Combining Jaccard and Mahalanobis Cosine Distance to Enhance the Face Recognition Rate

ABDELGHAFOUR ABBAD , HAMID TAIRI

Department of Computer Science
University Sidi Mohamed Ben Abdelah
LIIAN, Faculty of Sciences Dhar El Mahraz BP 1796, Fez
MOROCCO
gh.abbad@gmail.com

Abstract:- Facial recognition has become the most dynamic biometric technology. In recent years, it has become a very powerful tool for recognition and authentication of biometric systems. To increase the performance of face recognition algorithms, we propose a new face recognition method which consists in combining, Jaccard and Mahalanobis Cosine distance (JMahCosine). Recognition Rates obtained on a facial recognition system shows the interest of the proposed technique, compared to others methods of literature. Our system has been tested on different databases accessible to the public, namely ORL, YALE and Sheffield.

Key-Words:- Facial biometrics, Dissimilarity Measure, Jaccard distance, Mahalanobis Cosine distance, recognition Rate.

1 Introduction

Biometrics belongs to the category of strong authentication technologies. It is an identification and authentication technology far superior to other methods of confirmation of identity. Among the main biometric technologies we can cite face recognition that has become one of the most important and relevant to several computer scientists.

Several face recognition algorithms and systems have been proposed these last years [1], each one based on a particular representation of the face. We can identify three types of approaches: the global approach where the image of the face is regarded as a vector of characteristics and which is based on methods of reduction of space [2,3,4], and the local approach which consists in applying transformations to specific places of the image such as the corners of the eyes, the nose, or the mouth,… etc [5]. And finally, the algorithms based on hybrid approaches like the modular PCA [6].

All approaches of recognition go through the classification step in which several classifiers were adopted. Among them, there the neural networks [7], the Hidden Markov Models (HMM) [8] and the Support Vectors Machines (SVM) [9]. But the most used are those based on the Euclidean distance. However, the choice of similarity measures play an important role to test the performance of the recognition system [10]. In [11] , Sung-Hyuk Cha presents a variety of distance measures grouped into eight different families, and in [12] , Miller and al have classified the recognition rate of ten measurements of distance to show the success of each one in different database.

In this paper , we propose a new face recognition method which consists in combining, Jaccard and Mahalanobis Cosine distance (JMahCosine). JMahCosine method adopted by this article is tested in a system of face recognition based on Principal Component Analysis (PCA) using different databases- ORL Database, YALE Database and Sheffield Database. The results obtained by the proposed technique are very satisfied compared with results obtained by other distance existing in the literature.

The organization of this paper is as follows: In section 2 we present the similarity measures most commonly used in the field of face recognition. Section 3 presents the principle and the idea of our approach. In Section 4, we present some databases of faces. The last section is devoted to experimental tests with discussions followed by a conclusion.

Table 1. Some examples of distance measures grouped by family

family of Minkowski LP	family L1	family of Intersection	family of Inner Product	family of Fidelity or Squared-chord	family of Squared L2 or $\chi 2$	family of Shannon's entropy	family of the combinations
- Euclidean	- Gower	-Czekanowski	-Harmonic mean	-Hellinger	- Squared Euclidean	- Jeffreys	- Taneja
- City block	- Soergel	- Motyka	- Cosine	- Matusita	- Squared $\chi 2$	- Topsoe	- Avg (L1,L∞)
- Chebyshev	- Canberra	- Ruzicka	- Jaccard	- Fidelity	- Clark	- Jensen-Shannon	- Kumar Johnson
… etc	… etc	… etc	… etc	… etc	… etc	… etc	… etc

2 Dissimilarity measures

Many procedures of statistical analysis are based on the concepts of distance or dissimilarity for a pair of elements. These include clustering, multidimensional analysis, other algorithms of page layout and the methods of detection of the aberrant values. Distance measures or dissimilarity measures are used to compare two lists of numbers (for example vectors), and calculate a single number that evaluates the degree of dissimilarity between them. There are more than 60 different dissimilarity measures and many measures between them are used in the recognition of faces.

In [11], Sung-Hyuk Cha presents a wide variety of distance measures and have classified them on eight families, some of them have given below in table 1.

