A Robust Elastic and Partial Matching Metric for Face Recognition

Gang Hua Amir Akbarzadeh
Microsoft Corporate
One Microsoft Way, Redmond, WA 98052
{ganghua, amir}@microsoft.com

Abstract

We present a robust elastic and partial matching metric for face recognition. To handle challenges such as pose, facial expression and partial occlusion, we enable both elastic and partial matching by computing a part based face representation. In which N local image descriptors are extracted from densely sampled overlapping image patches. We then define a distance metric where each descriptor in one face is matched against its spatial neighborhood in the other face and the minimal distance is recorded. For implicit partial matching, the list of all minimal distances are sorted in ascending order and the distance at the αN-th position is picked up as the final distance. The parameter $0 \leq \alpha \leq 1$ controls how much occlusion, facial expression changes, or pixel degradations we would allow. The optimal parameter values of this new distance metric are extensively studied and identified with real-life photo collections. We also reveal that filtering the face image by a simple difference of Gaussian brings significant robustness to lighting variations and beats the more utilized self-quotient image. Extensive evaluations on face recognition benchmarks show that our method is leading or is competitive in performance when compared to state-of-the-art.

1. Introduction

Face recognition has been extensively studied in the community for several decades [1, 2, 8, 14, 17, 24, 29, 30]. It has been shown by the face recognition grand challenge [19] that under controlled settings, the recognition rate can be higher than 99% with false acceptance rate as low as 0.1%. Nevertheless, this is not the case when performing face recognition in uncontrolled real life photos [11]. Such photos include considerable visual variations caused by, for example, lighting [7, 21], difference in pose [21], facial expression [2] and partial occlusion [20].

It has been demonstrated that state-of-the-art photometric rectification techniques such as self-quotient image [26] can largely mitigate lighting variations except for extreme cases. In our experiments, we reveal that a simple difference of Gaussian (DoG) filter outperforms the more utilized self-quotient image method in handling normal lighting variations.

Figure 1. Examples of matched faces in our experiments. Notice the significant variations in lighting, pose, facial expression as well as partial occlusion. Each row shows two pair of matched faces.

Although there has been great research progress on face alignment [12], significant pose variations may still exist after alignment. Therefore, besides extreme lighting variation [7], pose, facial expression and partial occlusion remain great challenges. Intuitively, these challenges can largely be alleviated by designing robust face distance metrics that leverage both elastic and partial matching. By design principle, there are two types of face distance metrics: learning based metrics [24, 2, 8, 14, 30, 29] and hand-crafted metrics [1, 18, 15].

Inspired by the seminal work of Turk and Pentland on Eigen-Faces [24], learning based distance metrics have been a very active topic in face recognition. The predominant research efforts consist of identifying a discriminative embedding of faces in order to define a distance metric. Fisher-Faces [2], Laplacian-Faces [8] and their regularized variants [5, 4, 3] are all along this line. Other learned met-

rics include those based on SVMs [17] and Bayesian methods [16].

Nevertheless, learned metrics based on strongly supervised learning, such as linear discriminative embeddings [2, 8] or SVMs [17], need to be trained on the specific data-set that they deal with and most often the face images need to be aligned well to facilitate learning of meaningful structure of facial appearances. This makes them more suitable for controlled surveillance scenarios with limited subjects where labeled gallery faces are ready to be used for training. In addition, the training data-set has to be large enough to reduce the risk of over-fitting. For many face recognition tasks on real-life photos, these conditions may not be satisfied. Hence, it is desirable to have a plug-and-play distance metric, which does not need to be trained and can conveniently be used in face recognition tasks dealing with real-life photos.

Hand-crafted distance metrics do not suffer from the problems confronting the learned metrics. They also provide much more flexibility in incorporating elastic and partial matching schemes. For example, Elastic Bunch Graph Matching (EBGM) [28] represents each face by a graph, the nodes of which are a set of Gabor jets extracted from facial land-marks. Then a graph matching algorithm is designed to calculate the distance between two face representations. Notwithstanding their demonstrated success, graph matching is computationally intensive and the Gabor jets may not be discriminative enough for robust matching.

Ahonen et al. [1] proposed a distance metric that is calculated by a weighted sum of χ^2 distances, each of which is calculated between histograms of local binary patterns (LBP) on non-overlapping image partitions. However, it does not enable elastic matching at all, and hence, it is not robust to pose variations. Vivek and Sudha [18] proposed a partial Hausdorff distance metric where each pixel is represented as a binary vector similar to local binary pattern. The drawback of this method is that the spatial structure is largely discarded since a pixel in one face could be matched with any other pixel having the same local binary pattern in the other face. This is not desirable, especially when the faces are roughly aligned with each other.

