On Recognizing Human Faces from Frontal Views

Serdar YILMAZ, Metin ARTIKLAR
KSU, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Kahramanmaraş

ABSTRACT: This paper consists of two main parts. In the first one, we compared $L_{0.5}$, L_1, and L_2 metrics and showed which one is better suited for the area of face recognition for 2-dimensional gray level face images. In the second part, we have demonstrated the effectiveness of different facial expressions for the purpose of identification of individuals. We have shown that L_1 norm performs as good as other two metrics but takes almost as half time to process images. We have also compared recognition performance of four different expressions on two test sets and showed that blank and angry expressions better represent a face Simulation results are reported on a dataset of 600 people selected from CNNL face database.
Key Words: Face recognition; template matching; Similarity measures; Facial expressions.

Profilden İnsan Yüzü Tanıma Üzerine

ÖZET: Yayınımız iki ana bölümden oluşmaktadır. İlk bölümde resim işleme alanının bir alt dalı olan insan yüzü tanıma konusunda $L_{0.5}$, L_1, ve L_2 normlarından hangisinin daha uygun olduğunu araştırdık. İkinci bölümde ise kişilerin birden fazla resimlerinin bulunduğu insan yüzü veri tabanlarında hangi yüz ifadesinin kişileri tanıma amacı için en etkin olduğunu araştırdık. Çalışmalarımız sonucunda L_1 normunun diğer iki norma çok yakın sonuçlar verdiğini fakat onlara oranla yarıya yakın bir sürede resimleri işlediğini bulduk. Buna ek olarak, dört farklı yüz ifadesinden hangilerinin yüzü en iyi temsil ettiğini bulmak amacıyla iki test set üzerinde yaptığımız deneyler sonucunda normal ve kızgın yüz ifadelerinin bu amaç için en uygun olduğunu bulduk. Araştırmalarımız için CNNL insan yüzü veri tabanından 600 kişilik bir veri seti kullandık
Anahtar Kelimeler: İnsan yüzü tanıma, şablon eşleştirme, benzerlik ölçüleri, insan yüzündeki ifadeler

INTRODUCTION

Identification of human faces by computers has been studied extensively over last 25 years. Numerous algorithms have been developed including Eigenfaces (Sirovich and Kirby, 1987; Pentland and Turk, 1991), elastic graph matching (Lades et al., 1993) and other statistical and neural network based methods (Lawrance, 1997; Samaria, 1994). Among them, Eigenfaces and elastic graph matching received the most attention due to their high recognition performance. Although many of these algorithms focused on recognition of faces from frontal views, they, in most cases, lacked reporting performance results on a sufficiently large face database. This was partly due to absence of standard face databases other than the FERET (Philips et al.., 1998). However, when emphasis needs to be given on frontal facial views alone, even the FERET database can only provide images in such orientation up to 250 individuals. This still is not a sufficiently high number for a typical industrial application of face recognition system (a typical small size company would generally have 250 and up number of employees in today's world). Recently, Artiklar et. al. (2003) reported an algorithm based on 2 levels processing in order to address this problem. They have tested their system on the CNNL face database collected at Wayne State University computation and Neural Network Laboratory that includes more than 1200 individuals each having 10 different expressions. They reported over 97% classification performance on a memory set of 1000 individuals.

DATABASE

We have conducted all of our simulations on CNNL face database. A brief explanation of the database is given next. Detailed discussion of the construction of the database can be found in Artiklar, (2002).

CNNL face database consists of over 1200 individuals between ages of 15 to 60 years old. Subjects come from different ethnic backgrounds such as Caucasian-American, African-American, Asian, Indian, Middle Eastern, etc. Out of 1200 individuals, 733 are men and 467 are women. Figure 1 shows the distribution of ethnic backgrounds over 1200 individuals in our database.

The images were collected in a laboratory setting so as to minimize the amount of pre-processing that is necessary in order to reduce the effect of tilt rotation, translation, scaling, etc.

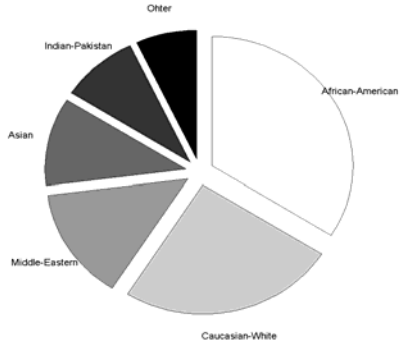


Figure 1. Distribution of ethnic-backgrounds of 1200 individuals in the CNNL database.