In our study we are interested to six distances measurements including Euclidean (L2), City Block (L1), Czekanowski, Hellinger, Jaccard and Mahalanobis Cosine. we chose these six distance because they are the most used compared to other distance. The first five distances are detailed in [11] while the Mahalanobis Cosine distance is given in [15].

Let u and v be two vectors of size N.

2.1 Euclidean distance

The Euclidean / L2 distance is the most common in many applications. The Euclidean distance between two vectors u and v in the image space is calculated by Equation 1

$$d_{Euclidean}(u,v)=\sqrt{\sum_{i=1}^{N}\left|u_i-v_i\right|^2} \quad (1)$$

2.2 City Block distance

The City block / L1 distance calculates the distance which would be navigated to go from one point to another following a grid-shaped path. The distance of City block (L1) between two vectors u and v in the space of the image is the sum of the difference of their corresponding elements, as in equation 2.

$$d_{Euclidean}(u,v)=\sum_{i=1}^{N}\left|u_i-v_i\right| \quad (2)$$

2.3 Czekanowski distance

The Czekanowski Distance between two vectors u and v in the space of the image is given by equation 3

$$d_{\text{Czekanowski}}(u,v)=1-\frac{2\sum_{i=1}^{N}\min(u_i,v_i)}{\sum_{i=1}^{N}\left|u_i+v_i\right|}=\frac{\sum_{i=1}^{N}\left|u_i-v_i\right|}{\sum_{i=1}^{N}\left|u_i+v_i\right|} \quad (3)$$

2.4 Hellinger distance

The Hellinger distance (also called Jeffries-Matusita) is similar to the L2 norm, but more sensitive to small changes. The Hellinger distance between two vectors u and v in the space of the image is calculated by equation 4.

$$d_{Hellinger}(u,v)=\sqrt{2\sum_{i=1}^{N}(\sqrt{|u_i|}-\sqrt{|v_i|})^2} \quad (4)$$

2.5 Jaccard distance

The Jaccard index or Jaccard similarity coefficient is used to compare the similarity of a set of data. The Jaccard index was proposed by Paul Jaccard in [16] and developed by Tanimoto for the non-binary case in [17].

The coefficient of similarity of Jaccard between two sets of objects is the result of a division between the number of objects in common by the number of distinct objects in both sets, otherwise said the cardinal of the intersection divided by the cardinal of the union.

We consider two vectors u and v of size N, the Jaccard index is given by:

