RECOGNITION OF OCCLUDED OR DISTORTED FACES USING PARTIAL SIMILARITY

Frank James

University of South Florida, Utah State University

ABSTRACT

This paper presents a recognition system for identifying faces that are partially occluded or distorted by variations in facial expression. The algorithm employs a partial similarity measure, which is useful since it captures the most similar features between two face images. This technique models that which is inherent in human perception. Two methods are explored in setting the threshold needed to determine similar features: one based on the *golden section* rule and the other based on a *maximum margin* value. The proposed method proves to be effective in handling expression changes and partial occlusions.

***Index Terms*—** Face recognition, machine learning, nonmetric similarity, partial similarity, similarity measure.

1. INTRODUCTION

Due to its wide applications in information security, law enforcement and surveillance, smart cards, access control, and others, face recognition techniques have received significantly increased attention from both the academic and industrial communities during the past several decades [1]. However, face recognition in uncontrolled situations remains one of the most important bottlenecks for practical face recognition systems [2].

The purpose of this paper is to address a face recognition problem where facial features are altered by partial occlusions or changes in facial expression. Such variations in facial appearance are commonly encountered in uncontrolled situations and may cause big trouble to the face-recognition-based security system but are less studied in the literature [3]. It should be noted, however, that this paper does not address other common variations that occur in uncontrolled conditions, such as changes in lighting or the effects of aging. These alterations tend to affect facial appearances in a global manner, whereas partial occlusions and expression variations are more local deformations. The challenge lies in the fact that these local variations may occur anywhere and in any size or shape in a face image.

This paper proposes to address this from the aspect of partial similarity matching by exploiting the spatial contiguousness nature of occlusions and other local deformations [4]. The best way to visualize this process is by looking at Fig. 1.

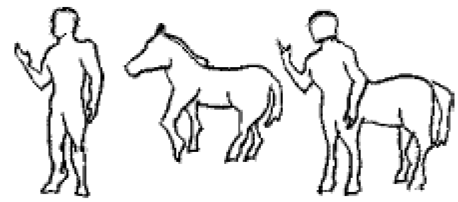

Fig. 1. Illustration of the intuition of partial similarity measures [5].

Most people will think that the man and horse are both similar to the centaur, even though the man and the horse do not share any similarities. A good explanation for this reasoning is that humans tend to “focus on the portions that are very similar and are willing to pay less attention to regions of great dissimilarity” [5].

The goal of this paper is to emulate the nature of human perception by comparing only those portions of face images that are similar, and disregarding the dissimilarities caused by occlusions or expression changes.

2. RELATED WORK

A significant amount of research seeking a reliable face recognition system has been conducted in the past. Many techniques attempt to reduce the influence of face variations as much as possible. Some of these methods include principal component analysis (PCA), linear discriminant analysis (LDA), Kernal PCA and LDA, Bayesian intra/extrapersonal classifier, and locality-preserving projection (LPP). All of these techniques tend to hold up well against “global” variations in face images, such as changes in lighting or age, however, they may not perform as well on images with large local deformations.

Ivanov *et al.* [6] introduced a “semi-local” method in which various components such as eyes, mouth, and nose are first detected by separate support vector machine (SVM) classifiers, and then a new (partial) face image is “reconstructed” with these components, which is further fed into another SVM classifier for final recognition [4]. This technique is similar to the one proposed in this paper, but it makes the assumption that the components of interest are not deformed or occluded, so that a reliable “reconstruction” can be acquired. In contrast, the proposed technique allows for any part of a face image to be deformed or occluded while performing recognition.

3. METHOD

The proposed face recognition algorithm consists of four steps (i) partitioning face images into sub-blocks, (ii) calculating a local pair wise similarity matrix between the probe image and every other image in the test set, (iii) computing the PD measures for the probe image, and (iv) using the PD measures to give the final identification of the image. Each step is explained in detail in the following subsections.

3.1. Partitioning Images into Sub-blocks

Each face image is partitioned into K (= $\mathrm{dim}_a/\mathrm{dim}_b$) nonoverlapping sub-blocks, where dim_a is the dimension of the whole image and dim_b is the dimension of the sub-block. Tan *et al.* [4] concluded that a sub-block size of 4 × 4 pixels is ideal. Each sub-block is represented by a local feature vector (LFV) by concatenating the pixels in the sub-block [4]. Fig. 2 loosely illustrates the partition process.

