Metrics Performance Evaluation: Application to Face Recognition

Naser Zaeri, Abeer AlSadeq, and Abdallah Cherri

Electrical Engineering Dept., Kuwait University, P.O. Box 5969, Safat 13060, Kuwait
{zaery, abeer, cherri}@eng.kuniv.edu.kw

***Abstract* – Countless number of applications varying from music, document classification, image and video retrieval, require measuring similarity between the query and the corresponding class. To achieve this, features, that belong to these objects are extracted and modified to produce an *N*-dimensional feature vector. A database containing these feature vectors is constructed, allowing query vectors to be applied and the distance between these vectors and those stored in the database to be calculated. As such, the careful choice of suitable proximity measures is a crucial success factor in pattern classification. The evaluation presented in this paper aims at showing the best distance measure that can be used in visual retrieval and more specifically in the field of face recognition. There exist a number of commonly used distance or similarity measures, where we have tested and implemented eight of these metrics. These eight metrics are famous in the field of pattern recognition and are recommended by the Moving Picture Experts Group (MPEG). More than 300 tests on 300 different databases were performed to consolidate our conclusion. The evaluation shows that the Euclidean and Minkowski distance measures are the best. On the other hand, the Canberra distance measure gives the worst results.**

***Index Terms* – Distance measure, Pattern classification, Face recognition.**

I. INTRODUCTION

Pattern recognition is the art of matching a feature vector of an object to a database of vectors. Interest in the area of pattern recognition has been renewed recently due to emerging many interesting and important applications. These applications include data mining (identifying a pattern, e.g., correlation, or an outlier in millions of multidimensional patterns), document classification (efficiently searching text documents), financial forecasting, organization and retrieval of multimedia databases, and biometrics (personal identification based on various physical attributes such as face and fingerprints) [1].

Human ability to recognize a face is tremendous. Human brain can identify a familiar object at a glance, despite internal changes or external environmental factors. On the other hand, the Computer – based face recognition is a very challenging dilemma, the recognition system has to be flexible with the external changes like; environmental light, person's orientations, distance from camera, and internal deformations (facial expression, aging and makeup).

One of the important fields of pattern recognition is the face recognition. Face recognition finds many important applications in many life sectors and in particular in commercial and law enforcement applications. In general, the process of recognizing a face passes through many steps such as segmentation of faces from cluttered scenes, extraction of face features, identification, and matching. Facial features extraction to find the most appropriate representation of face images is one of the most important components of a face recognition system for identification purposes. The Principal Component Analysis (eigenface) technique is one of the well known approaches to achieve a good representation of a face image [2, 3].

In addition, the matching procedure and technique has a great influence on the final result of the recognition system. Various similarity measures were proposed for the matching stage [3, 6]. In the different pattern recognition and classification areas, each one of these similarity (or dissimilarity) measures has its advantages, disadvantages, and applications.

In this paper, we study the most famous and important metrics and distance measures that are used in pattern recognition, where we focus on the face recognition field. We study eight of these metrics that are proposed and recommended by the MPEG [4]. We apply these eight metrics on 300 different databases. The databases are constructed and taken from the Olivetti Research Ltd. (ORL) database. More details are given in the following sections.

This work is organized as follows. A brief description of Principal Component Analysis (PCA) is given in Section 2. The study of the different metrics, their formulation, and characteristics is presented and discussed in Section 3. In Section 4, results of testing and implementing the study on the ORL database are presented. Conclusions are given in Section 5.

II. PRINCIPAL COMPONENT ANALYSIS

The PCA technique, proposed by Turk and Pentland [2], extracts the relevant information in a face image, encodes it as efficiently as possible, captures the variation in a collection of face images, and compares one face encoding with a database of models encoded similarly. The images of faces, being similar in overall configuration, will not be randomly distributed in the huge image space and, consequently, they can be described by a relatively low dimensional subspace.

This process is achieved by finding the principal components of the distribution of faces, or the eigenvectors of the covariance matrix of the set of face images. The eigenvectors are ordered, each one accounting for a different amount of the variation among the face images. These eigenvectors can be thought of as a set of features that together characterize the variation between face images. Each image contributes more or less to each eigenvector, so that the eigenvector is displayed as a sort of ghostly face which is called an eigenface. Example of some of the images used from the ORL database and some of the resultant eigenfaces using the PCA is shown in Figs. 2 and 3, respectively.

Fig. 1. Example of some of the images used from the ORL.

