Topology Preserving Structural Matching for Automatic Partial Face Recognition

Yueqi Duan, Jiwen Lu, *Senior Member, IEEE*, Jianjiang Feng, *Member, IEEE*, and Jie Zhou, *Senior Member, IEEE*

***Abstract*— In this paper, we propose a topology preserving graph matching (TPGM) method for partial face recognition. Most existing face recognition methods extract features from holistic facial images. However, faces in real-world unconstrained environments may be occluded by objects or other faces, which cannot provide the whole face images for description. Keypoint-based partial face recognition methods such as multi-keypoint descriptor with Gabor ternary pattern and robust point set matching match the local keypoints for partial face recognition. However, they simply measure the nodewise similarity without higher order geometric graph information, which are susceptible to noises. To address this, our TPGM method estimates a non-rigid transformation encoding the second-order geometric structure of the graph, so that more accurate and robust correspondence can be computed with the topological information. In order to exploit higher order topological information, we propose a topology preserving structural matching method to construct a higher order structure for each face and estimate the transformation. Experimental results on four widely used face data sets demonstrate that our method outperforms most existing state-of-the-art face recognition methods.**

***Index Terms*— Partial face recognition, keypoint extraction, structural matching.**

I. Introduction

FACE recognition is a longstanding computer vision problem and a variety of face recognition methods have been proposed over the past three decades [1], [4], [20], [37], [49], [68]–[70], [80]–[82], [84]. While most methods have achieved impressive performance under controlled conditions where frontal holistic face images are prealigned and normalized, there are still challenges for unconstrained face recognition in many real-world applications. Typical face recognition applications include smart surveillance systems and handheld devices, where faces may be occluded by objects or other faces under crowded scenes. In such scenarios, only partial faces can be obtained and face alignment may fail to work with facial landmarks occluded. Fig. 1 shows some examples of partial face images.

(a) (b)

Fig. 1. Several examples of partial faces. (a) Faces are occluded by objects (sunglasses and a microphone) from YouTube Face (YTF) dataset [77]. (b) Faces in the red ellipse are occluded.

Most face recognition methods including the state-of-the-art CNN approaches [60], [64], [68], [69] utilize the whole face images for recognition, which are not applicable to partial face recognition. On the one hand, they assume that each image have the same content of the aligned whole face and describe the holistic facial images for representation. However, the contents of images may be different even for the same person in partial face recognition, e.g. one image without the left eye and another without the mouth, which leads to large intra-class variations. On the other hand, the occluded objects are included in the representation, which may harm the discriminative power. Therefore, it is required to design a partial face recognition method which directly recognizes partial faces and is robust to occlusions.

In order to better recognize partial faces, the occluded face parts should be removed when computing the similarities between probe partial faces and gallery faces. An intuitive idea to address this problem is to align the partial faces and then exclude the occlusions. However, these alignment methods need to detect facial landmarks at first which may be occluded. More recently, a few keypoint-based methods have been proposed by computing the similarity between probe partial faces and gallery faces over the detected feature keypoints on the face images [44], [76], such as multi-keypoint descriptor with Gabor ternary pattern (MKD-GTP) [44] and robust point

Manuscript received July 31, 2017; revised November 29, 2017 and January 31, 2018; accepted February 1, 2018. Date of publication February 12, 2018; date of current version March 27, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant U1713214, Grant 61672306, Grant 61572271, and Grant 61527808, in part by the National 1000 Young Talents Plan Program, in part by the National Basic Research Program of China under Grant 2014CB349304, and in part by the Shenzhen Fundamental Research Fund (Subject Arrangement) under Grant JCYJ20170412170438636. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Domingo Mery. *(Corresponding author: Jiwen Lu.)*

The authors are with the Department of Automation, Tsinghua University, State Key Lab of Intelligent Technologies and Systems, and Beijing National Research Center for Information Science and Technology, Beijing 100084, China (e-mail: duanyq14@mails.tsinghua.edu.cn; lujiwen@tsinghua.edu.cn; jfeng@tsinghua.edu.cn; jzhou@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIFS.2018.2804919