Each individual is represented by 8 expressions in the database: Blank, smile angry, surprise, blank1, blank2, blank3, and blank4. Two additional expressions, blank and arbitrary, are obtained separately to generate additional test sets. For the arbitrary test expression, the subject is asked to give an expression that makes it difficult for computer to recognize the person. Both database and test sets are snapped as gray scale images at dimension of 115 x 82. Figure 2 shows these images of an individual from our database.



Figure 2. Sample images of an individual from CNNL database showing all 8 expressions and additional two test images.

ALGORITHM

Our goal in this paper is not to develop a new face recognition algorithm but to study characteristics of human face through different facial expressions. Hence, we'll use well known nearest neighbor algorithm in classification tests. One thing to be determined, though, is as to choose which metric needs to be used when implementing the nearest neighbor. L_2 norm which also known as Euclidian distance is the one that most popularly used among image processing community because of its convenient analytical properties. However, several researchers reported that beside the benefits, Euclidian distance tends to overemphasize the extremes in the image and suggested the use of L_1 norm instead.

In the remainder of this paper we'll first investigate the most appropriate distance measure among some of the well-known metrics for the face recognition applications and then study the effect of different facial expressions on the recognition performance of a face identification system.

FORMAT OF THE EXPERIMENTS

We have conducted extensive simulations on CNNL database for our study. All images are retained in their original form and no dimensionality reduction or preprocessing is applied prior to application of algorithms. (Note that images in our database were already taken under controlled environment to eliminate such preprocessing). The format of reporting simulation results is described below.

We have partitioned database into sets of 200, 400, and 600 individuals that are selected randomly from the database to form memory sets. Depending on the experiment, blank, smile, angry, or surprise expression of an individual is included in the memory. The blank and arbitrary expressions of those individuals in the test image dataset are used to form test sets.

When reporting an experimental result, we have repeated the experiment 5 times (where in each case images in the memory set are selected randomly) and averaged the results. Performance of algorithms is always measured based on the best matching memory image to an input test image.

EUCLIDIAN DISTANCE VS. CITY BLOCK

In this section, we will investigate on 3 different metrics that can be used in face recognition systems: Euclidian distance (L_2 norm), City Block distance (L_1 norm) and $L_{0.5}$ norm. Note that these metrics are by no means the only ones used in image processing community, yet the first two are quite popular among the researchers. The reason for adding the $L_{0.5}$ norm into our list is because it has been reported that for two-dimensional images $L_{0.5}$ has outperformed the previous two with a clear margin by Sim at all (2000).

All 3 distance measures mentioned above are known as L_p similarity measures in the literature and are defined as follows.

Let image $\mathbf{I}^i$ represent a 2 dimensional matrix of size n x m where $\mathbf{I}^i \in \mathbf{Z}^{n\times m}$ for gray level images and superscript i refers to index of the image in the database. Let also $I(x,y)$ denote the gray level pixel value of an image at row x and column y. Then, the distance between two images $\mathbf{I}^1$ and $\mathbf{I}^2$ is defined in terms of L_p similarity measure as:

$$L_p(\mathbf{I}^1 - \mathbf{I}^2) = \left(\sum_{x=1}^{n} \sum_{y=1}^{m} \left| I^1(x,y) - I^2(x,y) \right|^p \right)^{\frac{1}{p}}$$

Below, we demonstrate simulation results in order to compare the effect of these three metrics on recognition performance of our face recognition system. For the simulations, we have included only one image per person in the memory set which was the blank expression of subjects. This memory set is then tested against blank and arbitrary test expressions as to measure the classification performance of the system.

Figure 3 demonstrates the performance of the recognition algorithm while 200, 400 and 600 people were included in the memory and being tested against blank test expressions. From the figure, we could see that for the case of 200, Euclidian distance measure has performed 98.4% accuracy which is slightly better than the other two, whereas for 400 case, all 3 metrics seem to be performing around 97%. For the case of 600 individuals, though, $L_{0.5\ metric}$ performs little better than the other two with 96.9%.

Figure 4 demonstrates the performance of a similar case against arbitrary test images. Here, we see that when 200 subjects are included in the memory, $L_{0.5}$ performs 92.4%, which is better than the other two with

around 1% margin. For the case of 400, all three perform equally well and for 600 memory case, City Block distance slightly outperforms the other two with 88.5%.

