Robust Feature Set Matching for Partial Face Recognition

Renliang Weng[1], Jiwen Lu[2], Junlin Hu[1], Gao Yang[1] and Yap-Peng Tan[1]
[1]*School of EEE, Nanyang Technological University, Singapore*
[2]*Advanced Digital Sciences Center, Singapore*
Email: weng0017@e.ntu.edu.sg; jiwen.lu@adsc.com.sg

Abstract

Over the past two decades, a number of face recognition methods have been proposed in the literature. Most of them use holistic face images to recognize people. However, human faces are easily occluded by other objects in many real-world scenarios and we have to recognize the person of interest from his/her partial faces. In this paper, we propose a new partial face recognition approach by using feature set matching, which is able to align partial face patches to holistic gallery faces automatically and is robust to occlusions and illumination changes. Given each gallery image and probe face patch, we first detect keypoints and extract their local features. Then, we propose a Metric Learned Extended Robust Point Matching (MLERPM) method to discriminatively match local feature sets of a pair of gallery and probe samples. Lastly, the similarity of two faces is converted as the distance between two feature sets. Experimental results on three public face databases are presented to show the effectiveness of the proposed approach.

1. Introduction

A number of face recognition approaches have been proposed over the past two decades [22, 3, 1, 27, 18]. While these approaches have achieved encouraging results on some public databases, especially under controlled conditions, most of them use holistic face images to recognize people, where face images in both gallery and probe sets have to be pre-aligned and normalized to the same size before recognition. However, human faces are easily occluded by other objects in many real-world scenarios, especially in unconstrained environments such as smart visual surveillance systems. Hence, we have to recognize the person of interest from his/her partial faces, such as the examples shown in Figure 1. Therefore, it is desirable to develop a practical face recognition system which is able to process partial faces directly without any alignment and also robust to occlusions, variations of illumination and pose.

To make face recognition applicable in the real-life scenarios, several works have been presented to align probe facial images with training images automatically. Active Appearance Model (AAM) [8] endeavors to localize dozens of landmarks on facial images through an iterative search. Jia *et al*. [13] developed an automatic face alignment method through minimizing a structured sparsity norm. However, all these face alignment methods would fail to work if the probe image is an arbitrary face patch.

Figure 1. Several partial face samples. (a) Three partial face patches in the red ellipse are from the LFW database occluded by heads. [11]; (b) Partial faces with scarf and sunglasses occlusion in the AR dataset [19]. The objective of our study is to identify people from such partially occluded face images.

To deal with face occlusions, various algorithms based on sparse representation have been proposed recently [25, 28, 9, 16, 13], and [25] was the pioneer work in this area, where sparse representation was utilized to reconstruct occluded or stained facial images as well as to align probe face images to gallery images. While these approaches can achieve encouraging recognition performance in case of occlusions, they would fail if the probe image is an arbitrary face patch. In contrast to these methods, our approach processes partial face directly without manual alignment, which is more close to practical applications.

Feature set matching [7] has been a hot topic in pattern recognition. [24] was the first work that used graph matching for face recognition. However, their work relies heavily on manual landmarks labeling. Chui and Rangarajan [6] presented Robust Point set Matching (RPM) to align two feature sets according to their geometry distribution by learning a non-affine transformation function through iterative updates. However, it neglects textural information of

1550-5499/13 $31.00 © 2013 IEEE
DOI 10.1109/ICCV.2013.80

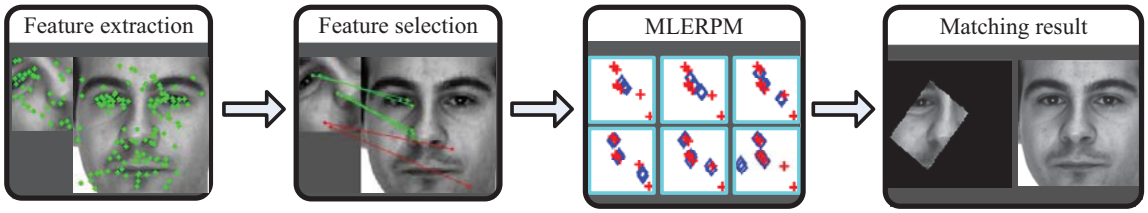

Figure 2. MLERPM-based partial face recognition framework. (a) Feature extraction: Keypoints detected by SIFT keypoint detector are marked out as green dots on both images. The left image is probe partial face image, and the right one is gallery face image. (b) Keypoint selection: correctly matched keypoints of these two images are connected by green lines, while two pairs of false matches are connected by red lines. (c) MLERPM process: point set of probe image marked out as blue diamond is iteratively aligned to the red-marked point set of gallery image from left top to the right bottom. Note that the two pairs of outliers are left alone after MLERPM while the rest are finely paired up. (d) Matching result: the left one is the warped image using the transformation parameters learnt from the matching process, the right one is the gallery image. Through MLERPM, the probe image is successfully aligned to the gallery image.

feature points. Liao *et al*. [15] utilized SRC to reconstruct probe local feature set with gallery feature sets, and they used the reconstruction error as distance metric. The main drawback of their method is that they neglected the geometry information of feature sets and their approach is computationally intensive.