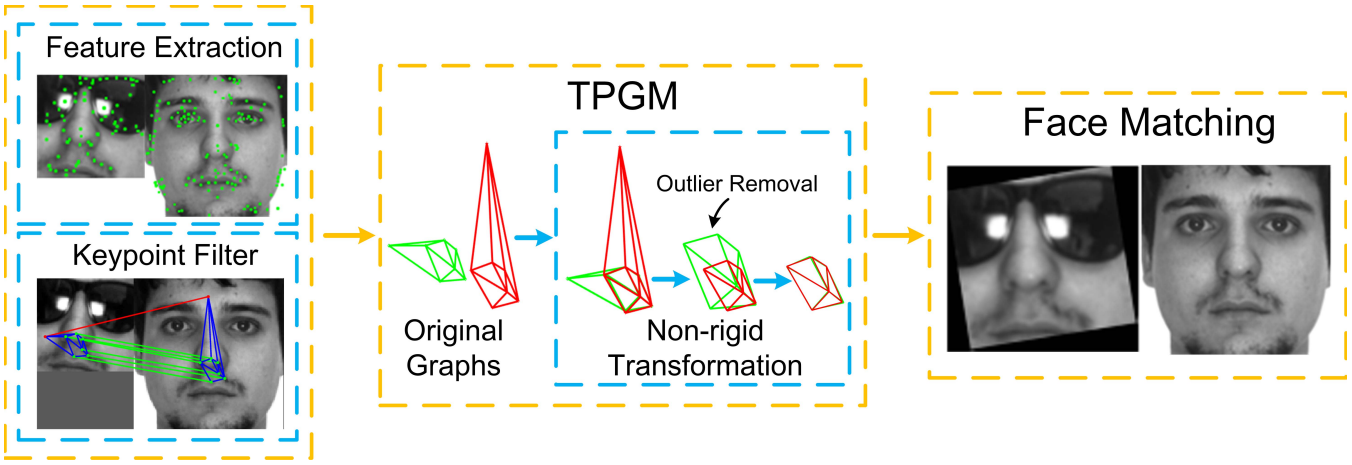


Fig. 2. The flowchart of our proposed TPGM approach for partial face recognition. For each pair of images, we first detect SIFT keypoint descriptors on both faces marked as green dots. Then, the keypoints are coarsely matched and selected based on Lowe's method, where green lines represent the correct matches while the red ones link the wrong matching keypoint pairs. We construct the graphs for both images by deploying Delaunay triangulation with blue lines. In TPGM procedure, a non-rigid transformation is estimated encoding the structural graph iteratively, with outliers removed during the iteration. Fig. 3 presents a more detailed illustration of the proposed TPGM. Finally, the probe partial face is aligned and recognized with the transformation and the matching result.

set matching (RPSM) [76], which achieve the state-of-the-art performance on partial face recognition. However, they only exploit node-wise similarity, which depend heavily on descriptors and are susceptible to illumination, deformation and other noises. Graph matching is an effective manner to address such unstableness by exploiting geometric structure of the graph, which enhances the robustness of feature matching in various of visual problems such as object categorization [15], feature tracking [28] and action recognition [5]. Inspired by the fact that graph matching delivers higher matching accuracy and stronger stableness, we propose a topology preserving graph matching (TPGM) method for partial face recognition by estimating a non-rigid transformation encoding the topological structure and measuring the correspondence between nodes and edges to compute the similarity between probe partial faces and gallery faces. Fig. 2 illustrates the pipeline of our proposed TPGM method. In order to exploit higher order structural information, we propose a topology preserving structural matching (TPSM) method by extending the second order geometric graph into higher order topological structure, so that more accurate and robust correspondence can be estimated. Experimental results on four widely used datasets including Labeled Face in the Wild (LFW), PubFig, AR and Extended Yale B (EYB) show the effectiveness of the proposed approach.

This paper is an extended version of our conference paper [10]. There are several new contributions:

1) We exploited higher order topological information to enhance the robustness of the transformation, where TPGM becomes a special case of the new TPSM.
2) We extracted the SiftSurfSILBPCNN descriptor rather than SiftSurfSILBP to present stronger discriminative power.
3) We conducted additional experiments on EYB and holistic face recognition settings of LFW and PubFig to show the effectiveness of the proposed methods.

II. Related Work

In this section, we briefly review three topics: 1) robust face recognition, 2) keypoint detection and description, and 3) graph matching.