Fig. 2. Some of the eigenvectors using conventional PCA.

In brief, if the training set of face images is $\Gamma_1, \Gamma_2, \Gamma_3, ..., \Gamma_{M'}$, then, the average face of the set is defined by

$$\Psi = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \qquad (1)$$

Each face differs from the average by the vector $\Phi_i = \Gamma_i - \Psi$. This set of vectors is subject to principal component analysis. These components are the eigenfaces of the covariance matrix

$$C = A^T A \qquad (2)$$

where the matrix $A = [\Phi_1 \ \Phi_2 \ ....\Phi_M]$. In practice, a *M'* eigenvectors smaller than *M* is sufficient for identification, since accurate reconstruction of the image is not a requirement. In this framework, identification becomes a pattern recognition task. The eigenfaces span an *M'* -dimensional subspace of the original image space N^2 . The *M'* significant eigenvectors are chosen as those with the largest associated eigenvalues.

Now, a face image (Γ) is transformed into its eigenface components by the following equation,

$$\omega_k = u_k^T(\Gamma - \Psi) \qquad (3)$$

for *k = 1, ..., M'*., where u_k is the k^{th} eigenvector of the covariance matrix. The weights form a vector $\Omega^T = [\omega_1, \omega_2, ..., \omega_{M'}]$ that describes the contribution of each eigenface in representing the input face image, treating the eigenfaces as a basis set for face images.

To determine which face class provides the best description of an input face image, we find the face class *k* that minimizes a certain distance

$$\varepsilon_k = Dis\{\Omega, \Omega_k\} \qquad (4)$$

where Ω_k is a vector describing the k^{th} of a face class. These classes are calculated by averaging the results of the eigenvector representation over a small number of face description vectors (as few as one) of each individual. A face is classified as belonging to class *k* if the corresponding ε_k is the minimum among all other ε_k 's.

III. DISTANCE MEASURES

The main task of good distance measures is to *reorganise* descriptor space in a way that media objects with the highest similarity are nearest to the query object. If distance is defined minimal, the query object is always in the origin of distance space and similar candidates should form clusters around the origin that are as large as possible.

The distance between two points is either greater than zero or equal to zero. It is greater than zero if the points are distinct; and it is equal to zero only when the two points are the same; in other words, when they are not distinct. We will make use of this property of distance to build our recognition system where the degree of similarity and dissimilarity between the tested image and the database is measured.

A. Euclidean Distance

The Euclidean distance between two vectors [5] takes the form

$$D_p(\vec{x}, \vec{y}) = \sqrt{\sum_{i=1..k}(x_i - y_i)^2} \qquad (5)$$

where x_i and y_i ; represent the coordinates of the two vectors X and Y in the plane. The Euclidean distance, so-called 2-norm distance, has been applied in many life and scientific fields due to its simplicity to be applied, and tends to obtain a convenient and flexible closed form solution in many complicated situations, while providing better accuracy under certain conditions.

B. Minkowski Geometry

Minkowski distance [5, 7] is a very popular method to measure image similarity. It is based on an assumption: the similar objects should be close to a query object in all dimensions. It consists of: the features of a scaled image that are of the same importance as the features of a cropped image. Suppose two images X and Y are represented by two k dimensional vector, $X=(x_1, x_2\dots x_p)$ and $Y=(y_1, y_2\dots y_p)$, respectively, the weighted Minkowski metric is defined as

$$D_p(\vec{x},\vec{y}) = (\sum_{i=1..k} w_i |x_i - y_i|^p)^{(1/p)} \quad (6)$$

Where p and k represent the Minkowski factor and number of dimensions (attributes) respectively, p need not be an integer, but it cannot be less than 1. w_i is the weighting to identify important features. If w_i is normalized then the above method will be more applicable and gives great results simultaneously. The Minkowski distance then is defined as

$$D_p(\vec{x},\vec{y}) = (\sum_{i=1..k} |x_i - y_i|^p)^{(1/p)} \quad (7)$$

Different values of the parameter p give us different distance measurements. If the p is set to be 2, we will get the Euclidean distance. The various forms of the Minkowski distance do not account for different metrics of the individual coordinates. If the coordinates span different ranges, the coordinate with the largest range will dominate the results. Therefore scale the data before calculating the distances is needed.