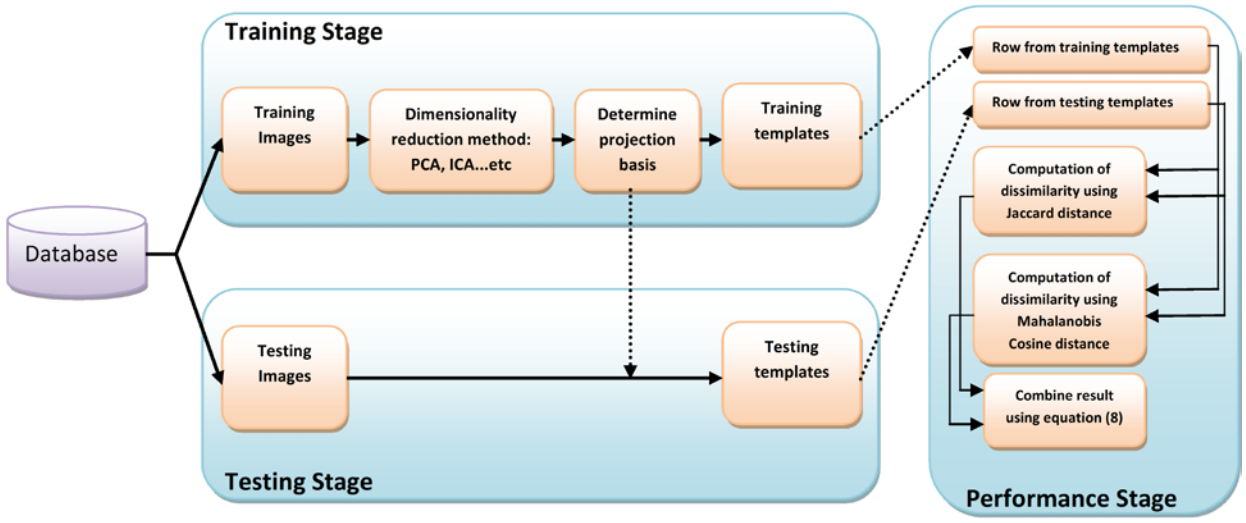


***Fig. 1**. General block diagram of proposed face recognition system*

$$S_{Jac} = \frac{\sum_{i=1}^{N} |u_i v_i|}{\sum_{i=1}^{N} u_i^2 + \sum_{i=1}^{N} v_i^2 - \sum_{i=1}^{N} |u_i v_i|} \quad (5)$$

With u_i and v_i are respectively the i^{th} element of u and v.

The Jaccard index is normalized (between 0 and 1). Plus it is close to 1 (or 100%), more the two individuals being compared are similar.

The Jaccard distance between two vectors u and v is obtained as follow:

$$d_{Jaccard}(u,v) = 1 - S_{Jac} = \frac{\sum_{i=1}^{N} (u_i - v_i)^2}{\sum_{i=1}^{N} u_i^2 + \sum_{i=1}^{N} v_i^2 - \sum_{i=1}^{N} |u_i v_i|} \quad (6)$$

More $d_{Jaccard}$ is close to 1 (or 100%), more the two individuals being compared are different.

2.6 Mahalanobis Cosine distance

Mahalanobis distance is introduced by the author P. C. Mahalanobis in 1936, it is a descriptive statistics based on the correlation between variables by which various data can be identified and analyzed. It differs from Euclidean distance, because it takes into account the correlations of the data set . Morover it is scale-invariant. In the area of face recognition Mahalanobis distance is used to calculate the distance between two vectors which presents projections of training images and projections of testing images, the scores of the matches are included between -1.0 and 1.0, with 1.0 being a perfect score.

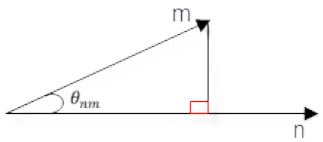


***Fig. 2**. The two vectors m and n in the Mahalanobis distance space*

Mahalanobis Cosine is the cosine of the angle θ between the images after they have been projected in the space of Mahalanobis and standardized by the estimation of the variance.

$$d_{MahCosine}(u,v) = -\frac{m.n}{|m||n|} \quad (7)$$

With m and n are two vectors of Mahalinobis space corresponding to u and v . The relation between these vectors is defined as follows:

$$m_i = \frac{u_i}{\sigma_i}, \qquad n_i = \frac{v_i}{\sigma_i}$$

With σ_i is the variance along i^{th} dimension.

3 Proposed technique

The Jaccard distance is known for its speed at the level of calculating such as Euclidean distance and it is powerful for the majority of the algorithm of recognition of faces; in addition, it gives very high rates of recognition compared to the other distances, but the

improper major of Jaccard is that it is sensitive to the confounding factors such as pose, illumination, expression, hair, glasses, or background. On the other hand, Mahalanobis Cosine distance gives promising results compared to the Jaccard distance when conditions are not controlled but Mahalanobis Cosine distance gives bad results when the data is very high.

In real-world pattern recognition problems, the data are generally very large and a facial image is often contaminated by one or more of confounding factors . In this case, we focus on the aspect combinations of measurements rather than on the approaches direct, this combination aims at overcoming the problems of each distance and to keep their points of force.

The block diagram of our idea is shown in Fig.1. The proposed system includes different components. First is to separate the data set into training images and testing images ; Second is the extracting features from the training templates and testing templates by PCA ,ICA,...etc; The last step is classification using equation (8).

The main novelty of this paper is that it compares features of training templates and testing templates in classification stage based on two distance Jaccard and Mahalanobis Cosine distance. Firstly the Jaccard dissimilarity $d_{Jaccard}(u,v)$ calculate the distance between two vectors u and v extracted from training and testing templates respectively. Secondly we use Mahalanobis Cosine distance $d_{MahCosine}(u,v)$ to calculate the dissimilarity of these two vectors u and v.