To address the partial face recognition problem, we propose a new partial face recognition approach by using feature set matching, and devise a Metric Learned Extended Robust Point set Matching (MLERPM) approach to register the extracted local features according to their geometric distribution and textural information. Based on the matching result, a point set distance metric is proposed to describe the similarity of two faces. Our approach doesn't require manual face alignment and is robust to occlusions as well as illumination changes. Experimental results on three public face databases are presented to show the effectiveness of the proposed approach.

2. Proposed Approach

We propose to use local features instead of holistic features for partial face representation. Specifically, we apply the Scale-Invariant Feature Transform (SIFT) [17] feature detector to detect local feature keypoints, which are then concatenated with the Speeded Up Robust Features (SURF) [2]. Before matching, keypoints selection is performed to filter out obvious outliers. These selected keypoints of probe and gallery images are then matched by our MLERPM based on their geometric distribution and textural information, through which we obtain a one-to-one point set correspondence matrix to indicate the genuine matching pairs, as well as a non-affine transformation function to register geometric distributions of these matched keypoints. With matched keypoint pairs at hand, we design a point set distance metric to describe the difference between two faces based on MLERPM, where the lowest matching distance achieved would be reckoned as positive match. The face matching process is illustrated in Figure 2. Throughout the rest of the paper, matrix transposition is denoted by $'$.

2.1. Feature Extraction

Since there exist rotation, translation, scaling and even occlusions between probe image and gallery images of same identity, it is very difficult to normalize them to eye positions. Without proper face alignment, holistic features would fail to work. Hence, we proposed to use local features. Firstly, we detect keypoints with SIFT feature detector. Normally for a typical 128×128 face image, SIFT feature detector could output hundreds of feature points. The geometric feature of each keypoint, denoted as g, records its relative position in the image frame.

To describe the texture features of these detected keypoints, we combined the strength of SIFT and SURF keypoint descriptor by simple concatenation. SURF keypoint descriptor was introduced as a complement to SIFT for its greater robustness against illumination variations [14]. Hence, this augmented texture feature, denoted as t, is robust against in-plane rotation, scale as well as illumination change.

2.2. Keypoint Selection

As we have indicated previously, the number of keypoints of facial image could be up to hundreds. Matching point sets at this scale is computationally intensive. Moreover, irrelevant keypoints might hamper point set matching process, such as misleading the matching process to a local minimum, especially when genuine matching pairs are few among all matching features. Hence, it's beneficial to filter out obvious outliers before point matching.

We applied the idea of Lowe's matching scheme [17] for keypoint selection, which is to compare the ratio of distance of the closest neighbour to the one of the second-closest neighbour to a predefined threshold. The threshold was set as 0.5 in our experiments. These coarsely matched keypoint

pairs are then selected for our MLERPM for finer matching.

2.3. Metric Learned Robust Point Matching

After feature extraction and keypoints selection, for the probe partial face image, its geometry feature set is $\{g_1^P, g_2^P, ...g_{N_P}^P\}$, with its correspondent texture feature set as $\{t_1^P, t_2^P, ...t_{N_P}^P\}$, where N_P is the number of keypoints in probe feature set. Similarly, for the gallery image, we have $\{g_1^G, g_2^G, ...g_{N_G}^G\}$ and $\{t_1^G, t_2^G, ...t_{N_G}^G\}$ correspondingly. To align a probe partial face image to a gallery image automatically, we need match their correspondent geometric features and textural features respectively, which should have three characteristics:

- Subset matching: since the probe image and gallery images are not identical, some keypoints in the probe image couldn't find their correspondences in the gallery image. Likewise, not all keypoints in gallery images are ensured to be matched. Hence, this point set matching is a subset point matching problem.
- One-to-one point correspondence: this trait is obvious as keypoints of different positions in the probe image shouldn't be matched to a single keypoint in the gallery image.
- Non-affine transformation: the appearance of face changes when the perspective or facial expression changes. Such changes, when projected into the 2D image, are non-affine.

