice of the number of clusters: The method used recently is called “elbow” [9]. It consists of launching the K -means algorithm with several K values and choosing the value of k which minimizes the inter-class distance (BSS) and maximizes the intra-class distance (WSS). Thus, the value $k = 51$ is chosen in the framework of our experiment using the Silhouette index [7].

(2)Initialization of the different centroids: One of the recent versions of k-means algorithm was implemented in 2019, offering another way of initializing centroids [15]. In the implementation of this algorithm, we put two convergences criteria: **setting the maximum number of iterations above which the algorithm stops, and if there is no more mobility between the different centroids, then there is convergence and we finish.**



Figure 1: Application of the k -means algorithm to some image of face

Fig. 1 illustrates as well how the grouping of the different pixels of the image was carried out. After this step, each partition a corresponding histogram, then from the histogram derive the Markov chains, built the initial model and then we process to the training model.

3.3 Proposed algorithm

The idea of this algorithm 1 is to start from an HMM associated with a face image, to produce a vector by combining the properties of the model.

Algorithm 1 : $\text{HMM_To_Vector}(\text{Vec}, S, \Phi, E, \lambda = \{A, B, \pi\})$.

Require: $S, \Phi, E, \lambda = \{A, B, \pi\}$ ▷ S represents the set of sequences.
1: Initialize $k \leftarrow 0$
2: **for each** $seq \in S$ **do** ▷ seq represents the set of observations.
3: **for each** $obs \in seq$ **do** ▷ obs represents an observation.
4: $\gamma \leftarrow 0$
5: **for each** $st \in E$ **do** ▷ E represents the set of states of the model
6: $\gamma + = \Phi[st] * B[st][obs]$
7: **end for**
8: $Vec[k] \leftarrow \gamma$
9: $k \leftarrow k + 1$
10: **end for**
11: **end for**
Ensure: Vec ▷ The transforming of HMM λ into vector

The algorithm therefore takes as parameters an initially empty vector, the set of possible observation sequences for a face image, the stationarity Φ of the model which is in fact a vector whose content is the behavior of the model studied in the long term, all

the states of the model as well as the model λ in question.

The algorithm begins by initializing a variable k at 1 on line 1 to count the different values to provide in the output vector. In lines 2 and 3, the algorithm chooses an observation sequence of S , then for each observation of the sequence, the algorithm calculates in the variable γ from line 4 to line 7 the proportion of time the model spends observing a certain symbol taken from the long-term sequence, regardless of the state from which the symbol is observed. At line 8, the algorithm keeps this proportion in the vector that will be output and increments k at line 9.

This algorithm that we propose has a complexity of $\Theta(|S| \times (\text{Max}_{s \in S} |s|) \times |E|)$ where $|S|$ is the total number of observation sequences for an image, $|s|$ is the number of observations for a given observation sequence and $|E|$ the number of states that make up the image model.

IV EXPERIMENTAL RESULTS

4.1 Proposed model for the FR system

The image comparison methodology that we propose is given in the following figure 2. In