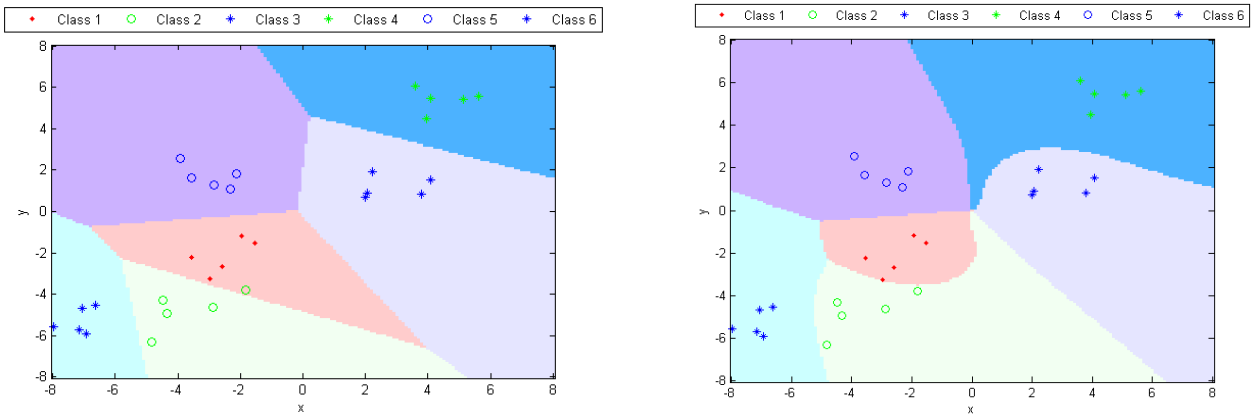
Finaly the result obtened by jaccard and Mahalanobis Cosine distance are combined using the following formula:

$$d_{JMahCosine}(u,v) = d_{Jaccard}(u,v) + e^{d_{MahCosine}(u,v)} \quad (8)$$

The exponential in the equation (8) is used to synchronize the values obtained by Mahalanobis Cosine with the values obtained by the Jaccard distance.

The combination of these two different distance allows us to exploit different sources of information presented by the values. The robustness of Jaccard is obtained by combining the properties of cosine distance and Euclidean distance. The cosine distance measures the similarity between two vectors based only on direction, and ignoring the impact of the distance between them while the Euclidean distance considers only the impact of the distance between the vectors, regardless of the direction of the vectors. Mahalanobis Cosine distance measures the degree of overlap between two vectors based on the correlation between them. The proposed method takes into account three parameters very important the direction, the impact between the vector and the degree of overlap which make it robust and adaptive to different use.

The Fig.3. plot the classifier decision boundaries of six classes using Euclidean distance and proposed approach. In this example we used Gaussian-distributed points with five training samples for each class.



(a) **(b)**

***Fig.3**. Classification by **(a)** The Euclidean distance and **(b)** The Proposed method*

Comparing result obtained by the Euclidean distance in Fig.3.(a) with the proposed method in Fig.3.(b), we see that the new method adopted by this article performs very well for all classes. However the Euclidean distance is lacking some precision (Classe 2).

4 Databases

Three databases of reference are used to carry out the tests. The first database is ORL (Olivetti Research Laboratory) [18]. It contains 400 images of 40 individuals, for each person, we have 10 images of size 112×92 pixels. For certain individuals, the images were captured at different times. The facial expressions and facial appearances vary too. The ten images of the first five people of ORL database are presented in Fig.4.

The second database is the database Yale [19]. This database contains 165 grayscale images representing the faces of 15 individuals, with 11 images / person of size 243 × 320 pixels. The images contain different facial expressions and conditions illumination for each individual. Eleven images of five people from the YALE database are presented in Fig.5.

The third database is the database Sheffield (formerly UMIST) [20] this database consists of 564 images in grayscale from 20 people of different races, gender and appearance. The size of each image is about 220 x 220 pixels. Each individual is represented in a variety of poses from the profile to the frontal views. Fifty-four images of a person of the Sheffield database are presented in Fig.6.