The work of Chui and Rangarajan [6] could meet most requirements listed above. However, its framework only considers feature points' geometric information. Hence we extended that framework to directly match textural features by introducing metric-learned texture distance as a regularizing term. Moreover, Chui's framework utilizes Thin-Plate Splines (TPS) [4] as non-affine transformation model. TPS tries to minimize a global bending energy function, which has a global nature, *i.e.* in order to match a non-smiling mouth in the probe face image to a smiling one of gallery image, it will tilt the whole probe image to make its mouth part smile, which however, would make the rest part of image highly distorted. Hereby we utilize radial basis function as the kernel function for the non-affine transformation.

The objective function of our proposed MLERPM algorithm is:

$$\begin{aligned} J = \min_{f,m} & \sum_{i,j} m_{ij} \left(\|f(g_i^P) - g_j^G\|_2^2 + \lambda_1 \|t_i^P - t_j^G\|_M^2 \right) \\ & - \tau \sum_{i,j} m_{ij} + C \sum_{i,j} m_{ij} \log m_{ij} + \lambda_2 \Psi(f) \\ s.t. & \sum_{j=1}^{N_G} m_{ij} \leq 1, \sum_{i=1}^{N_P} m_{ij} \leq 1, m_{ij} \geq 0 \end{aligned} \quad (1)$$

where m is the correspondence matrix and m_{ij} denotes the correspondence from keypoint i of probe image to keypoint j of gallery image. M is the metric matrix which would be detailed in section 2.5, f is the geometric non-affine transformation function and $\Psi(f)$ calculates the energy of its non-affine portion, both of which would be specified later.

In the above cost function, the first summation measures the total weighted cost of matching probe keypoint set and gallery keypoint set based on geometric and textural information. The second summation penalizes the case where only few point correspondences are established, and the third summation makes point correspondence fuzzy, that is m_{ij} could have any value between 0 and 1. Parameter C controls the fuzziness of correspondence matrix: as the value of C gradually decreases, m_{ij} moves towards to either 0 or 1, such that the correspondence between two point sets becomes more definite. τ, λ_1 and λ_2 are parameters which control tradeoffs between penalties.

Applying Chui's framework, we update the correspondence matrix and transformation parameters alternatively embedded in an annealing process:

Step1. Correspondence matrix update:

Correspondence between probe feature point i and gallery feature point j is updated by

$$m_{ij} = \exp\left(-\frac{\|f(g_i^P) - g_j^G\|_2^2 + \lambda_1 \|t_i^P - t_j^G\|_M^2}{2C} \right) \quad (2)$$

after which, rows and columns of correspondence matrix are iteratively normalized until convergence.

Step 2. Update the transformation parameters:

Our geometric non-affine transformation function is:

$$f(g_i^P) = A \times g_i^P + Q \times \phi(i) + b; \quad (3)$$

where A is a 2×2 affine transformation matrix and b is a translation vector, Q is a weight matrix associated with $\phi(i)$, the latter of which is a $k \times 1$ vector recording internal geometry structure of probe point set, defined as

$$\phi(i) = [\exp(-\frac{\|g_i - f_1\|_2^2}{\sigma}), ..., \exp(-\frac{\|g_i - f_k\|_2^2}{\sigma})]' \quad (4)$$

in which f_i is one of the k randomly selected anchor points from probe keypoint set, and σ controls the influence of anchor points: the larger σ is, the more global the transformation would be, which means anchor points far away from point g_i could have impact on it as well.

After dropping the terms independent of A, b and Q, the cost function of Eq. (1) becomes,

$$\min_{A,b,Q} \sum_{i,j} m_{ij} \left(\|f(g_i^P) - g_j^G\|_2^2 \right) + \lambda_2 \operatorname{tr}(Q \Phi \Phi' Q') \quad (5)$$

where Φ is the RBF kernel matrix whose ith column is $\phi(i)$, and $\operatorname{tr}(Q\Phi\Phi'Q')$ calculates the trace of $Q\Phi\Phi'Q'$, which