A. Robust Face Recognition

There have been extensive work on robust face recognition with occlusion in recent years [11], [12], [22], [26], [31], [42], [48], [50], [58], [73], [78]. Lahasan *et al.* [33] summarized the recent strategies to overcome three major challenges in face recognition: occlusion, single sample per subject and expression. For example, Lahasan *et al.* [34] proposed a harmony search oriented-elastic bunch graph matching (HSO-EBGM) method by exploiting the optimal facial landmarks, which obtained outstanding performance with the occluded faces. Yang *et al.* [85] proposed a regularized robust coding (RRC) method for robust face recognition by regressing signals based on the maximum a posterior (MAP) principle under the assumption that the coding residual and vector are respectively independent and identically distributed. Zheng *et al.* [88] also proposed a sparse representation dictionary for robust face recognition. All these methods require face alignment, which has been proved to be critical by Ekenel and Stiefelhagen [16]. However, these face alignment based methods fail to work well with unknown missing facial portions. Also, a number of part-based representations have been proposed, which can be mainly classified into subregion-based methods [1], [49], [53], [59] and component-based methods [6], [21], [61]. Subregion-based methods divide face images into blocks and then compute the similarity between faces by integrating the matching results for the subregions. Component-based methods detect facial components such as eyes, nose and mouth, and then fuse the similarity of components for face matching. However, as occluded face regions in real-world applications are exceedingly unstructured, both

TABLE I
SUMMARIZATION OF RECENT FACE RECOGNITION METHODS

Approach	Category	Image Type	Geometry
SR [78]	Sparse Coding	Partial Face	Unexploited
RRC [85]	Sparse Coding	Partial Face	Unexploited
LH-ESRC [88]	Sparse Coding	Partial Face	Unexploited
MKD-GTP [44]	Keypoint Matching	Partial Face	Unexploited
RPSM [76]	Keypoint Matching	Partial Face	First Order
HSO-EBGM [34]	Graph Matching	Partial Face	Second Order
VGG [67]	Deep Learning	Holistic Face	Unexploited
FaceNet [64]	Deep Learning	Holistic Face	Unexploited
Center Loss [74]	Deep Learning	Holistic Face	Unexploited
Light CNN [79]	Deep Learning	Holistic Face	Unexploited
SphereFace [46]	Deep Learning	Holistic Face	Unexploited
TPGM	Graph Matching	Partial Face	Second Order
TPSM	Graph Matching	Partial Face	High Order

subregion-based methods and component-based methods may fail with incorrect occlusion detection or occluded facial components. More recently, a few keypoint-based approaches have been proposed [44], [76] which remove the occluded facial regions by computing on the detected feature keypoints. Liao *et al.* [44] proposed a MKD-GTP method which was the first general formulation of the partial face recognition. Weng *et al.* [76] applied point set matching method by considering the first-order compatibility between point sets, which achieves the state-of-the-art performance. However, these methods only exploit node-wise unary similarity without higher order geometrical graph information, therefore, they depend largely on descriptors and are susceptible to illumination, deformation and other noises. Table I summarizes the widely-used face recognition methods for clear demonstration.

B. Keypoint Detection and Description

Local keypoint detection and description are two essential steps for image matching, where keypoint detection finds repeatable image regions despite of changes and keypoint description captures distinctive and robust information of the interest regions. Early keypoint detector can be traced back to the work of Moravec [56]. Harris and Stephens [19] improved the Moravec detector by making it more repeatable with small image transformation and near edges, which becomes the well-known Harris corner detector. Mikolajczyk *et al.* [55] summarized and evaluated competitive detectors at the time. As Harris corner detector is scale-variant, Rosten and Drummond [62] proposed features from FAST criterion for keypoint detection, which was further improved by AGAST [51]. After detecting regions of interest, a descriptor is needed to describe each local keypoint. SIFT [47] is the most famous keypoint descriptor in the literature, as it provides high distinctiveness and robustness. In order to improve the efficiency of SIFT, Ke and Sukthankar [30] proposed PCA-SIFT to reduce the dimension of the descriptor from 128 to 36. Bay *et al.* [3] also presented SURF to obtain faster detection and description. Mikolajczyk and Schmid [54] proposed the GLOH descriptor to improve the distinctiveness of SIFT with higher computation cost. More recently, several learning based local descriptors have been proposed, which are more data-adaptive. For example, Hussain *et al.* [24] presented a local quantized pattern (LQP) method by improving LBP with a learned coding strategy. Lu *et al.* [49] proposed a compact binary feature descriptor (CBFD) by learning a hashing filter to project image patches to compact binary codes. Lin *et al.* [45] presented a deep learning approach to extract local descriptors in an unsupervised manner. Duan *et al.* [13] learns deep binary descriptor with multi-quantization (DBD-MQ) to minimize quantization loss.