5 Discussions and experimental results

To compare our approach against existing distances we chose four widely known distances in the literature (Euclidean, City Block, Czekanowski, Hellinger), Thus the two distances that we have combined (Mahalanobis Cosine and Jaccard). All distances have been used in a facial recognition system. This system uses three database ORL, YALE and Sheffield as the basis for the tests.

These databases are used without any pretreatment and no standardization. To measure the robustness of the proposed approach , we used a well-known in the field of face recognition algorithm namely PCA.

The number of training images used for each class is variable (n = 1, 2,...) to analyze the impact of each distance on the recognition results, the training images are chosen randomly and the rest images to do the test. In order to have a fair comparison, the proposed approach and all distances use the same images of training and the same images of test to each faith. All results are summarized in tables 2, 3 and 4. In addition, a visual comparison of different methods presented in Fig.7.

Fig. 4*. Five samples of database ORL, each person has 10 images*

Fig. 5*. Five samples of YALE database, each person has 11 images*

Fig. 6*. A sample of Sheffield database, the person has 54 images*

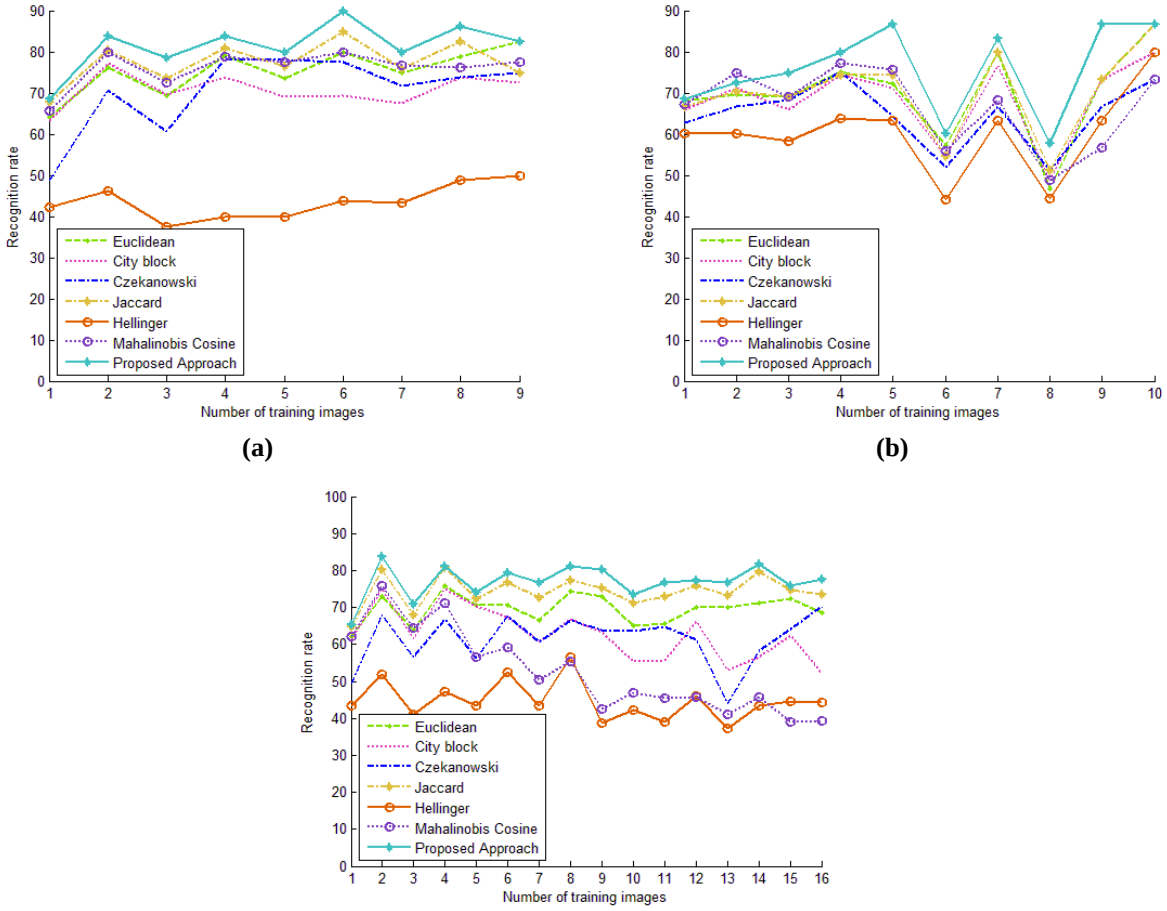


Fig. 7. *Graphical representation of* ***(a)*** *Table 2 (ORL database)* ***(b)*** *Table 3 (YALE database) and* ***(c)*** *Table 4 (Sheffield database)*