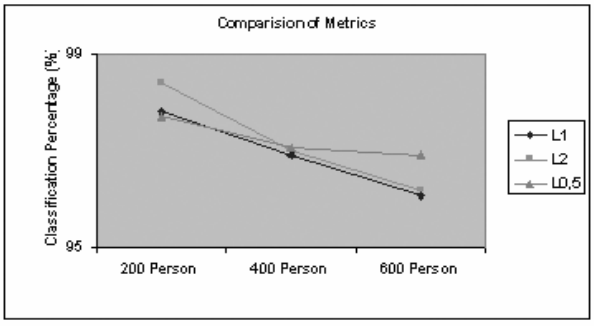


Figure 3. Comparison of metrics using blank test expression.

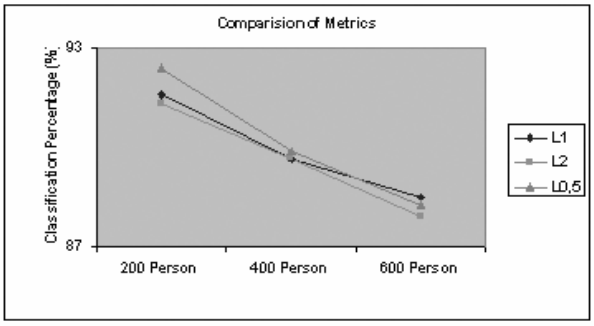


Figure 4. Comparison of metrics using arbitrary test expressions.

The next figure compares the time that it takes for each metric to process a test image. As it can be seen from the Figure 5, the processing time of identifying a test image against a database of 200 individuals for the case of City Block metric is 0.22 sec. (These simulations are run using a PC with AMD XP 2.1 GHz processor and 512 Mbytes DDR 333 MHz RAM). This value is 0.37 sec for the remaining two metrics. For databases of 400 and 600 individuals, we again see a similar trend.

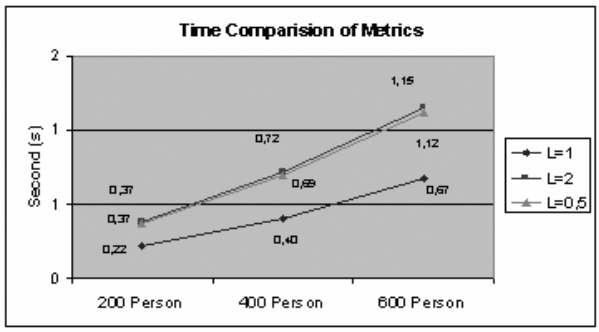


Figure 5. Comparison between processing time of an input image for $L_{0.5}$, L_1 and L_2 metrics.

The following comments can be made after an analysis of above figures: In terms of recognition performance, none of the 3 metrics clearly outperformed the other two in all cases. On the other hand, by looking at the time requirement, we could say that for applications that need faster retrieval time, City Block distance seems to be a better choice since it performs very close to others in most cases but requires a lot less time to process a test image. However, note that Euclidian distance has analytical properties that make it suitable for the cases where algorithm developments require extensive math. Due to this reason, this metric is quite popularly used.

For the remaining of this paper, our goal is to compare expressions in order to investigate which facial expression is the best suited to represent a human face. Since the choice of the metric used in distance calculations will not affect the outcome, we have decided to use Euclidian distance due to its analytical properties in the remaining simulations.

EFFECT of FACIAL EXPRESSIONS

In this section, we demonstrate which expression is the most useful in identifying humans. In order to answer this question, we included only one expression of each individual in the memory. These were blank, smile, angry and surprise expressions. We have then measured performance of the algorithm on blank and arbitrary test images. Simulation results for both cases are given below.

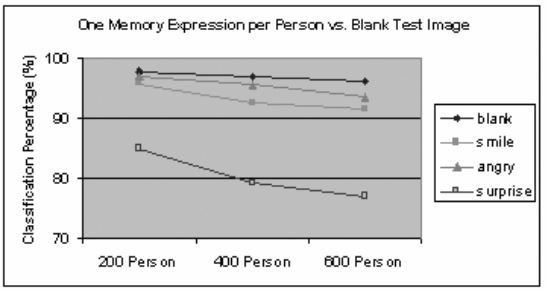
In the below simulations, the Figure 6 shows classification performance results for the cases of blank and arbitrary test expressions. For the blank test images case, the blank memory expression performs the best as compare to remaining three with 97.8% recognition rate when 200 individuals are used to form the memory set. Angry expression gives a performance of 96.9% and is the second best. The smile and surprise expressions come in the third and fourth places respectively.