C. Graph Matching

Recent years have witnessed a number of graph matching methods, which can be mainly classified into two categories: second-order graph matching [8], [9], [18], [38], [39], [89], [90] and hyper-graph matching [14], [35], [83], [86]. Second-order graph matching methods estimate the correspondence through the geometric similarity between nodes and edges, where the objective quadratic assignment problem is NP-hard. For example, Leordeanu *et al.* [39] optimized the objective function in the discrete domain. Cho *et al.* [8] presented a reweighted random walk graph matching (RRWM) framework by iteratively optimizing the candidate correspondences. Zhou and Torre [89], [90] factorized the affinity matrix into a Kronecker product of small matrices. One step further, hyper-graph matching methods exploit the higher-order relations, where third-order is the most frequently used. For example, Lee *et al.* [35] developed the RRWM [8] framework into higher-order and proposed a reweighted random walk hyper-graph matching method (RRWHM). Yan *et al.* [83] presented a discrete hyper-graph matching method by approximating the problem into a first-order assignment and optimizing the objective function in the integer domain.

III. PROPOSED APPROACH

In this section, we first model the partial face recognition as a geometric graph matching task and present the TPGM method. Then, we propose the TPSM approach to exploit higher order structural information. Lastly, we introduce how to use TPGM and TPSM for partial face recognition.

In order to incorporate the structural information, we utilize Delaunay triangulation to construct a graph for each image after keypoint extraction. There are four key properties to describe each graph: the location of nodes, the distance vector of edges, the descriptor for each node and the node-edge relations. Therefore, we present each graph as a 4-tuple $g = \{\mathbf{P}, \mathbf{Q}, \mathbf{T}, \mathbf{G}\}$, accordingly.

Let $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_n] \in R^{2\times n}$ be the set of nodes and $\mathbf{Q} = [\mathbf{q}_1, \ldots, \mathbf{q}_m] \in R^{2\times m}$ be the set of edges, where the edges are the difference vectors between the coordinates of connected nodes, and n and m are the number of keypoints and edges, respectively. Also, we denote the textural features of each node as $\mathbf{T} = [\mathbf{t}_1, \ldots, \mathbf{t}_n] \in R^{d_t\times n}$ and the node-edge relation matrix as $\mathbf{G} \in \{0, 1\}^{n\times m}$ to encode the topology of the graph, where $g_{ic} = 1$ if the cth edge connects the ith node. In particular, $g^P = \{\mathbf{P}^P, \mathbf{Q}^P, \mathbf{T}^P, \mathbf{G}^P\}$ represents the 4-tuple facial graph from the probe set, and $g^G = \{\mathbf{P}^G, \mathbf{Q}^G, \mathbf{T}^G, \mathbf{G}^G\}$ is from the gallery. The objective of TPGM and TPSM is to estimate a transformation from the probe graph to the gallery

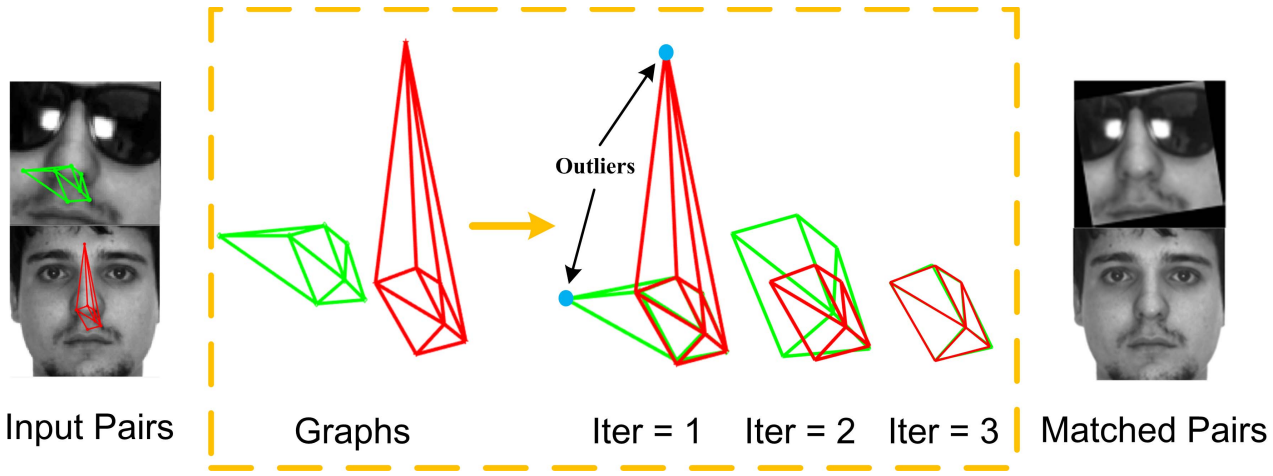


Fig. 3. An illustration of the proposed methods. For each image pair (with the probe partial face on the top and the gallery face on the bottom), we first construct the geometric graphs where the green structure represents the probe face and the red one for the gallery face. Then, the TPGM approach iteratively estimates the transformation to minimize both the textural and the geometric cost, and the outliers are removed during iteration. The probe graph gradually aligned to the gallery graph. Finally the graphs as well as the face images are perfectly matched in iteration 3.

graph to simultaneously minimize the textural and geometric costs, removing the outlier keypoints through the iterations. Fig. 3 illustrates the procedure of the proposed methods.