C. Canberra Distance

The Canberra distance metric [12] is used for similarity/dissimilarity comparison. It examines the sum of series of a fraction differences between coordinates of a pair of objects according to the following equation

$$D(\vec{x},\vec{y}) = \sum_{i=1}^{k} \frac{|xi - yi|}{|xi + yi|} \quad (8)$$

The Canberra distance is suitable for variables taking non-negative values and is sensitive to small changes. It completely ignores comparison when the two coordinates of concerned data are equal to zero. Due to robustness of Canberra technique, it is a highly recommended method in medical researches, especially in the cases of missing data.

D. Chebyshev Distance

In Chebyshev distance [5, 9] we take the maximum difference between the compared coordinates of two objects. Its mathematical representation is

$$D_p(\vec{x},\vec{y}) = \max\{|x_1 - y_1|, |x_2 - y_2|, \dots, |x_n - y_n|\} \quad (9)$$

The geometric analysis of Chebyshev distance is given for the functioning of discrete detectors with applications for two-dimensional event recognition problems. A big disadvantage of the Chebyshev method is that if one element in the vectors has a wider range than the other elements then that large range may *dilute* the distances of the small-range elements.

E. Correlation Coefficient Similarity

Correlation measures [7, 9] the extent of relationship between the two variables; where simple linear regression provides the prediction equation that quantifies such relationship at the best way. Correlation is measured by the correlation coefficient

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} \quad (10)$$

The correlation coefficient r has a value between -1 and +1, and equals zero if the two variables are not associated. It is positive if the two continuous X and Y variables have direct relation (i.e. they increase or decrease together). If r is negative, the relationship is indirect (i.e. if X values increase, Y values decrease). The larger the value of r, the stronger is the association. The maximum value of 1 occurs in case of perfect correlation.

F. Divergence Coefficient Clark (DCC)

Divergence coefficient Clark method [5] can be represented as

$$D(\vec{x},\vec{y}) = \sqrt{\sum_{i=1}^{k} \frac{(xi - yi)^2}{xi + yi}} \quad (11)$$

Due to the existence of square root formula, this method suffers from extreme sensitivity to negative values at the denominator. Numbers should be chosen carefully then.

G. Meehl Distance

The figurative meaning of Meehl Distance [11] can be shown through the following equation

$$D(\vec{x},\vec{y}) = \sum_{i=1}^{k} ((x_i - x_{i+1}) - (y_i - y_{i+1}))^{\wedge 2} \quad (12)$$

From this equation, we can see that the distance depends on the two consecutive points in the feature vector.

H. 2D Cosine Similarity

This similarity [8] does not depend on the coordinates of the points but it depends on the angle measured between them. It is known that two $\bar{x}$ and $\bar{y}$ arrays can meet in a single point to form an arbitrary angle α measured in degrees.

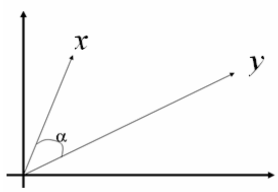

Fig. 3. Angle measurement between two vectors.

Suppose two images X and Y are represented by two k dimensional vector, $\overline{X} = (x_1, x_2, ..., x_k)$ and $\overline{Y} = (y_1, y_2, ..., y_k)$, respectively. The two vectors X and Y are similar when $\alpha = 0$, otherwise dissimilar. Cosine similarity or in other word inner product distance can be figured as:

$$sim(x, y) = \cos\alpha = \frac{x * y}{|x||y|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2}\sqrt{\sum_{i}^{n} y_i^2}} \tag{13}$$

The advantage of this method is that we can normalize the zero mean and the unit variance of the output by dividing each of its components by any related length, (e.g.: Euclidean). The major advantage of this method over the previous metric distances is that it measures similarity by the angle between the compared object and does not depend on the variance of their coordinates. This characteristic is very important especially in strings objects.

IV. EXPERIMENTS

Although we can apply the eight similarity/ dissimilarity measuring method illustrated before in a huge pattern recognition process, we are here interesting in applying this whole process to face recognition specifically, where our aim is to find the best metric that suits the face recognition.

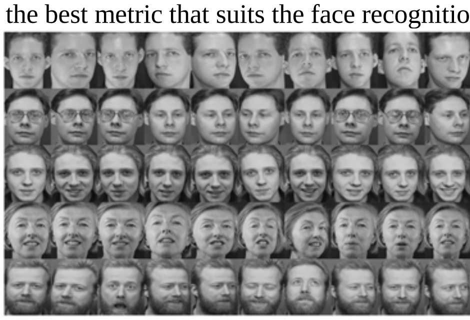

Fig. 4. Examples from the ORL database used in the evaluation of the different metrics.