Table 2. Recognition results according to the number of training images for ORL database

Methods		Number of training samples per class								
		1	2	3	4	5	6	7	8	9
KNN+	Euclidean	64,17	76,25	69,29	79,17	73,50	80,00	75,00	78,75	82,50
	City Block	63,33	77,19	69,64	73,75	69,00	69,38	67,50	73,75	72,50
	Czekanowski	48,89	70,63	60,71	78,33	78,00	77,50	71,67	73,75	75,00
	Jaccard	67,78	80,31	73,57	80,83	76,50	85,00	75,83	82,50	75,00
	Hellinger	42,22	46,25	37,50	40,00	40,00	43,75	43,33	48,75	50,00
	Mahalanobis Cosine	65,56	80,00	72,50	78,75	77,50	80,00	76,67	76,25	77,50
	Proposed Approach	68,61	83,75	78,57	83,75	80,00	90,00	80,00	86,25	82,50

Table 3. Recognition results according to the number of training images for YALE database

Methods		Number of training samples per class									
		1	2	3	4	5	6	7	8	9	10
KNN+	Euclidean	68,00	69,63	69,17	75,24	72,22	57,33	80,00	46,67	73,33	86,67
	City Block	66,00	71,11	65,83	74,29	71,11	54,67	76,67	48,89	73,33	80,00
	Czekanowski	62,67	66,67	68,33	75,24	64,44	52,00	66,67	51,11	66,67	73,33
	Jaccard	66,67	70,37	69,17	74,29	74,44	54,67	80,00	51,11	73,33	86,67
	Hellinger	60,00	60,00	58,33	63,81	63,33	44,00	63,33	44,44	63,33	80,00
	Mahalanobis Cosine	67,33	74,81	69,17	77,14	75,56	56,00	68,33	48,89	56,67	73,33
	Proposed Approach	68,67	72,59	75,00	80,00	86,67	60,00	83,33	57,78	86,67	86,67

Table 4. Recognition results according to the number of training images for Sheffield database

Methods		Number of training samples per class											
		1	2	3	4	5	6	7	8	9	10	11	12
KNN+	Euclidean	62,10	73,05	63,76	75,97	70,50	70,63	66,63	74,41	73,08	65,15	65,78	69,95
	City Block	61,29	75,10	61,55	74,89	70,18	67,49	60,55	66,67	63,22	55,54	55,68	66,19
	Czekanowski	49,60	68,11	56,62	66,95	56,25	67,60	60,78	66,43	63,82	63,67	64,77	61,4
	Jaccard	64,82	80,35	68,07	80,90	72,26	76,68	72,59	77,46	75,36	71,06	72,98	75,91
	Hellinger	43,55	51,85	41,18	47,10	43,53	52,58	43,35	56,46	38,70	42,24	39,02	45,98
	Mahalanobis Cosine	62,00	75,82	64,39	71,24	56,69	59,08	50,46	55,40	42,67	46,80	45,33	45,85
	Proposed Approach	65,22	83,74	70,90	81,22	74,23	79,26	76,83	81,22	80,29	73,40	76,64	77,2

First of all, we note from the three tables 2, 3 and 4 that the rate of recognition of PCA using the proposed method is better compared to other distances in three database. In the ORL database the rate of recognition of PCA based on the proposed method is reached 90.00% while the best rate of recognition based on existing work does not exceed 85%. In the Yale database the recognition rate of PCA with the proposed method improves the rate of 68.67% for a training image to 86.67% with ten training image. Regard to Sheffield database the recognition rate of PCA using the proposed approach rises from 65.22% with an image for the training to 83,74% with two image for the training.