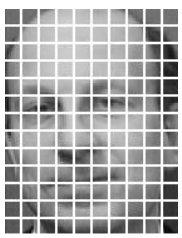

Fig. 2. Example of image partitioning.

3.2. Calculating the Local Pairwise Similarity Matrix

After the local regions of the face images have been defined, the local pairwise similarity matrix D between a probe face $\boldsymbol{q}$ and every face $\boldsymbol{x}_i$ in the training set T can be constructed. Let the set of sub-blocks of $\boldsymbol{q}$ and $\boldsymbol{x}_i$ be denoted as $\{\boldsymbol{q}\}_{k=1}^{K}$ and $\{\boldsymbol{x}_i\}_{k=1}^{K}$, respectively, where K is the number of sub-blocks. Each element of D can now be calculated as

$$d_{ki} = d(\boldsymbol{q}^k, \boldsymbol{x}_i^k) \tag{1}$$

which is the local pairwise distance between the two corresponding, kth sub-blocks from the probe face and the ith training face. The similarity matrix can be written as

$$D(\boldsymbol{q}, T) = \begin{bmatrix} d(\boldsymbol{q}^1, \boldsymbol{x}_1^1) & \dots & d(\boldsymbol{q}^1, \boldsymbol{x}_N^1) \\ d(\boldsymbol{q}^2, \boldsymbol{x}_1^2) & \dots & d(\boldsymbol{q}^2, \boldsymbol{x}_N^2) \\ \vdots & \vdots & \vdots \\ d(\boldsymbol{q}^K, \boldsymbol{x}_1^K) & \dots & d(\boldsymbol{q}^K, \boldsymbol{x}_N^K) \end{bmatrix} \tag{2}$$

where N is the number of training faces in T. Each column of D contains the local similarity distances between the probe face and the ith training face. The global distance $d(\boldsymbol{q}, \boldsymbol{x}_i)$ between the probe face $\boldsymbol{q}$ and $\boldsymbol{x}_i$ face can be calculated by summing the values in the corresponding column. The equation follows:

$$d(\boldsymbol{q}, \boldsymbol{x}_i) = \sum_{k=1}^{K} d_{ki} = \sum_{k=1}^{K} d(\boldsymbol{q}^k, \boldsymbol{x}_i^k). \tag{3}$$

3.3. Computing the PD Measure

A threshold τ can be used to divide the set of local pairwise distances $\{d_{ki}\}_{k=1}^{K}$ into two subsets, such that

$$S = \{\, k \mid d_{ki} \le \tau\,,\ k = 1,\ldots,\mathrm{K}\} \tag{4}$$

$$F = \{\, k \mid d_{ki} > \tau\,,\ k = 1,\ldots,\mathrm{K}\}. \tag{5}$$

S is the subset of similar local pairwise distances and F is the subset of dissimilar local pairwise distances. The threshold τ allows the proportion of these two subsets to be adjusted. Clearly, a larger threshold will result in a greater number of similar sub-blocks.

Equation (3) can be rewritten as

$$d(\boldsymbol{q}, \boldsymbol{x}_i) = \sum_k d_{ki} \sum_{k\epsilon S} d_{ki} + \sum_{k\epsilon F} d_{ki}. \tag{6}$$

The global distance between two face images is equal to the sum of the similar and dissimilar local pairwise distances. Therefore, equation (6) can be generalized to

$$d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta) = \beta \sum_{k\epsilon S} d_{ki} + (1-\beta) \sum_{k\epsilon F} d_{ki}. \tag{7}$$

$\beta \in [0,1]$ and balances the weight of the similar and dissimilar portions and is defined to be

$$\beta = \min\left(1, \frac{|S|}{|F|}\right) \tag{8}$$

where $|S|$ is the number of similar sub-blocks and $|F|$ is the number of dissimilar sub-blocks.

The distance defined in (7) is referred to as *partial distance* (PD). The problem of how to set the threshold τ will be addressed in Section 3.5.

One variant of the PD distance is called *continuous partial distance* (cPD). Rather than relying on a similarity threshold τ, cPD simply sums the first α most similar pairwise distances. The first step in calculating cPD is to sort the columns of the pairwise similarity matrix of (2) in increasing order such that

$$d_{ki} \le d_{\alpha i}, \quad \forall 1 \le k < \alpha \le \mathrm{K}. \tag{9}$$

Now, the cPD measure can be computed by using the following formula:

$$d_{cPD}(\boldsymbol{q}, \boldsymbol{x}_i) = \sum_{k \le \alpha} d_{ki}. \tag{10}$$

3.4. Using the PD Measure to Identify an Image

The process of identifying a probe face image is simple and can be achieved by

$$\text{label}(\boldsymbol{q}) = \min_{i=1\ldots N}(d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta)) \quad (11)$$

where label($\boldsymbol{q}$) returns the class label of the probe face image $\boldsymbol{q}$ and $d_{PD}(\boldsymbol{q}, \boldsymbol{x}_i, \beta)$ can be calculated using (7) for PD or (10) for cPD.