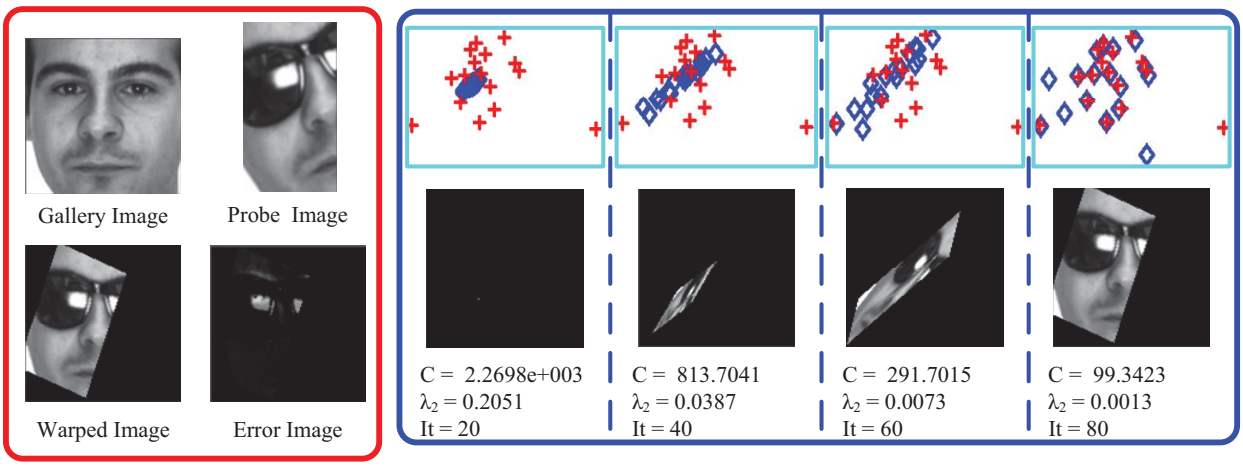

Figure 3. MLERPM matching process. In the red rectangle: the upper two images are gallery image and probe image, the lower two are warped image after MLERPM matching and error image. Error image records the absolute values of pixel-wise difference between gallery image and warped image. In the blue rectangle: Each column indicates the status of one iteration, the first row shows the matching process of geometric feature sets of gallery image and probe image, where the blue diamonds denote probe keypoints and red crosses denote gallery keypoints. The second row displays the warped image derived at each iteration, and the third row lists parameters' values, note that "It" is the iteration number.

is the $\Psi(f)$ in Eq. (1). Note that λ_2 controls the energy of non-affine transformation, a large λ_2 constrains the non-affine transformation part, while a small λ_2 encourages the image to transform freely around its anchor points. Hence it would be prudent to set λ_2 to a large value in the beginning, and gradually decrease it during the iteration process, as it's beneficial to align the matching images with affine transformation first before we get into detailed local warping (non-affine transformation).

For notational clarity, the probe geometric feature set is grouped into one matrix as X, where its ith column is g_i^P. Similarly, the gallery geometric feature set is grouped into Y. Furthermore, all transformation parameters are grouped into one matrix H, where $H = [A, b, Q]$, whose optimal value could be derived below:

$$H = Ym'\bar{X}'(\bar{X}\,\mathrm{diag}(m \times e_{N_G})\bar{X}' + \lambda_2\hat{X}\hat{X}')^{-1} \quad (6)$$

where e_{N_G} is an all-one vector with dimension as N_G. $\mathrm{diag}(v)$ is a diagonal matrix whose (i,i)th element is the ith element in vector v. $\bar{X}$ and $\hat{X}$ have the same size:

$$\bar{X} = \begin{pmatrix} X \\ e'_{N_P} \\ \Phi \end{pmatrix}, \; \hat{X} = \begin{pmatrix} \boldsymbol{O} \\ \Phi \end{pmatrix}$$

where $\boldsymbol{O}^{3\times N_P}$ is an all zero matrix.

We update between step 1 and step 2 alternatively, while gradually decreasing the values of C and λ_2, so that transformation parameters would be gradually refined and correspondences between two point sets would be more definite. The whole matching process is tabulated as Algorithm 1.

Algorithm 1: The MLERPM Algorithm:

Input: g^P, g^G, t^P, t^G, Φ
Output: A, b, Q, m
Parameters: $\lambda_1, \lambda_2, C, M, It_{max}, \epsilon$
Initialize A, b, Q
for $It = 1 : It_{max}$ **do**
  Step 1: update m using (2);
  Step 2: update (A, b, Q) using (6);
  Calculate J^{It} using (1);
  Decrease C and λ_2;
  if $|J^{It} - J^{It-1}| < \epsilon$ **then**
    break;
  end
end
return *A,b,Q,m.*

An example of our MLERPM matching process is illustrated in Figure 3. Note the probe facial image is not only rotated, scaled and translated from the gallery face image, it's occluded by sunglasses as well. During matching, temperature C is gradually decreased, so is λ_2. Meanwhile, transformation of the probe image keypoint set gradually becomes delicate: from shrinking the whole probe point set to a contracted single point cluster without knowing where to expand (in iteration 20) to the final refined perfect match (in iteration 80). It's also evident that outliers are automatically detected and left out during the matching process. This example shows that our MLERPM is robust to occlusions, rotation, translation and scaling.