We have applied our experiments on the ORL database. It consists of 40 different individuals with each individual represented by 10 different images. These images were taken randomly under different external environmental factors like the intensity of light, the back ground effect and with different pose orientations. Further, some internal factors were also considered: different facial features, facial hair, with or without glasses.

We have performed 300 experiments. In each one of these experiments, we take a part of the image for each one of the 400 images under test. This part is the same for all the images in the same experiments. In each one of these experiments, this part differs in location and size from other parts. As explained in Section II, each image is represented by a number of weights forming a vector $\Omega^T = [\omega_1, \omega_2, ..., \omega_{M'}]$ that describes the contribution of each eigenface in representing the input query image.

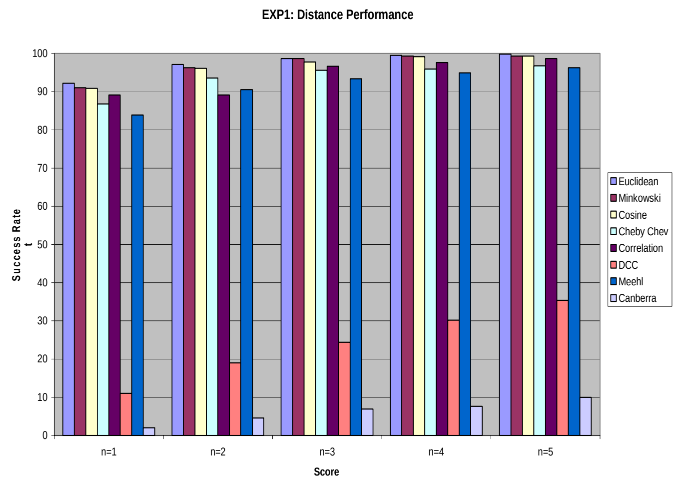

Fig. 5. Recognition rate for different values of cumulative matching score for the various distance measurements for experiment no. 1.

All the images, or the image *parts* are represented with vectors of size 10. To determine which face *part* class provides the best description of an input face *part* image, we find the face *part* class k that minimizes a certain distance , as explained in equation (4). Figures (5-7) show the results and the performance of each of the distance measures that were explained in the previous section. The figures shown are randomly selected for the 5 different experiments out of the 300 experiments that were performed. Note that the figures show the results as a function of the cumulative match score "*n*". Cumulative match score (CMS) is an evaluation methodology proposed by the developers of FERET In this case, identification is regarded as correct if the true object is in the top Rank *n* matches.

As the charts illustrates Euclidean measure tends to get the best detection results, however Minkowski with p=1 gives very close results to the Euclidean case. That proves that both give best results. As the power p of Minkowski increases, the performance degrades. The low performance of Chebyshev may be explained by the *dilution* of the small values of similarity and the domination of the largest value. The lowest performance was for Canberra distance. This is somewhat expected, since this distance measure suffers from an extreme sensitivity of negative values. Figure 8 summarizes the performance of distances after performing the 300 experiments.

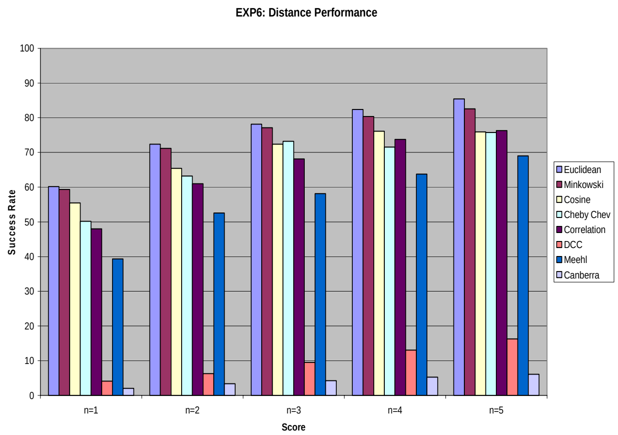

Fig. 6. Recognition rate for different values of cumulative matching score for the various distance measurements for experiment no. 6.