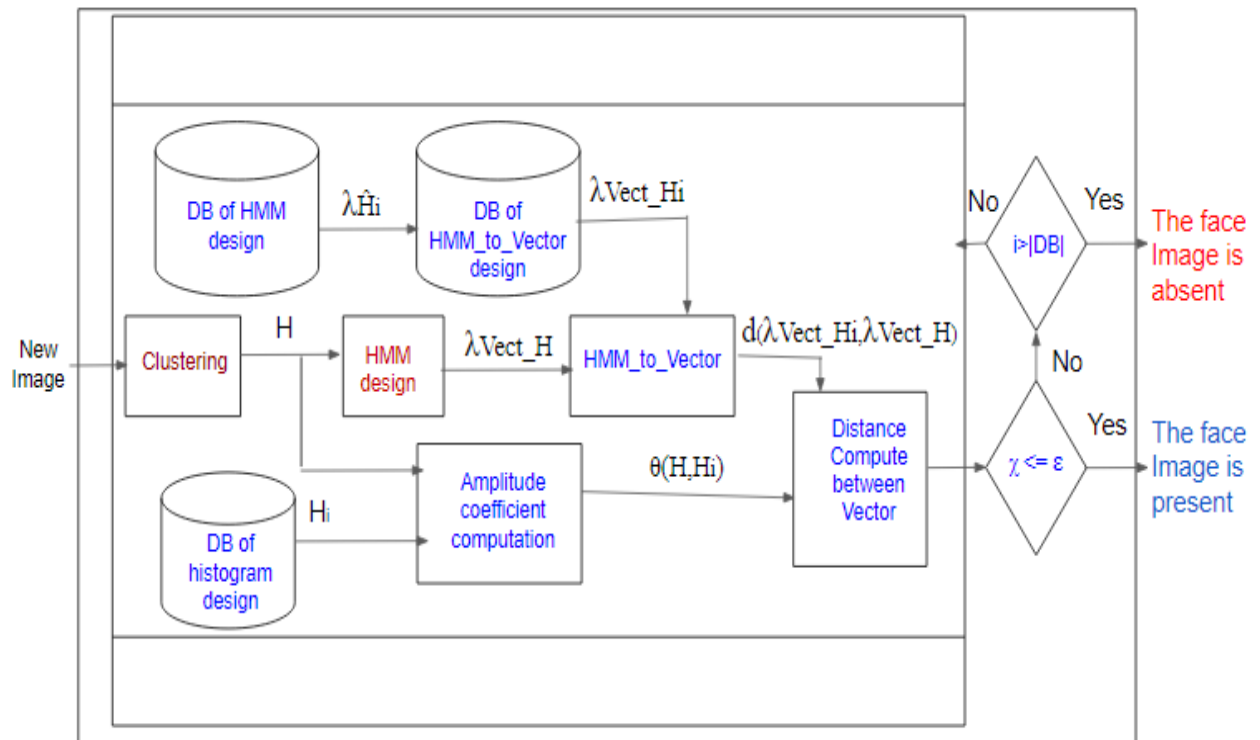


Figure 2: New Image comparison methodology based on our proposed algorithm 1 .

the enrolment, three databases are created for each face image, a database that contains the different histograms of the clusters obtained from the images, a database of the trained

models of these images, and a database that contents HMM of each image transformed into a vector. The system work like this:

1. Clustering of the captured facial image of the individual in question, resulting in the set of derived cluster histograms with the associated HMM and trained model;
2. The system will, therefore, use its three databases, it will choose an image already enrolled in the system, i.e. choose a set of histograms and the corresponding trained model from the respective databases;
3. From this data chosen from the corresponding databases, the system applies the aglorithm 1 to transform the model obtained into a vector associated with the image taken as input. The vector corresponding to the image to which the comparison must be made in the third database is chosen. We thus evaluate the distance between the two vectors, either the Euclidean distance (2), or the Manhattan distance (1).
4. Meanwhile, the corresponding image histogram is chosen from the corresponding database, and with the image histogram taken as input from the system, an amplitude coefficient between the two histograms is calculated;
5. This coefficient is used later to weight the distance calculated in the previous step in order to preserve the starting space of the image data.
6. Based on this weighted distance, a decision is made, if it is below a certain threshold, then the image is recognized by the system. Otherwise, a new image is chosen from the database, ie the set of histograms and the associated model, then the process is repeated until the images in the database are exhausted. If at the end no image corresponds to that taken at the input of the system, the image will be rejected and the person will not be able to access the system.

The results we obtained are therefore recorded in table 1. An **F1-Score** of **99.76%** obtained by our model allows us to understand that globally, an RF system based on this model learns strongly about the entirety of the individuals enrolled in the system, and this is all the more remarkable when we look at the **Precision** and **Accuracy** of this model. A result of **100.00%** of true positives against **99.98%** of false negatives also testifies that the method allows the recognition of what is real face and what is not.

4.2 Synthesis of the results of the different methods

When we look at the summary of the table 2, compared to the approach using statistics and linear algebra, stochastic approaches produce good performances, this is due to the information concerning the color of the image which is lost when switching from a color image to a grayscale image, and the impact of switching from HMM to an vector. This color information has been preserved in our approach using HMMs, but considering the results provided by neural networks, we understand that there is still a lot more information to take into account in our HMM associated to the face image, but also more

Table 1: Comparison of the three methods. (results giving in %)

Methods	TPR	FPR	TNR	FNR	Se	Sp	Accuracy	Precision	F1-Score
DMM	77.14	22.84	79.05	20.95	<i>78.64</i>	<i>77.57</i>	78.10	77.16	77.89
<i>Grid + HMM</i>	<i>100.00</i>	<i>0.82</i>	<i>99.95</i>	<i>0.00</i>	<i>100.00</i>	<i>99.18</i>	<i>99.59</i>	<i>99.19</i>	<i>99.59</i>
Grid + HMM (Grayscale Image)	83.81	16.19	86.67	13.33	86.27	84.26	85.24	83.81	85.03
Enhanced HMM	100.00	0.47	99.98	0.00	100.00	99.53	99.77	99.53	99.76
Enhanced HMM (Grayscale Image)	89.57	11.43	89.23	10.71	89.21	88.65	88.98	88.68	88.90

regarding the Benchmark that we could not have.

V CONCLUSION AND PERSPECTIVES

In this paper, we proposed to enhance the HMMs by initially integrating a clustering step, and the results reveal all the interest of applying an unsupervised learning algorithm to the image in the partitioning phase, to proceed with the other steps. This work used the RGB color space, which made it possible to take into account almost all the information on the image. A perspective that we note for this work thus consists in combining texture

Table 2: Synthesis of the results of the different methods.