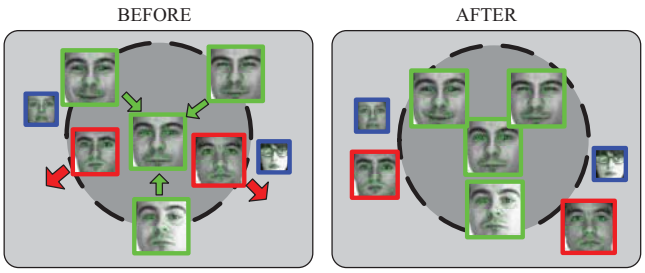

Figure 4. Illustration of one gallery image's neighborhood (the inner area rounded by dashed lines) variation before metric learning (left) versus after metric learning (right). After metric learning: (i) its three nearest gallery images with same labels are drawn nearer; (ii) similar gallery images with different labels are pulled farther.

2.4. Point Set Distance

Having obtained the transformation parameters between probe and gallery feature sets, we define point set distance metric S of two facial images as:

$$d_f = \frac{\sum_{i,j} m_{ij} \left(\|f(g_i^P) - g_j^G\|_2^2 + \lambda_1 \|t_i^P - t_j^G\|_M^2 \right)}{\sum_{i,j} m_{ij}}$$
$$S = \frac{d_f(1 + d_A)}{\sum_{i,j} m_{ij}} \tag{7}$$

where d_f calculates the average difference between matched keypoints, and d_A calculates the skewness of transformation matrix A. Defined as:

$$d_A = \left(\frac{k_2}{k_1 + \epsilon} - 1\right) \times \max\left(\frac{1}{k_1 + \epsilon}, (k_1 - 1)^2\right) \tag{8}$$

where k_1 and k_2 are eigenvalues of $A'A$, and $k_1 \leq k_2$, ϵ is a very small number, introduced to make the division meaningful in case that k_1 is close to zero.

The point set distance defined above is proportional to the skewness of transformation involved, and to the average matching difference. It's inversely proportional to the number of matched point pairs. This distance metric has intuitive interpretation: the number of matched point pairs indicates the area of two faces which are alike, the average matching difference points out the average resemblance of two faces share, and the skewness of transformation shows the facial shape dissimilarity of two faces.

2.5. Metric Learning for Point Set Distance

As mentioned previously, the augmented features we extracted were from concatenation of two feature descriptors, which in essence are features of different modalities, simple concatenation in Euclidean space cannot effectively represent the information carried by different features. Metric Learning [26] could exploit potential discriminating information of concatenated features through introducing a positive semi-definite matrix M, with which the new distance metric is defined as

$$\|t_i - t_j\|_M^2 = (t_i - t_j)' M (t_i - t_j) \tag{9}$$

Inspired by the work of Weinberger *et al*. [23], we proposed a point set metric learning scheme, so that the learned point set distance between feature sets from similar faces of same identity would be as small as possible, while distance between feature sets from similar faces belonging to different identities be enlarged. See Figure 4. Specifically, N gallery images covering all identities in the training data were selected to the metric learning process. For gallery point set G_p and G_q, according to Eq. 7 their point set distance is:

$$S_{pq} = \lambda_1 (1 + d_A) \frac{\sum_{i,j} \left(m_{ij}^{pq} \|t_i^{G_p} - t_j^{G_q}\|_M^2 \right)}{\sum_{i,j} m_{ij}^{pq}} + d \tag{10}$$

where m^{pq} is the correspondence matrix from point set G_p to G_q, and d is a constant unrelated to M. For each G_p, according to the derived point set distances, its n nearest neighbours with the same identity are selected to form its positive neighbourhood assembly, denoted as N_p^+. Similarly, its n nearest neighbours with different labels are chosen to form its negative neighbourhood assembly, denoted as N_p^-. With these information at hand, the metric learning cost function is

$$\begin{aligned} \min_M & \sum_{pq} \eta_{pq} S_{pq} + \zeta \sum_{pql} \xi_{pql} \\ s.t. \quad & S_{pl} - S_{pq} > 1 - \xi_{pql}, \\ & \xi_{pql} > 0, \eta_{pq} = 1, \eta_{pl} = -1 \end{aligned} \tag{11}$$

where η_{pq} denotes the relationship between G_p and G_q, if $G_q \subset N_p^+$, $\eta_{pq} = 1$. Similarly, if $G_q \subset N_p^-$, $\eta_{pq} = -1$. ζ is the cost associated with the penalty.

3. Experiments

To verify the effectiveness of our partial face recognition approach, we conducted partial face recognition for arbitrary face patch on the LFW dataset [11]. To comprehensively demonstrate the pros and cons of our approach, we did experiments of disguised and occluded partial face recognition on the AR [19] and Extended Yale B [10] datasets, respectively.