In search for a robust face distance metric to handle all the challenges in face recognition, we take a part based face representation to enable elastic and partial matching, where a set of N local image descriptors [13, 27, 9, 23] are extracted from overlapping and densely sampled image patches. Then, in the matching process, each local image descriptor in one face image is compared against descriptors in its spatial neighborhood in the other face image and the minimal distance is recorded.

To perform partial matching, the list of all recorded minimal distances are then sorted in ascending order and the distance at the αN-th position is picked up as the final distance metric, where $0 < \alpha \leq 1$ is a control parameter on how much pixel degradation, facial expression changes and partial occlusions we would allow in the face images. The optimal parameter settings of our distance metric is extensively studied on real life photos obtained from several people's "family & friends" photo collections. With these optimal parameter settings, the proposed distance metric exhibits great robustness to pose variation, partial occlusion, as well as facial expression changes. This is demonstrated in our experiments on various face recognition benchmarks. We present in Figure 1 some matched faces in our experiments on real-life photo collections to demonstrate how robust our distance metric is to all the different visual variations.

The design of our distance metric follows the fundamental principle of the generalized Hausdorff distance. However, unlike its previous applications in face recognition, in which it was used to match edge points in the image space [6, 22], our distance metric is defined in the feature space, i.e., the space of the local image descriptors. Furthermore, we reinforce constraints to only allow each local image descriptor to be matched with its spatial neighbors in the image space. This spatial constraint is essential for matching two face images as shown in our experiments. This way, we perform explicit elastic matching and implicit partial matching in an efficient and robust way. Our main contributions are two fold:

1) We propose a novel robust partial matching metric for face recognition, which performs explicit elastic matching and implicit partial matching and shows leading performance when compared to the state-of-the-art.

2) We empirically show that a simple difference of Gaussian filter outperforms the more utilized self-quotient image and brings significant lighting invariance.

2. Part-based Face Representation

Figure 2 presents our pipeline used to extract the representation of a face. As illustrated in the figure, given an input image containing a face, we first run a variant [31] of the Viola-Jones face detector [25]. Next, the detected face image patch is fed into an eye detector, a convolutional neural network regressor, which localizes the left and right eye locations.

Our geometric rectification step is conducted by warping the face patch with a similarity transformation that places the two eyes into canonical positions in a patch of size $w \times w$ ($w = 128$ in our settings). The geometrically rectified face patch $\mathcal{I}$ is then passed through a DoG filter to obtain the photometrically rectified face patch $\hat{\mathcal{I}}$, i.e.,

$$\hat{\mathcal{I}} = \mathcal{I}_{\sigma_1} - \mathcal{I}_{\sigma_2}, \tag{1}$$

where $\mathcal{I}_\sigma$ is produced by smoothing $\mathcal{I}$ with a Gaussian kernel $\mathcal{G}_\sigma$ with standard deviation of σ pixels. Our empirical

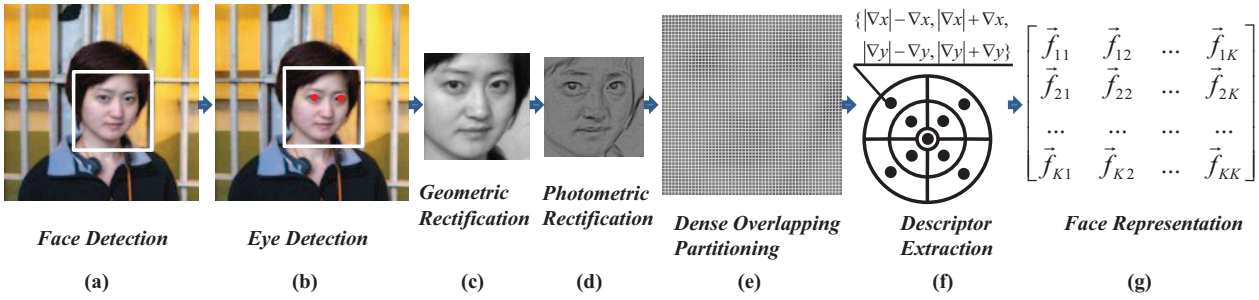


Figure 2. Our processing pipeline to extract the face representation.

investigations reveal that the combination of $\sigma_1 = 0$, which simply means no smoothing (i.e., $\mathcal{I}_{\sigma_1} = \mathcal{I}$), and $\sigma_2 = 1$ is optimal.