Approaches	Method	Dataset used	Accuracy (in %)
Stat. and L. Alg.	PCA + LDA [2, 17]	IFACE	92.64
	High-dim LBP [2]	LFW	95.17
	LG+PCA+LDA [17, 19]	FERET	97.30
	EBGM [17]	LFW	95.25
Neural N.	VGGface [6]	LFW, VGGface Training Set	98.95
	SphereFace [8]	LFW, CASIA-WebFace Training Set	99.42
	Arcface [10]	LFW, CASIA-WebFace Training Set	99.63
Stochastic	DMM [1]	FERET	78.10
	Grid+HMM	FERET	99.59
	Enhanced HMM	FERET	99.77

characteristics based on a model [12], Gabor filters and co-occurrence matrices to better feed the HMM with much more information, and reduce the number of sequential steps in our proposed FR system.

REFERENCES

Publications

- [1] Samaria, F. Fallside, and Frank. *Automated face identification using hidden markov models*. Olivetti Research Limited, 1993.
- [2] Delac, K. G., Mislav, G., and Sonja. “Independent comparative study of PCA, ICA, and LDA on the FERET data set”. In: *International Journal of Imaging Systems and Technology* 15.5 (2005), pages 252–260.
- [3] Sahraeian, S. M. E., and Y. Byung-Jun. “A novel low-complexity HMM similarity measure”. In: *IEEE Signal Processing Letters* 18.11728279 (2011), pages 87–90.
- [4] Y. Taigman, M. Yang, M. R., and L. W. “Deepface: Closing the gap to human-level performance in face verification”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pages 1701–1708.
- [5] Bergounioux and Martine. *Introduction au traitement mathématique des images-méthodes déterministes*. Volume 76. Springer, 2015.
- [6] A. Z. O. M. Parkhi A. Vedaldi. “Deep face recognition.” In: *In: BMVC* (2015).
- [7] Starczewski, Artur, Krzyżak, and Adam. “Performance evaluation of the silhouette index”. In: *International Conference on Artificial Intelligence and Soft Computing*. Volume 9120. Springer. 2015, pages 49–58.
- [8] Liu, Weiyang, Wen, Yandong, Yu, Zhiding, Li, Ming, Raj, Bhiksha, Song, and L. “Sphereface: Deep hypersphere embedding for face recognition”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE, 2017, pages 212–220.
- [9] D. Marutho, S. H. Handaka, E. Wijakay, and Muljono. “The Determination of Cluster Number at k-mean using Elbow Method and Purity Evaluation on Headline News”. In: *2018 International Seminar on Application for Technology of Information and Communication (iSemantic)* (2018).
- [10] Deng, J. G., J. Xue, N. Zafeiriou, and Stefanos. “Arcface: Additive angular margin loss for deep face recognition”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE/CVF, 2019, pages 4690–4699.
- [11] S. Iloga, R. Olivier, and M. Tchuente. “An accurate HMM-based similarity measure between finite sets of histograms.” In: *Pattern Analysis and Applications* 22 (2019), pages 1079–1104.

- [12] Chaki, J. D., and Nilanjan. *Texture Feature Extraction Techniques for Image Recognition*. Springer, 2020.
- [13] O. Arriaga, P. G. Plóger, and M. Valdenegro. “Real-time Convolutional Neural Networks for Emotion and Gender Classification”. In: *In: rXiv:1710.07557v1 [cs.CV]* (20 Oct 2017).
- [14] Ding, Changxing, and T. Dacheng. “A comprehensive survey on pose-invariant face recognition”. In: *ACM Transactions on intelligent systems and technology (TIST)* 7.3 (February 2016), pages 1–42.
- [15] M. K. Gupta and P. Chandra. “P-k-Means: k-Means Using Partition Based Cluster Initialization Method”. In: *Proceedings of International Conference on Advances in Computing* (October 1, 2019).
- [16] S. Iloga. *Apprentissage à l’aide des modèles de Markov cachés et application*. Note de cours, Ecole de Mathématiques Africaine (EMA), Juillet 2019.
- [17] W. D. M. Wang. “Deep Face Recognition: A Survey”. In: *Neurocomputing* 429 (March 2021), pages 215–244.
- [18] Marutho, Dhendra, Handaka, S. Hendra, Wijaya, Ekaprana, et al. “The determination of cluster number at k-mean using elbow method and purity evaluation on headline news”. In: *2018 International Seminar on Application for Technology of Information and Communication*. IEEE. 29 November 2018, pages 533–538.
- [19] C. S. and A. M. S. “Face recognition using complex wavelet moments”. In: *Optics & Laser Technology* 47 (April 2013), pages 256–267.
- [20] M. Wang and W. Deng. “Deep Face Recognition: A Survey”. In: *arXiv:1804.06655v9* (1 Aug 2020).
- [21] G. Wilson and D. J. COOK. “A Survey of Unsupervised Deep Domain Adaptation”. In: *ACM Journals* 11.5 (July 2020), pages 1–46.