3.5. Setting the Similarity Threshold

The problem still remains of how to properly set the similarity thresholds α and τ, which are used in calculating cPD and PD, respectively.

The method proposed for setting the cPD threshold α involves the *golden section* strategy. Two quantities are said to be in the golden ratio if "the whole (that is, the sum of the two parts) is to the larger part as the larger part is to the smaller part" [4]. As discussed in the previous section, a face image has been partitioned into the similar subset S and the dissimilar subset F. If these partitions are golden proportion-compliant, then the similarity threshold α can be computed by

$$\frac{\text{total size of } x}{\text{total size of } S} = \frac{\text{total size of } S}{\text{total size of } F}$$

$$\begin{aligned}\text{total size of } S &= \frac{\sqrt{5}-1}{2} \times (\text{total size of } \boldsymbol{x}) \\ &= 0.618 \times (\text{total size of } \boldsymbol{x}) \end{aligned} \quad (12)$$

Therefore, a similar portion occupies approximately 61.8% of the entire face image.

Setting the threshold τ for use in the PD measure is a more involved process. Intuitively, good thresholds should be class-dependent in nature, i.e., different thresholds should be set for different persons (class) [4]. The proposed technique to accomplish this is based on the *maximum margin* criterion.

Let y_i denote the class label of the training image $\boldsymbol{x}_i$. The index set of the training examples belonging to the cth class is $H_c = \{i \mid y_i = c, i = 1,\ldots,N\}$, and the index set of the examples from the other class is $\overline{H}_c = \{i \mid y_i \neq c, i = 1,\ldots,N\}$. Then the training set of the cth class is denoted as $X_c = \{\boldsymbol{x}_i, i \in H_c\}$ with size $|H_c|$. Furthermore, denote the threshold of the cth class as τ_c [4].

The optimization process of τ_c is as follows: first, choose one image $\boldsymbol{x}_i \in X_c$ as the validation example, and leave every other example in the training set T as prototypes. Next, attempt to identify this example using PD under some test threshold τ_c using (11). If the identification result is $\text{FOO}(\boldsymbol{x}_i, \tau_c)$, then the average margin of the cth class is

$$\begin{aligned}\overline{m}_c(X_c, \tau_c) = \frac{1}{|H_c|} \sum_{i \in H_c} \{ 1(FOO(\boldsymbol{x}_i, \tau_c) = y_i) \\ \times [\min_{j \in \overline{H}_c} d_{\text{PD}}(\boldsymbol{x}_i, \boldsymbol{x}_j, \tau_c) \\ - \min_{j \in H_{c\, j \neq i}} d_{\text{PD}}(\boldsymbol{x}_i, \boldsymbol{x}_j, \tau_c)]\}. \end{aligned} \quad (13)$$

The above equation says that if a training example $\boldsymbol{x}_i$ can be successfully identified in the leave-one-out validation, then the classification confidence can be measured by the margin between the nearest training example outside of the class (the first term) and the nearest example (except for $\boldsymbol{x}_i$ itself) within the cth class (the second term).

The classification confidence for a class increases as this value grows larger. Therefore, equation (13) is used as the cost function which must be optimized in the training phase. Its output will be the class-dependent similarity threshold τ_c.

4. EXPERIMENTAL RESULTS

The ORL database is the sole source of test images for the proposed partial similarity technique. This database contains grayscale images of 40 people with 10 images per person. The images include faces that have glasses or no glasses, closed or opened eyes, and faces that are tilted or rotated up to 20°. No preprocessing has been performed on the images prior to testing and the native dimensions of 112 × 92 pixels have been maintained.

The ORL database is not known for providing largely occluded faces; however, it is the only database available for testing. The purpose of the proposed technique is to recognize faces under partial occlusions; therefore, a subset of the images has been "artificially" occluded for use in one of two experiments that follow.

4.1 Testing the Affects of Larger Training Sets

Ideally, classification accuracy should increase as larger training sets are used in the threshold learning phase. This experiment tests that hypothesis.