3.1. Data Sets

LFW Dataset: The Labeled Face in the Wild (LFW) dataset [11] contains 13233 labeled faces of 5749 people, in which 1680 people have two or more face images. Images in this dataset exhibit large appearance variations as

Figure 5. Example face images from the LFW dataset. First row: gallery images. Second row: probe partial face images randomly generated from another image of the same subject.

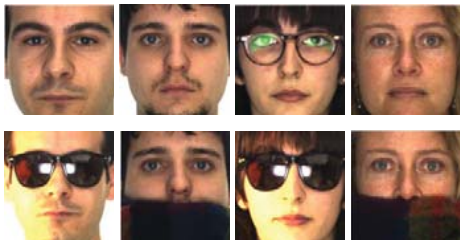

Figure 6. Samples from the AR dataset. First row: gallery images. Second row: probe images occluded by sunglasses and scarf.

Figure 7. Sample probe images in Extended Yale B dataset with random block occlusion with their correspondent occlusion levels are listed underneath.

they were taken from uncontrolled settings, including variations in scale, viewpoint, lighting condition, background, make-up, dress, expression, color saturation, image resolution, focus, *etc*., which pose a great challenge to our recognition task.

AR dataset: The AR dataset [19] contains 126 subjects, including 70 male and 56 female, respectively. For each subject, there are 26 face pictures taken in two different sessions (each session has 13 face images). In each session, there are 3 images with different illumination conditions, 4 images with different expressions, and 6 images with different facial disguises (3 images wearing sunglasses and 3 images wearing scarf, respectively).

Extended Yale B: There are 2414 frontal face images of 38 identities photographed under varying controlled illuminations in the Extended Yale B database. The public available cropped Yale database was used directly, whose image size is 192×168.

3.2. Experiment Settings

For the subjects in the LFW dataset, we chose the identities with no less than 10 images, from which we found 158 subjects. For subjects with more than 10 images, their first 10 pictures were selected for the experiment. These chosen images were then converted to gray-scale. For each subject, we randomly selected one image of him or her to synthetically produce a probe partial face image, while the other 9 images formed gallery set. For the gallery set, all images were normalized to 128×128 pixels according to the eye positions. Figure 5 shows some example normalized gallery face images (the first row). Note that our method is able to work on non-aligned gallery images as well.

Before extracting local features, we generated partial faces in a random way. Firstly we randomly rotated the whole image with rotation angle uniformly distributed in $[-10°, 10°]$, after which, the rotated image would undergo a random scaling between 0.8 to 1.2. Lastly, this scaled image was randomly cropped to $h \times w$, both of which were distributed within $[64, 128]$ uniformly. Some sample partial face images are shown in Figure 5 (the bottom row).

For the AR database [19], a subset containing 50 male subjects and 50 female subjects were selected from the first session in the AR dataset as in [25, 28]. For each identity, 14 images (without occlusion) were used for training, while 6 images with sunglasses and 6 images with scarves were selected for testing. For fair comparison with existing holistic methods, all these probe images and gallery images were cropped to 128×128 pixels and properly aligned. Figure 6 shows several cropped face images from the AR dataset.

For extended Yale B dataset, we randomly chose 32 images of each subject for training, and the remaining 32 for testing. In our experiments we synthesized contiguous-block-occluded images with occlusion levels ranging from 10% to 50%, by superimposing a correspondingly sized unrelated image randomly on each probe image, as in Fig. 7.

We used the same parameter setting scheme for all these three datasets: $\lambda_1 = 200,000/\operatorname{tr}(M)$, and the initial value of $\lambda_2 = 1$, the annealing rate for λ_2 is 0.92, which means, after each alternative update, the value of λ_2 would be decreased to $0.92\lambda_2$. For the other parameters related to matching we used the same setting as Chui's work [6]. Metric learning process was conducted respectively and we set n as 3 and penalty parameter ζ to 10.