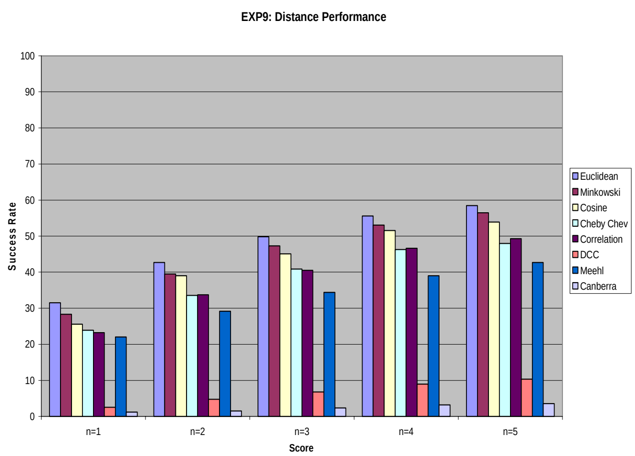

Fig. 7. Recognition rate for different values of cumulative matching score for the various distance measurements for experiment no. 9.

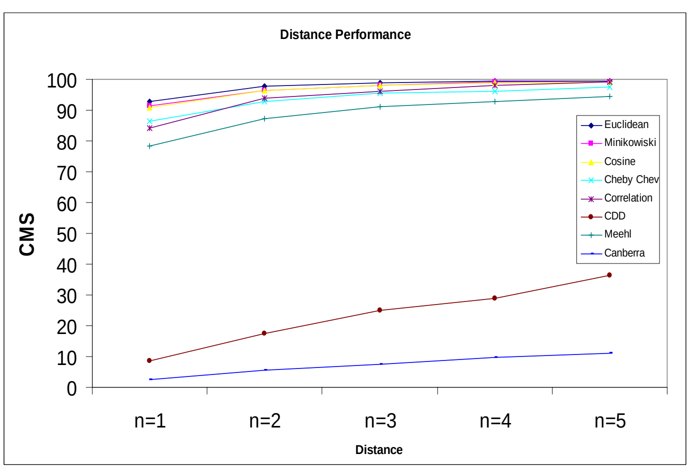

Fig. 8. Average Cumulative matching score for the various distance measurements for the 300 experiments performed.

V. CONCLUSION

The evaluation presented in this paper aims at testing the recommended distance measures and finding better ones for the basic visual descriptions, in particular for face recognition. Eight different distance measures were implemented, where 300 different databases were used all taken and derived from the main ORL database. The choice of the right distance function for similarity measurement depends on the pattern classification problem under study. The results show that the best performance was for the Euclidean distance and Minkowski, respectively. On the other hand, the Canberra gave the worst results.

REFERENCES

[1] R. Chellappa, C. L. Wilson, and N. Sirohey “Human and machine recognition of faces, a survey”, *Proc. IEEE* 83, 705-740 (1995).

[2] Turk and Pentland, “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, March 1991.

[3] B. Moghaddam and A. Pentland, “Probabilistic Matching for Face Recognition”, *IEEE*, 1998.

[4] Farzin Mokhtarian and Miroslaw Bober, *Curvature Scale Space: Theory, Applications, and MPEG-7 Standardization*, Kluwer Academic Publishers, Netherlands, 2003.

[5] Marvin Jay Greenberg. "Euclidean and Non-Euclidean Geometries, Development and History". Third Edition .W.H Freeman and Company, 1999.

[6] H. Kong, X. Li, J.-G. Wang, E.K. Teoh, C. Kambhamettu, “Discriminant Low-dimensional Subspace Analysis for Face Recognition with Small Number of Training Samples”, *British Machine Vision Conference (BMVC)*, Oxford, UK, Sept. 5-9, 2005.

[7] Eidenberger, H., and Breiteneder, C. Visual similarity measurement with the Feature Contrast Model. In Proceedings SPIE Storage and Retrieval for Media Databases Conference (Santa Clara CA, January 2003), SPIE Vol. 5021, 64-76.

[8] Gower, J.G. Multivariate analysis and multidimensional geometry. The Statistician, 17 (1967),13-25.

[9] Santini, S., and Jain, R. Similarity measures. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21/9, 871-883, 1999.

[10] Cohen, J. A profile similarity coefficient invariant over variable reflection. Psychological Bulletin, 71 (1969), 281-284.

[11] Meehl, P. E. The problem is epistemology, not statistics: Replace significance tests by confidence intervals and quantify accuracy of risky numerical predictions. In Harlow, L.L., Mulaik, S.A., and Steiger, J.H. (Eds.). What if there were no significance tests? Erlbaum, Mahwah NJ, 393-425.

[12] Lance, G.N., and Williams, W.T. Mixed data classificatory programs. Agglomerative Systems Australian Comp. Journal, 9 (1967), 373-380.