First, two separate training sets are created from the ORL database. One contains three images per person and the other contains five images per person. Class-dependent thresholds are then learned from each of these training sets using the *maximum margin* technique. Next, a test set is created from the remaining images in the ORL database, containing five test images per person. The training sets and the test set are disjoint. The recognition process is then performed on each of the test set images. Since there are five images per person, the top four matches (or ranks) are returned. The average accuracy for the entire test set is displayed in Fig. 3. The results for the same test set using cPD have been included for comparison.

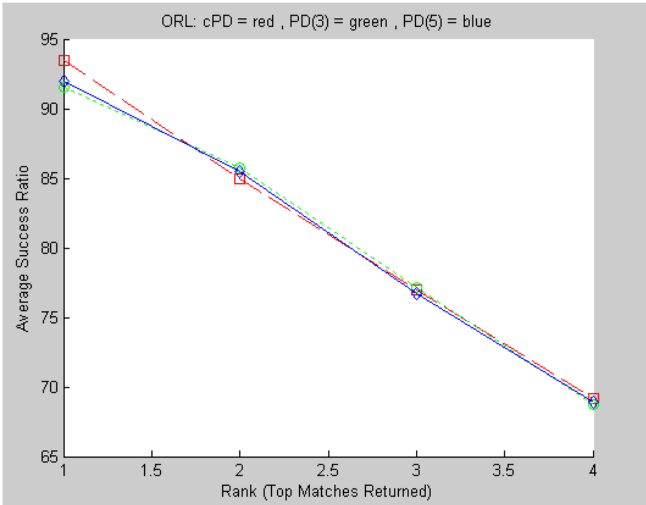

Fig. 3. Comparing the success of cPD, PD using thresholds learned from a training set of three images per person, and PD using thresholds learned from a training set of five images per person.

Fig.3. shows that all three methods perform almost equally on a test set of five images per person; however, it is unclear how much of an impact larger training sets have on the PD recognition process.

A second test is conducted using a training set and test set of ten images per person. Since there are ten test images per person, the top nine matches (or ranks) are returned. The results are displayed in Fig. 4. Again, cPD is included for comparison.

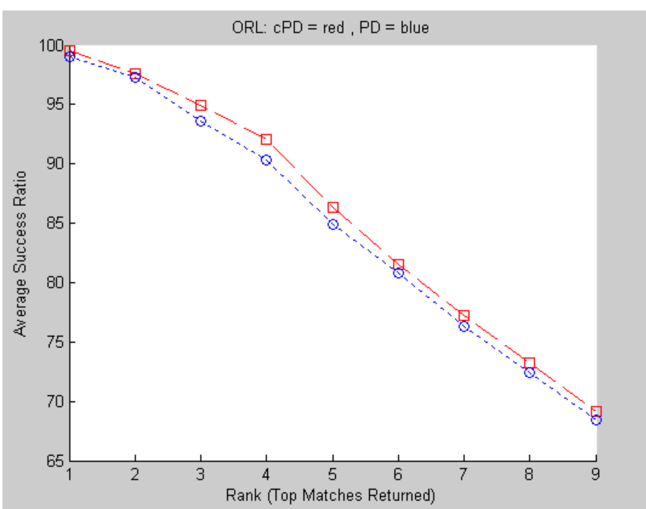

Fig. 4. Comparing the success of cPD and PD using thresholds learned from a training set of ten images per person.

The affect of using larger training sets is made more apparent when comparing the PD results from Fig. 4. to those in Fig. 3. The data from both charts is included in Table I for a better analysis.

Rank	cPD(5)	cPD(10)	PD(3)	PD(5)	PD(10)
1	93.5%	99.5%	91.5%	92%	99%
2	85%	97.6%	85.8%	85.5%	97.3%
3	77%	94.9%	77.2%	76.7%	93.6%
4	69.1%	92.1%	68.8%	68.9%	90.4%
5	X	86.4%	X	X	84.9%
6	X	81.5%	X	X	80.8%
7	X	77.3%	X	X	76.3%
8	X	73.3%	X	X	72.4%
9	X	69.2%	X	X	68.4%

Table I. Summary of the results shown in Fig. 3. and Fig. 4.

Table I makes it easy to see that using larger training sets in threshold does in fact increase the recognition accuracy for the PD technique. One can also see that cPD is quite robust as well.