3.3. Results and Analysis

Experiment 1: Partial Face Recognition on Arbitrary Patch. We conducted partial face recognition for arbitrary face patch on the LFW dataset. To demonstrate the effectiveness of our matching approach, we designed two groups of methods for comparison. The first group of comparing algorithms were designed to demonstrate the strength of metric learning and the merits of combining SIFT and SURF features. Specifically we added a variant of MLERP-M for comparison, wherein its metric matrix was simply an identity matrix. Hence we name this metric-learning free

Table 1. Recognition accuracy (%) of the comparing algorithms at various ranks on LFW

Method	Rank 1	Rank 10	Rank 20
RPM-SIFTSURF	0.63	3.16	6.96
HausDist-SIFTSURF	2.53	8.23	18.99
EMD-SIFTSURF	3.80	23.41	32.28
Lowe-SIFTSURF	24.68	49.37	55.06
ERPM-SURF	36.68	55.92	63.13
ERPM-SIFT	39.68	53.51	60.13
ERPM-SIFTSURF	42.09	58.92	66.74
MLERPM-SIFTSURF	**50.72**	**67.34**	**72.75**

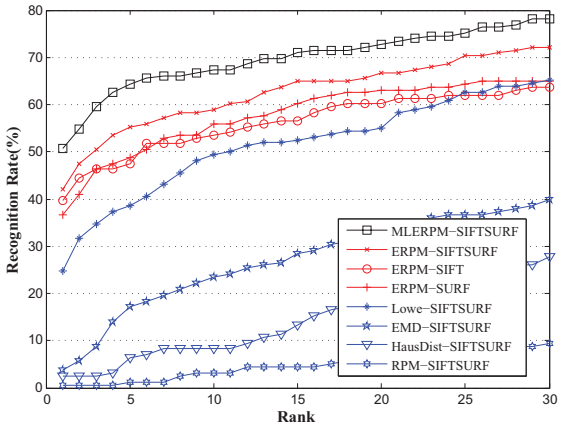

Figure 8. Recognition rates of the comparing algorithms at various ranks on LFW

method as Extended Robust Point set Matching (ERPM).

The second group of comparing algorithms work on either geometric features or textural features of concatenated SIFT and SURF feature sets (SIFTSURF). For geometry features, we deployed RPM [6] scheme directly, and used our point set distance metric as distance measurement (where we set λ_1 as 0). For textural features, we added three baseline methods. The first one was using Lowe's matching method to match textural features sets of gallery images and probe images, the number of matching pairs was set as similarity criterion. Hausdorff distance (HausDist) [12] was the second method which calculates the largest distance between closures of two texture feature sets. The third method was Earth Mover's Distance (EMD)[20], which measures the minimum cost of transforming one distribution of textural feature set into the other, where we set number of K-means clusters to 10 as it was the setting achieving the best recognition result. Table 1 and Figure 8 show the experimental results. Note that we put methods' names to the left of '-' and their correspondent features to the right of '-'. From the results we could make several observations:

1. Our method MLERPM-SIFTSURF obtained the best recognition rates. Note that it performed consistently better than ERPM-SIFTSURF at all ranks, showing the benefits of using metric learning for boosting the discriminating power of local features.

2. Within the ERPM-based category, ERPM-SIFTSURF performed the best, which proved that by combining SIFT and SURF descriptors, the invariance of local features to illumination, viewpoint, pose variations could be enhanced.

3. MLERPM-based and ERPM-based methods received better results than RPM, HausDist, EMD and Lowe's matching approaches. This is because matching only on geometry features or on texture features merely exploits partial information of face image, whereas both the geometry information and texture information of feature sets were considered by ERPM and MLERPM, resulting in a much more robust feature set matching.

4. RPM-SIFTSURF performed the poorest among all. This might be related to the fact that human faces generally share similar geometric structure. Based on these highly correlated geometric features alone, RPM approach could barely discriminate faces of different identities. Likewise, texture features alone are not robust enough for discrimination, which explains the poor performance of Lowe-SIFTSURF, EMD-SIFTSURF, and HausDist-SIFTSURF.

Experiment 2: Partial Face Recognition under Disguise: The AR dataset was selected for our partial face recognition under disguise. Table 2 records the recognition accuracy on the AR dataset with sunglasses, scarf and both, respectively. Our proposed method shows superior performance over the other state-of-the art methods on the AR dataset, which could be credited to our subset matching scheme: the correspondence values of keypoints located among occlusion parts, such as sunglasses and scarf, were gradually set to zero during the matching process, hence outliers' impacts on final distance metric were minimized. Only those matched keypoints in facial area were selected to point set distance calculation.

Experiment 3: Partial Face Recognition with Random Block Occlusion: The Extended Yale B dataset was selected for our partial face recognition under random block occlusion. We compared our algorithm with SRC [25], where we obtained some interesting results, as in Table 3. Before occlusion level arrived at 40%, our method performed comparably with SRC, but it degraded drastically when the occlusion percent is larger than 40%, while in the dataset of AR, our method did nearly perfectly where the percent of disguise for scarf is 40%. This is because in the experiment of AR dataset, disguise is either laid on the upper half or lower half of the face, discriminative features are

Table 2. Recognition accuracy (%) of the comparing algorithms at various ranks on AR.