After the photometric rectification step, we densely partition the face image into $N = K \times K$ overlapping patches of size $n \times n$ ($n = 18$ in our settings), with both the horizontal and vertical step set to s pixels ($s = 2$ in our settings). To obtain our part-based representation, we compute a local image descriptor for each of the size $n \times n$ small patches. We adopt a variant of the descriptor T2-S2-9 proposed by Winder and Brown [27], which essentially accumulates 4 dimensional histograms of rectified image gradients $\{|\nabla_x| - \nabla_x, |\nabla_x| + \nabla_x, |\nabla_y| - \nabla_y, |\nabla_y| + \nabla_y\}$ over 9 spatial pooling regions, as shown in Figure 2(f). This descriptor provides excellent performance when matching image patches subject to different lighting and geometric distortions.

After we extract the local image descriptor for each of the local image patches, our final face representation is a matrix of $N = K \times K$ local image descriptors, i.e.,

$$F = [\vec{f}_{mn}],\ 1 < m < K,\ 1 < n < K \tag{2}$$

where $\vec{f}_{mn}$ corresponds to the descriptor extracted from the patch at location $(m \cdot s, n \cdot s)$ in pixel coordinates. Given such face representations we now proceed to define our distance metric.

3. Robust Elastic and Partial Matching Metric

We would like to utilize both elastic and partial matching to handle the different visual variations in face images. To calculate the distance between two face representations $F^{(1)}$ and $F^{(2)}$, we first perform elastic matching for each local descriptor $\vec{f}_{ij}$ in $F^{(1)}$. This is done by finding that descriptor's best match among its spatially neighboring descriptors in $F^{(2)}$. More formally, for each $1 \le i, j \le K$, we have

$$d(\vec{f}_{ij}^{(1)}) = \min_{k,l:|i\cdot s-k\cdot s|\le r,\ |j\cdot s-l\cdot s|\le r} \|\vec{f}_{ij}^{(1)} - \vec{f}_{kl}^{(2)}\|_1. \tag{3}$$

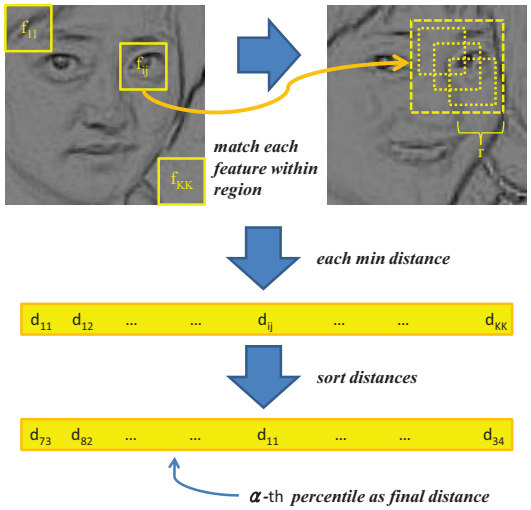


Figure 6. Illustration of our robust distance metric. Note however that the α-th percentile selection can be implemented with a quick selection algorithm, i.e. no explicit sorting is needed.

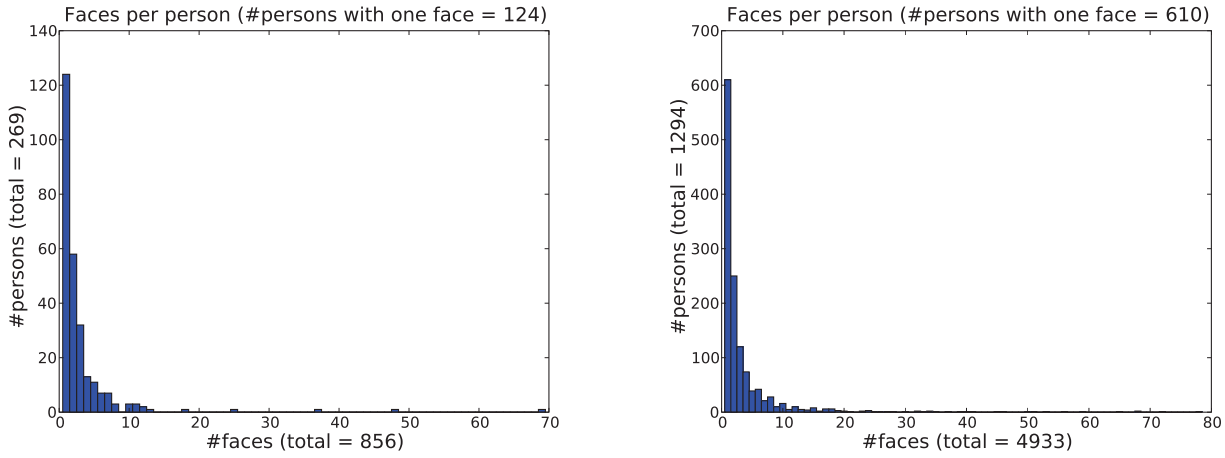
where $\|\cdot\|_1$ stands for the L1 norm, and r is a parameter controlling how much elasticity we would allow during matching. We name the neighborhood defined by r as r-neighborhood. Then, let

$$[d_1, d_2, \ldots, d_{\alpha N}, \ldots, d_N] = Sort_{ascend}\{d(\vec{f}_{ij}^{(1)})\}_{i,j=1}^{K} \tag{4}$$

be the sorted distances of all $d(\vec{f}_{ij}^{(1)})$ in ascending order, we define