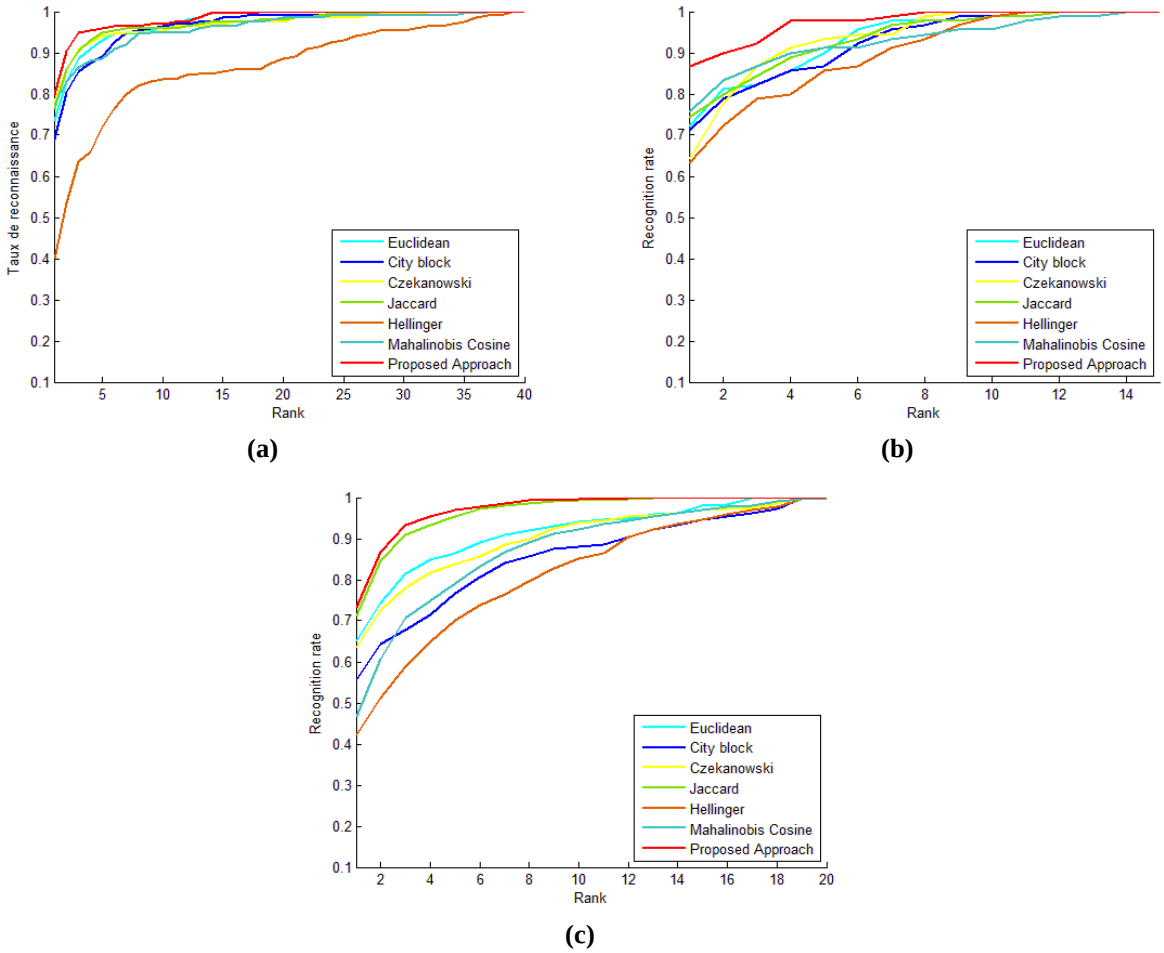


Fig. 8. *CMC curves using various metric on database* ***(a)*** *ORL* ***(b)*** *YALE and* ***(c)*** *Sheffield*

These results are confirmed by the Cumulative Match Characteristic curve (CMC). The CMC curve is used to measure the performance of an identification system that uses an ordered list of candidates. This curve gives the probability that the correct class of the example of test is presented in a row of the list. It is said that a system recognizes the rank 1 when it chooses the closest image as a result of the recognition. We say that a system recognizes the rank 2, when choosing among two images that best matches the input image, etc. CMC curves correspond to our approach and different distances in the database (ORL, YALE and Sheffield) are presented in Fig.8.