For the case of arbitrary test images, we see a similar pattern. Blank expression performs the highest with 93.2%, which is followed by angry expression performing at 91.0%. The remaining two expressions of smile and surprise give recognition rates of 90.3% and 86.6% and come in the third and fourth places respectively.

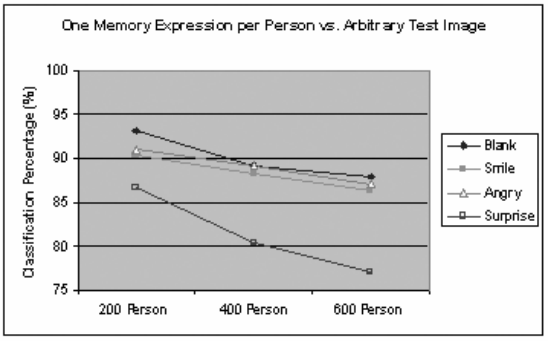
Note that during the collection of database, subjects are asked to give an unusual expression when snapshot is taken for the arbitrary test image. Hence, this expression is quite different from the other expressions. However, the fact that the blank expression is performing the best in both test image cases clearly indicates that this expression is the most suitable for identification of a human face.

The fact that angry expression comes as the second best is to be expected because subjects tend to exaggerate the smile expression when they are asked to pose since it is easier to imitate. This causes large amount of changes around mouth area of the face and in turn makes it little harder to recognize him/her. Angry poses, on the other hand, cause changes around

eyebrows/eye region of the face and are minor when compared to those in the smile case. It also worth mentioning that during the generation of the database, subjects generally commented on difficulty of posing angry expression when they are not in that mood naturally.



(a)



(b)

Figure 6. Recognition performance results for including one expression in the memory for the cases of (a) blank test images and (b) arbitrary test images.

CONCLUSION

In this paper, we have demonstrated through extensive simulation results that for face identification problems, L_2 and $L_{0.5}$ norms take almost same times to process an input image and their effect on the classification performance is also very close. Hence one does not yield a better distance computation than the other. The L_1 norm, on the other hand, needs almost as half of the time as compared to other two in order to calculate a distance between an input and a database image. The effect of the L_1 norm on the performance is a little worse than the other two in most cases but the difference is very small. This feature of L_1 norm would make it very appropriate for real time applications.

In the representation of human face, we have found that blank and angry expressions are the most effective. These two performed the highest when only one expression is used to describe a face.

In the future work, we will further investigate this topic by experimenting on multiple expressions per individual in the memory and by measuring recognition system performance versus image size.

REFERENCES

Artıklar M., Mu X., Hassoun M., Watta P. 2003. Local Voting Networks for Human Face Recognition. International Joint Conference on Neural Networks , , July 20-24, Portland.

Artıklar, M. 2002. Capacity Analysis of Voting Networks with Application to Human Face Recognition, Wayne State University College of Engineering, Ph.D. Thesis, 44s

Belhumeur P., Hespanha J., Kriegman, D. 1997. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 19(7): 711-720.

Lades M., Vorbruggen J. C., Buhmann, J., Lange J., Malsburg C.V.D., Wurtz R. P., Konen W. 1993. Distortion Invariant Object Recognition in Dynamic Link Architecture. IEEE Transactions on Computation, 42: 300-311.

Lawrance S., Giles C. L., Tsoi A. C., Back A. D. 1997. Face Recognition: A Convolutional Neural Network Approach. IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition, 8(1): 98-113.

Phillips P. J., Wechsler H., Huang J., Rauss P. J. 1998. The FERET Database and Evaluation Procedure for Face Recognition Algorithms. Image and Vision Computing, 16: 295-306.

Samaria F. S., Harter A. C. 1994. "Parameterization of a Stochastic Model for Human Face Identification", Proceedings of the 2nd IEEE Workshop on Applications of Computer Vision, Sarasota.

Sim T., Sukthankar R., Mullin M., Baluja S. 2000. Memory-based face recognition for visitor identification. Proceedings of 4th International Conference on Face and Gesture Recognition, Grenoble.

Sirovich L., Kirby M. 1987. Low Dimensional Procedure for Characterization of Human Faces. Journal of Optical Society of America, 4: 519-524.

Turk M., Pentland A. 1991. Eigenfaces for Recognition. Journal of. Cognitive Neuroscience, 3(1): 71-86.