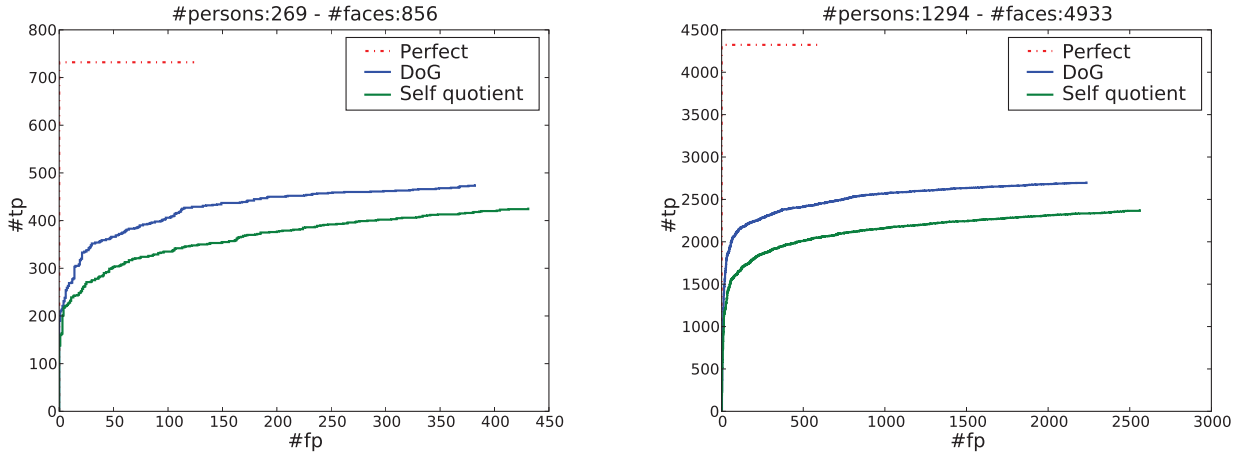
$$d(F^{(1)} \to F^{(2)}) = d_{\alpha N} \tag{5}$$

as the directional distance from F^1 to F^2, where $0 \le \alpha \le 1$ is a control parameter for partial matching. Note that it is not needed to do an explicit sorting in the implementation, we can instead use a quick selection algorithm. The parameter α controls how much pixel degradation, partial occlusion or facial expression changes we expect in the face images. Figure 6 visually illustrates how the distance metric is computed.



(a) FF1 (b) FF2

Figure 3. Histogram of the number of subjects owning a specific number of faces.



(a) FF1 (b) FF2

Figure 4. Comparison of DoG filter with self-quotient image.

Similarly, we can also define $d(F^{(2)} \rightarrow F^{(1)})$ and it is clear that most often $d(F^{(1)} \rightarrow F^{(2)}) \neq d(F^{(2)} \rightarrow F^{(1)})$. To make our distance symmetric, our final robust distance metric is defined as

$$D(F^{(1)}, F^{(2)}) = \max(d(F^{(1)} \rightarrow F^{(2)}), d(F^{(2)} \rightarrow F^{(1)})). \quad (6)$$