According to Fig.8 we find that the CMC curve correspond to the proposed approach in all databases is fast compared to the other CMC curves. All of these results brings us to the conclusion that our approach is the best compared to the other distances in this study.

6 Conclusion

Face recognition is still an active area of research. We have proposed a new technique for the face recognition based on the combination of two distances Jaccard and Mahalanobis Cosine. This combination allows us to overcome the disadvantages of these two distances.

We tested our method on three database ORL, YALE and Sheffield. The comparative study has shown the interest of the new technique with respect to some distance, such as Euclidean distance and Czekanowski distance.

References

[1] Zhao, W., Chellappa, R., Phillips, P. J., & Rosenfeld, A. (2003). Face recognition: A literature survey. Acm Computing Surveys (CSUR), 35(4), 399-458.

[2] Scholkopft, B., & Mullert, K. R. (1999). Fisher discriminant analysis with kernels. Neural networks for signal processing IX.

[3] S. Mika, G. Rätsch, J. Weston, B. Schölkopf, A. J. Smola, and K.-R. Müller. (2000).Invariant feature extraction and classification in feature spaces. In S. A. Solla, T. K. Leen, and K.-R. Müller, editors, Advances in Neural Information Processing Systems, volume 12, pp. 526-532. MIT Press.

[4] Schölkopf, B., Smola, A., & Müller, K. R. (1998). Nonlinear component analysis as a kernel eigenvalue problem. Neural computation, 10(5), 1299-1319.

[5] Ahonen, T., Hadid, A., & Pietikäinen, M. (2004). Face recognition with local binary patterns. In Computer vision-eccv 2004 (pp. 469-481). Springer Berlin Heidelberg.

[6] Gottumukkal, R., & Asari, V. K. (2004). An improved face recognition technique based on modular PCA approach. Pattern Recognition Letters, 25(4), 429-436.

[7] Lawrence, S., Giles, C. L., Tsoi, A. C., & Back, A. D. (1997). Face recognition: A convolutional neural-network approach. Neural Networks, IEEE Transactions on, 8(1), 98-113.

[8] Nefian, A. V., & Hayes III, M. H. (1998). Hidden markov models for face recognition. choice, 1, 6.

[9] Guo, G., Li, S. Z., & Chan, K. L. (2000). Face recognition by support vector machines. In Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on (pp. 196-201). IEEE.

[10] Zhao, W., Krishnaswamy, A., Chellappa, R., Swets, D. L., & Weng, J. (1998). Discriminant analysis of principal components for face recognition. In Face Recognition (pp. 73-85). Springer Berlin Heidelberg.

[11] Cha, S. H. (2007). Comprehensive survey on distance/similarity measures between probability density functions. City, 1(2), 1.

[12] Miller, P., & Lyle, J. (2008). The effect of distance measures on the recognition rates of PCA and LDA based facial recognition. Digital Image Processing.

[13] Turk, M., & Pentland, A. (1991). Eigenfaces for recognition. Journal of cognitive neuroscience, 3(1), 71-86.

[14] Jolliffe, I. (2005). Principal component analysis. John Wiley & Sons, Ltd.

[15] Beveridge, R., Bolme, D., Teixeira, M., & Draper, B. (2003). The CSU face identification evaluation system user's guide: version 5.0. Computer Science Department, Colorado State University, 2(3).

[16] Real, R., & Vargas, J. M. (1996). The probabilistic basis of Jaccard's index of similarity. Systematic biology, 380-385.

[17] Tanimoto, T. (1957). Internal report: Ibm technical report series. Armonk, NY: IBM.

[18] The ORL Database of Faces, Available: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html

[19] The YALE database of Faces, Available :http://vision.ucsd.edu/content/yale-face-database

[20] The University of Sheffield. The umist face database, available: http://www.sheffield.ac.uk/eee/research/iel/research/face. 1998.