4.2 Testing cPD and PD on Occluded Images

As previously mentioned, the ORL database is not known for offering images of largely occluded faces. This second experiment tests the robustness of the partial similarity techniques on ORL images that have been "artificially" occluded by applying various sized black patches at random locations in the images.

First, one face image per person is removed from the database. A square, black patch is then inserted into this image at a random location. The newly "occluded" images are the probe set, while the remaining 360 "clean" images serve as the test set.

Five different patch sizes are used to test how well the cPD and PD methods hold up to occlusions. The patches range from 15 × 15 pixels to 75 ×75 pixels, incrementing in steps of 15 pixels. It should be noted that the largest patch size of 75 ×75 pixels covers approximately 55% of a face image. The physical effects of the patching can be seen in Fig. 5. In this experiment, the PD method uses thresholds learned from the size ten training set found in the previous experiment. The results are displayed in Fig. 6.

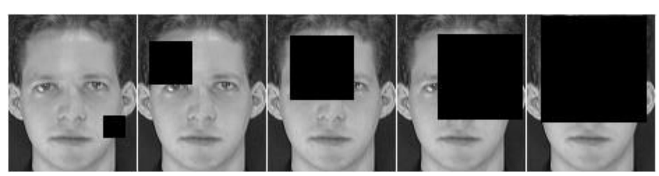
Fig. 5. "Artificial" occlusion patching used on the ORL database.

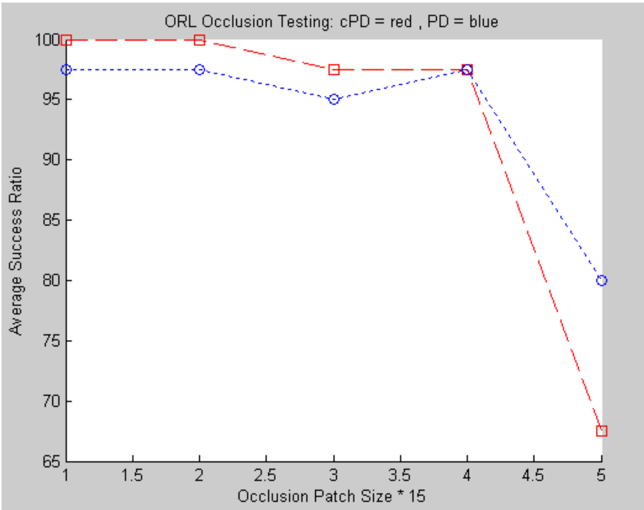

Fig. 6. Matching success of occluded faces using cPD and PD.

Fig. 6. shows that cPD and PD hold up quite well in the recognition of occluded images. The most notable result is the success of PD for the largest patch size of 75 × 75 pixels. The technique is able to achieve an 80% recognition success rate when over 50% of each probe image is occluded.

5. FUTURE WORK

The current system could be improved by increasing the run time efficiency so that the system could run in real time on video feeds. This would allow for the partial similarity face recognition techniques to be implemented in real time security systems. For example, the techniques could be used to identify known criminals in an attempt to prevent further crime. Combining this improvement with robotics could also prove useful in military applications.

6. CONCLUSION

This paper presented a robust system for identifying faces that are partially occluded or distorted by variations in facial expression. The algorithm compares similar portions of face images by using two types of similarity thresholds. Both the *continuous partial distance* (cPD) and *partial distance* (PD) methods produced impressive face recognition results and, while it was not the intention of this research, the techniques also proved to hold up quite well to face images taken at an angle.

7. REFERENCES

[1] W. Zhoa, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Survey*, vol. 34, no. 4, pp. 399-485, 2003.

[2] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, San Diego, CA, 2005, pp. 947-954.

[3] R. Gross, J. Cohn, and J. Shi, "Quo vadis face recognition," in *Working Notes of the 3rd Workshop on Empirical Evaluation Methods in Computer Vision*, Kauai, HI, 2001.

[4] X. Tan, S. Chen, Z. Zhou, and J. Liu, "Face recognition under occlusions and variant expressions with partial similarity," in *IEEE Transactions on Information Forensics and Security*, vol. 4, no. 2, June 2009.

[5] D. W. Jacobs, D. Weinshall, and Y. Gdalyahu, "Classification with non-metric distances: Image retrieval and class representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 6, pp. 583-600, June 2000.

[6] Y. Ivanov, B. Heisele, and T. Serre, "Using component features for face recognition," in *Proc. 6th Int. Conf. Automatic Face and Gesture Recognition*, pp. 421-426, May 2004.