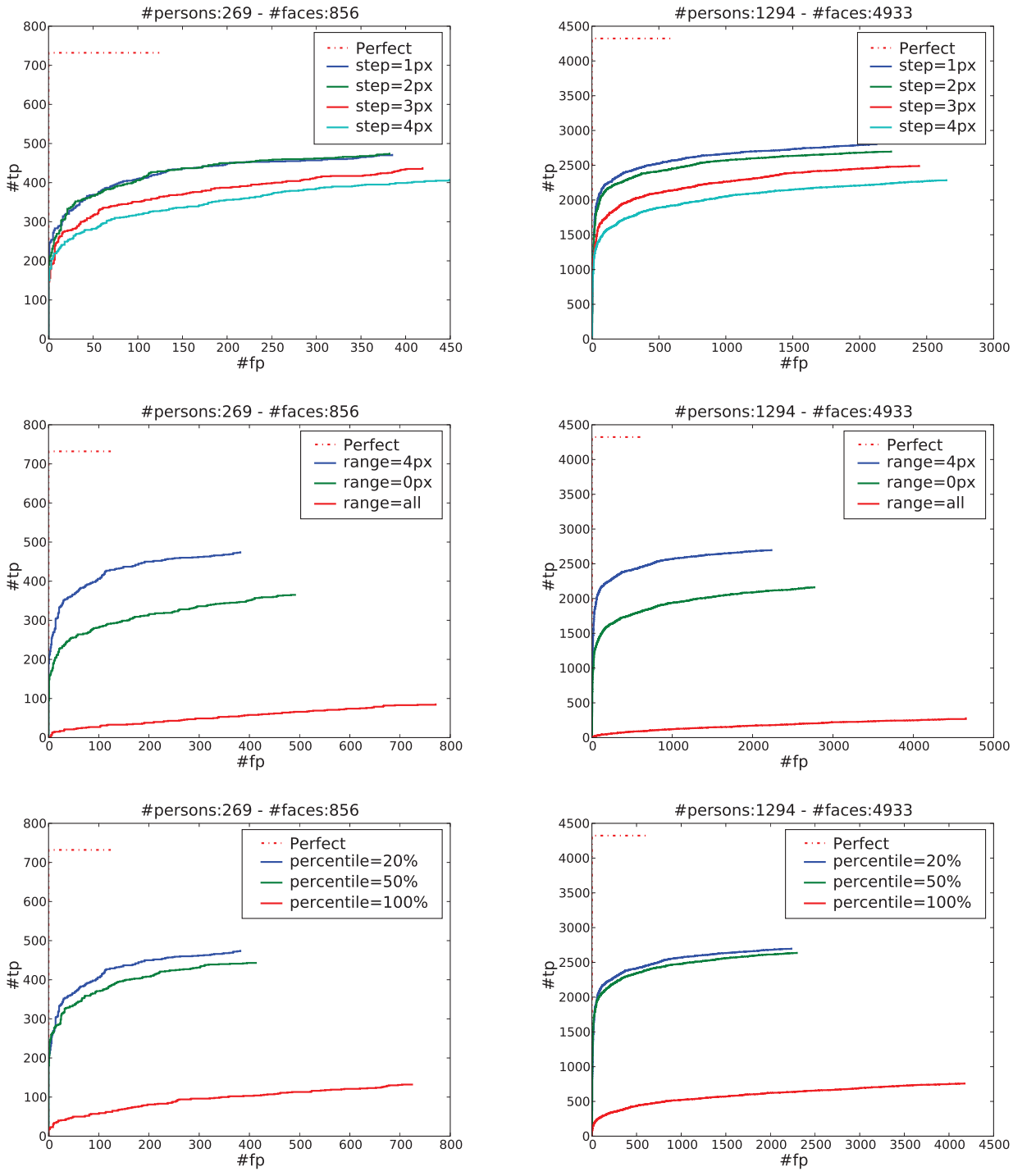
The following property of the proposed distance metric is trivially realized.

Property 3.1 *If $D(F^{(1)}, F^{(2)}) < V$, then at least α portion of the local image descriptors in $F^{(1)}$ ($F^{(2)}$) have a matched local descriptor in their r-neighborhood in $F^{(2)}$ ($F^{(1)}$) with distance less than V.*

Property 3.1 reflects how the proposed distance metric performs partial matching, i.e., the distance represents how well α portion of the face images are matched.

4. Experiments

We present extensive experiments to validate the quality of our face distance metric. Our somewhat optimized C++ implementation executes the distance metric at a speed of $0.23ms$ for a pair of faces (on a machine with a single core 3.0GHz CPU). This is excluding the time for extracting the face representations. Below, we first explore the parameter values of our distance metric on real-life photo collections. Then, fixing the optimal parameter values, we perform ex-



(a) FF1

(b) FF2

Figure 5. From top row to bottom row, the ROC curves for different values of the parameters s, r, and α, respectively. Both on FF1 (Column (a)) and FF2 (Column (b)).

tensive evaluation on benchmarks such as Labeled Faces in the Wild (LFW) [11], the Olivetti Research Laboratory (ORL) database [20], the Yale face database [2], and the CMU PIE database [21]. To make the presentation more

concise, we call them LFW, ORL, Yale, and PIE, respectively.

4.1. Parameter Exploration

There are three important parameters in our distance metric that we need to explore in order to obtain optimal performance. The first one is the patch sampling step parameter s in the face representation pipeline. It determines how many densely sampled patch descriptors we generate. The parameter s should be carefully chosen to balance between speed and recognition quality. The second parameter is the elasticity range parameter r, which defines the r-neighborhood an individual descriptor would match against. The third parameter is the partial matching control parameter α. As discussed before, α controls how much pixel degradation, partial occlusion, or facial expression changes we would expect.