Method	Sunglass	Scarf	Sunglass + Scarf
SRC [25]	87.00	59.50	73.25
CRC [28]	68.50	90.50	79.50
RoBM [21]	84.50	80.70	82.60
Stringfaces [5]	88.00	96.00	92.00
NNCW [16]	88.44	62.19	75.32
ℓ_1-ℓ_{struct} [13]	92.50	69.00	80.80
MLERPM	**98**	**97.00**	**97.50**

Table 3. Recognition accuracy (%) between SRC and MLERPM on Extended Yale B.

Occlusion	0%	10%	20%	30%	40%	50%
SRC [25]	100	100	99.8	**98.5**	**90.3**	**65.3**
MLERPM	**100**	**100**	**100**	98.3	80.2	30.2

almost half retained, while in this experiment, occlusion occurred randomly, *i.e*. in Figure 7, when occlusion percent is 50%, most part of face area is occluded, making face match extremely difficult. Hence our method is suitable for scenarios where sufficient discriminative facial areas are available.

4. Conclusion

In this paper, we have proposed a partial face recognition method by using robust feature set matching. We proposed to use local features instead of holistic features, and these local feature point sets were matched by our MLERPM approach, the outcome of which were a point set correspondence matrix indicating matching keypoint pairs and a non-affine transformation function. This transformation function could align the probe partial face to gallery face automatically. Moreover, a point set distance metric was designed, based on which, a simple nearest neighbor classifier could recognize input probe faces robustly even at presence of occlusions. Experimental results on three widely used face datasets were presented to show the efficacy and limitations of our proposed method, the latter of which pointed out the direction for our future work.

Acknowledgement

Jiwen Lu was partly supported by the research grant for the Human Sixth Sense Program (HSSP) at the Advanced Digital Sciences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore.

References

[1] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, 2002. 1

[2] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. *ECCV*, 2006. 2

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *PAMI*, 1997. 1

[4] F. L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *PAMI*, 1989. 3

[5] W. Chen and Y. Gao. Recognizing partially occluded faces from a single sample per class using string-based matching. In *ECCV*. 2010. 8

[6] H. Chui and A. Rangarajan. A new point matching algorithm for non-rigid registration. *CVIU*, 2003. 1, 3, 6, 7

[7] D. Conte, P. Foggia, C. Sansone, and M. Vento. Thirty years of graph matching in pattern recognition. *PAMI*, 2004. 1

[8] T. F. Cootes, G. J. Edwards, and C. J. Taylor. Active appearance models. *PAMI*, 2001. 1

[9] E. Elhamifar and R. Vidal. Robust classification using structured sparse representation. In *CVPR*, 2011. 1

[10] A. Georghiades, P. Belhumeur, and D. Kriegman. From few to many: Illumination cone models for face recognition under variable lighting and pose. *PAMI*, 2001. 5

[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. 1, 5

[12] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the hausdorff distance. *PAMI*, 1993. 7

[13] K. Jia, T.-H. Chan, and Y. Ma. Robust and practical face recognition via structured sparsity. In *ECCV*. 2012. 1, 8

[14] L. Juan and O. Gwun. A comparison of sift, pca-sift and surf. *International Journal of Image Processing*, 2009. 2

[15] S. Liao and A. K. Jain. Partial face recognition: An alignment free approach. In *IJCB*, 2011. 2

[16] Y. Liu, F. Wu, Z. Zhang, Y. Zhuang, and S. Yan. Sparse representation using nonnegative curds and whey. In *CVPR*, 2010. 1, 8

[17] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 2004. 2

[18] J. Lu, Y.-P. Tan, and G. Wang. Discriminative multimanifold analysis for face recognition from a single training sample per person. *PAMI*, 2013. 1

[19] M. Martinez and R. Benavente. The AR face database, 1998. 1, 5, 6

[20] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *IJCV*, 2000. 7

[21] Y. Tang, R. Salakhutdinov, and G. Hinton. Robust boltzmann machines for recognition and denoising. In *CVPR*, 2012. 8

[22] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *CVPR*, 1991. 1

[23] K. Q. Weinberger, J. Blitzer, and L. K. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2006. 5

[24] L. Wiskott, J.-M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *PAMI*, 1997. 1

[25] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *PAMI*, 2009. 1, 6, 7, 8

[26] E. P. Xing, A. Y. Ng, M. I. Jordan, and S. Russell. Distance metric learning, with application to clustering with side-information. *NIPS*, 2002. 5

[27] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *PAMI*, 2007. 1

[28] L. Zhang, M. Yang, and X. Feng. Sparse representation or collaborative representation: Which helps face recognition? In *ICCV*, 2011. 1, 6, 8