To explore the effects of these parameters, we collected two real-life photo collections from several people's "family & friends" photo albums. We manually tagged all the faces that were detected by our face detector in these photos. We call them "family & friends" data set one and two, or in short FF1 and FF2, respectively. FF1 has a total of 856 faces of 269 subjects. While FF2 contains 4933 faces of 1294 subjects. The number of faces per subject is not uniform in these data sets. Figure 3 presents the distribution of how many faces images each subject has. The horizontal axis of the figure displays the number of faces, the vertical axis represents the number of subjects that have that number of faces. For example, in Figure 3 (a), we can see that there are 124 subjects in FF1 who only have 1 face, and in Figure 3 (b), we can see that there are around 250 subjects in FF2 who have 2 faces.

With each data set, we split the faces half and half per subject into two subsets. One subset is used as the gallery set and the other is used as the probe set. The recognition rate is evaluated by 1 nearest neighbor classification. A ROC curve is generated by picking a threshold on a ratio of the distances between the query face to the best matched gallery face and the distance between the query face to the second best matched gallery face with a different identity than the best matched one. If the ratio is below a certain threshold, then we accept the match, otherwise we do not accept the match. The horizontal and vertical axis present the number of falsely and correctly recognized faces among the accepted matches, respectively. Both subsets are served as gallery set once so the final ROC curve is the aggregation of two tests.

We have exhaustively run evaluations with all possible combinations of the two photometric rectification methods, i.e., DoG filter and the self-quotient image [26], with different settings of the parameters s, r and α. Our conclusion is that using the DoG filter, with $s = 2$ pixel, $r = 4$ pixel, and $\alpha = 0.2$ is the optimal setting. To better understand this investigation, in Figure 4 and 5, we present comparative ROC curves by setting each parameter to be a different value than the optimal one, while keeping the other settings at their optimal values. We also present in these figures the ground truth ROC curve (red dotted line) that a perfect face recognizer would achieve. Such a recognizer would refuse to accept a match for faces that do not have a corresponding gallery face, while correctly matching all the other faces which do have corresponding gallery faces.

More specifically, Figure 4 presents the recognition ROC curves using either the DoG filter (blue curve) or self-quotient image (green curve) for photometric rectification. We can clearly observe, on both FF1 and FF2, that using DoG achieves significantly better recognition rate than using self-quotient image. Similarly, the first, second, and third row in Figure 5 presents the effect of different values of the parameters s, r, and α, respectively. We can clearly observe how the recognition performance degrades when the value is not the optimal one.

We did note that overall $s = 1$ (blue curve) obtained slightly better performance than $s = 2$. However, $s = 2$ saves almost 4 times the computation time for extracting the local descriptors and also makes the face representation size 4 times smaller. Therefore, we choose to use $s = 2$ at the slight sacrifice of recognition performance. Another observation is that the optimal setting for α, an optimal value of 0.2, implies that the best matched 20% region in the face images largely determines the identity of the face.

Last but not least, we would like to emphasize the importance of using the parameter r to control the amount of elasticity we would allow in the distance metric. As clearly observed in the second row of Figure 5, neither allowing no elasticity $r = 0$ (blue line in the figure) nor allowing maximal elasticity $r > K$ (red line in the figure) is desired. The former does not handle pose variation well. While the latter does not take into consideration that face images are very structured, especially after they are roughly aligned. The matching of the local descriptors should not be over the whole face image.

In the following, we exclusively use the identified optimal settings for our distance metric to compare with the state-of-the-art on various face recognition benchmarks.

4.2. Experiments on Recognition Benchmarks

4.2.1 Face Recognition on LFW

We first present our recognition result on the LFW dataset [11]. We followed the test settings specified in [11]. The evaluation of the quality of a face recognition algorithm on LFW is to classify a pair of faces as either match or non-match based on the distance between them. From different threshold settings, a ROC curve is generated. Figure 7

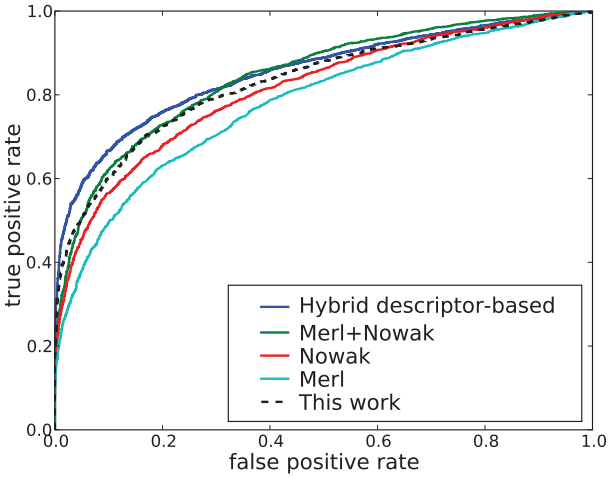


Figure 7. Our performance on the faces in the wild data set. Note that parameters for our algorithm is not tuned on this data set.

presents the ROC curve from our approach (black dotted line), along with the ROC curves of previous methods such as the hybrid descriptor based method proposed by Wolf et al. [29], Nowak's method on image matching applied to face recognition [17], the MERL face recognizer [10], and the combined MERL+Nowak method [10].

As shown in Figure 7, our method outperforms both the MERL recognizer and Nowak's method in both the low and high false positive regions. While the hybrid descriptor based method has the leading performance in the low false positive region and the MERL+Nowak method is leading in the high false positive region. Moreover, our method outperforms MERL+Nowak in the low false positive region and is comparable to the hybrid descriptor method in the high false positive region. We must emphasize that in order to show the generalization ability of our method, we did not leverage the training data provided in LFW to tune the parameters of our algorithm, we simply used the parameters reported above. Also notice that our algorithm is the top performing non-combination method, i.e. unlike hybrid descriptor-based and Merl+Nowak it does not combine the output of several algorithms.

4.2.2 Face Recognition on ORL, Yale, and PIE

We perform extensive experiments on three face recognition benchmarks including Yale, ORL, and PIE[1]. We briefly describe each of the data-sets and our experimental setting:

• The Yale database contains 165 faces of 15 subjects. There are 11 faces per subject, which are of different facial expressions. We randomly select 5 faces per person to form the set of gallery faces, and the rest are used as the probe faces.

• The ORL database contains 400 faces of 40 subjects. There are 10 faces per subject, which were taken at different time, lighting, and facial expressions. We randomly select 5 images per subject to form the set of gallery faces, and the rest are used to form the probe set.

• The PIE dataset contains 41368 images of 68 subjects with 13 poses, 43 illumination conditions, and 4 expressions. We used the images of the 5 nearly frontal poses (C05, C07, C09, C27, C29) under all illumination conditions and expressions, i.e., a subset of 11560 face images with 170 images per person. We randomly selected 30 images per person as the gallery faces, and the rest were used for testing.

In our experiments, the face images were all resized to 32×32 with the eyes aligned in the same canonical location. For each data set, we randomly split it 50 times and we report the average recognition error rate in percentage, see Table 1. The matches are calculated by 1 nearest neighbor classification. Note because the faces are already aligned, we bypassed the geometric rectification step when calculating our face representation.

Since the various discriminative embedding methods show the state-of-the-art performances on these dataset, we compare the performance of our algorithm mainly to them, including the Fisher-Face (LDA) [2], Laplacian-Face (LPP) [8], spatially smooth Fisher-Face (SLDA) [5], and the regularized Fisher-Face (RLDA) [3]. We also present two baseline results: the first one is taking Euclidean distances between the raw face images, and the second one is using Eigen-Faces (PCA) [24]. The recognition error rates of all the methods are also summarized in Table 1[2]. All the learned metrics are trained on the gallery faces. The face image vectors are all normalized to be unit vectors in preprocessing, as suggested in [5, 3].

From Table 1, it can be seen that our face distance met-

	Yale	**ORL**	**PIE**
Base	43.8 ± 4.1	13.7 ± 2.4	28.0 ± 0.6
PCA	$43.8 \pm 4.1_{74}$	$13.7 \pm 2.4_{199}$	$28.0 \pm 0.6_{1006}$
LDA	$21.2 \pm 3.4_{14}$	$7.2 \pm 1.7_{39}$	$7.9 \pm 0.3_{67}$
LPP	$21.1 \pm 3.4_{14}$	$6.8 \pm 1.6_{39}$	$7.5 \pm 0.3_{135}$
RLDA	$17.4 \pm 3.3_{14}$	$3.6 \pm 1.2_{39}$	$4.3 \pm 0.2_{67}$
SLDA	$14.9 \pm 3.2_{14}$	$2.3 \pm 1.0_{39}$	$3.6 \pm 0.2_{67}$
Ours	$\mathbf{9.4} \pm 3.1$	$\mathbf{1.6} \pm 0.9$	$\mathbf{2.4} \pm 0.2$

Table 1. Error rates on three academic benchmarks. The table enumerates the average face recognition error rate and the standard deviation in percentage over 50 runs. The subscripts present the optimal dimension of different embedding methods.

[1]The cropped and aligned faces of these datasets are obtained from Dr. Deng Cai at http://www.cs.uiuc.edu/homes/dengcai2/Data/FaceData.html

[2]To generate the comparison results, we used the Matlab script from Dr. Deng Cai at http://www.cs.uiuc.edu/homes/dengcai2/Data/FaceData.html.

ric outperforms all the learned metrics by a significant margin in all tests. This is even though we did not perform any training or parameter tuning on these data sets. For example, our method achieves error rates of $\mathbf{9.4}\% \pm 3.2\%$, $\mathbf{1}\% \pm 0.9\%$, $\mathbf{2.4}\% \pm 0.2\%$ on Yale, ORL, and PIE, respectively, while the second best results achieved by SLDA are $14.9\% \pm 3.1\%$, $2.3\% \pm 1.0\%$, and $3.6\% \pm 0.2\%$. These results clearly demonstrate the performance and generalization ability of our method.

5. Conclusion and Future Work

In this paper, we propose a robust distance metric for face recognition. The distance metric incorporates the ideas of both elastic and partial matching. The metric can deal with all various visual variations found in real life face photos. This includes lighting variation, difference in pose, facial expression, as well as partial occlusion. We extensively studied the effects of different parameters when dealing with photos from several people's real life photo collections. The distance metric with the optimal parameter settings was then evaluated on various face recognition benchmarks. The extensive evaluation demonstrated the excellent performance of our proposed approach. We also demonstrated that a simple DoG filtering is better than the more utilized photometric rectification method self-quotient image in handling lighting variations. Future research includes further exploration of using different local image descriptors as well as experimentation with the photometric rectification algorithm to handle more extreme lighting variations.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. In *ECCV*, 2004.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Trans. PAMI*, 19(7):711–720, 1997. Special Issue on Face Recognition.

[3] D. Cai, X. He, and J. Han. Spectral regression for efficient regularized subspace learning. In *ICCV*, 2007.

[4] D. Cai, X. He, J. Han, and H.-J. Zhang. Orthogonal laplacianfaces for face recognition. *IEEE Trans. Image Processing*, 15(11):3608–3614, November 2006.

[5] D. Cai, X. He, Y. Hu, J. Han, and T. Huang. Learning a spatially smooth subspace for face recognition. In *CVPR*, 2007.

[6] Y. Gao and M. K. Leung. Face recognition using line edge map. *IEEE Trans. PAMI*, 24(6):764–779, 2002.

[7] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. PAMI*, 23(6):643–660, 2001.

[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang. Face recognition using laplacianfaces. *IEEE Trans. PAMI*, 27(3):328–340, March 2005.

[9] G. Hua, M. Brown, and S. Winder. Discriminant embedding for local image descriptors. In *ICCV*,2007.

[10] G. B. Huang, M. J. Jones, and E. Learned-Miller. Lfw results using a combined nowak plus merl recognizer. In *Faces in Real-Life Images Workshop in ECCV*, 2008.

[11] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller. Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[12] L. Liang, R. Xiao, F. Wen, and J. Sun. Face alignment via component based discriminative search. In *ECCV*, 2008.

[13] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60:91–110, 2004.

[14] S. Lucey and T. Chen. Learning patch dependencies for improved pose mismatched face verification. In *CVPR*, 2006.

[15] J. Luo, Y. Ma, E. Takikawa, S. Lao, M. Kawade, and B.-L. Lu. Person-specific sift features for face recognition. In *Proc. ICASSP*, 2007.

[16] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.

[17] E. Nowak and F. Jurie. Learning visual similarity measures for comparing never seen objects. In *CVPR*, 2007.

[18] V. E. P and N. Sudha. Robust hausdorff distance measure for face recognition. *Pattern Recognition*, 40(2):431–442, 2007.

[19] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *CVPR*, 2005.

[20] F. Samaria and A. Harter. Parameterization of a stochastic model for human face identification. In *WACV*, 1994.

[21] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE Trans. PAMI*, 25(12):1615–1618, December 2003.

[22] B. Takacs. Comparing face images using the modified hausdorff distance. *Pattern Recognition*, 31(12):1873–1880, 1998.

[23] E. Tola, V. Leptit, and P. Fua. A fast local descriptor for dense matching. In *CVPR*, Anchograge, AK, June 2008.

[24] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, pages 586–591, June 1991.

[25] P. Viola and M. J. Jones. Robust real-time face detection. *IJCV*, 57:137–154, 2004.

[26] H. Wang, S. Z. Li, and Y. Wang. Generalized quotient image. In *CVPR*, 2004.

[27] S. Winder and M. Brown. Learning local image descriptors. In *CVPR*, 2007.

[28] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Trans. PAMI*, 19(7):775–779, 1997.

[29] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *Faces in Real-Life Images Workshop in ECCV*, 2008.

[30] J. Wright and G. Hua. Implicit elastic matching with randomized projections for pose-variant face recognition. In *CVPR*, 2009.

[31] C. Zhang and P. Viola. Multiple-instance pruning for learning efficient cascade detectors. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *NIPS*, 2008.