A. Feature Extraction and Keypoint Filtering

We first extract keypoints and their descriptions on facial images to construct the graph, and then filter the candidate keypoints through coarse matching.

First, we apply Scale-Invariant Feature Transform (SIFT) to detect keypoints [47], which is widely used for keypoint extraction. Then, we follow the "SiftSurfSILBP" [76] and combine the CNN descriptor to construct "SiftSurfSILBPCNN" for keypoint description. More specifically, we concatenate the SIFT descriptor with the Speeded Up Robust Features (SURF) to enhance the robustness to illumination variations [29], and further apply the Scale Invariant LBP (SILBP) features [43] to exploit the detailed textual information. For SILBP, we utilize its uniform pattern $LBP^{riu2}_{P,R}$ with four different values of {P,R}: {8,1}, {8,2}, {16,2} and {16,3}. Lastly, we apply VGG-16 [67] trained on the ImageNet, reducing the dimension to 128 with PCA. The proposed "SiftSurfSILBPCNN" combines the advantages of SIFT, SURF, SILBP and CNN, showing robustness to rotation, scale and illumination, which is important for keypoint description.

The number of keypoints detected by SIFT can be up to hundreds in a typical 128 $\times$ 128 facial image, which would suffer from heavy computational cost if we directly apply graph matching on all keypoints. Therefore, we apply Lowe's matching approach to filter out obvious outliers at first, and then perform our graph matching methods on the candidate keypoints.

B. Topology Preserving Graph Matching

Let $f^p(\cdot)$ and $f^q(\cdot)$ be the non-rigid transformation function of nodes and edges, respectively, where $f^q(\mathbf{q}_c) = f^p(\mathbf{p}_i) - f^p(\mathbf{p}_j)$ for $\mathbf{q}_c = \mathbf{p}_i - \mathbf{p}_j$. We denote $\mathbf{X} \in \{0,1\}^{n_p \times n_g}$ be the correspondence matrix of nodes where $\mathbf{X}_{ij} = 1$ for the matched ith probe keypoint and jth gallery keypoint, so that we can represent the node-wise mapping as $h^p(\mathbf{p}_i^P) = \mathbf{P}^G \mathbf{X}_i^{\mathrm{T}}$. Similarly, we can formulate the edge-wise mapping $h^q(\cdot)$ with the correspondence matrix of edges $\mathbf{Y} \in \{0,1\}^{m_p \times m_g}$. For $\mathbf{q}_i^P = \mathbf{p}_k^P - \mathbf{p}_l^P$ and $\mathbf{q}_j^G = \mathbf{p}_{k'}^G - \mathbf{p}_{l'}^G$, we set $\mathbf{Y}_{ij} = 1$ if both $\mathbf{X}_{kk'} = 1$ and $\mathbf{X}_{ll'} = 1$:

$$\mathbf{Y} = \mathbf{1}\Big[(\mathbf{G}^P)^{\mathrm{T}} \mathbf{X} \mathbf{G}^G > 1\Big], \tag{1}$$

where $\mathbf{1}(\text{true}) = 1$ and $\mathbf{1}(\text{false}) = 0$. The physical meaning of $(\mathbf{G}^P)^{\mathrm{T}} \mathbf{X} \mathbf{G}^G \in \{0,1,2\}^{m_p \times m_g}$ is the number of matching nodes connected with the corresponding edges, where edges are matched only if both the two pairs of connected nodes are matched.

With the denotations above, we formulate the objective function for partial face matching as follows:

$$\begin{aligned} \min J = {} & K_t(\mathbf{t}^P, h^t(\mathbf{t}^P)) + \lambda_p K_p(f^p(\mathbf{p}^P), h^p(\mathbf{p}^P)) \\ & + \lambda_q K_q(f^q(\mathbf{q}^P), h^q(\mathbf{q}^P)), \end{aligned} \tag{2}$$

where the $f(\cdot)$ function is the non-affine transformation, and the $h(\cdot)$ function links the input probe nodes (edges) to the output corresponding gallery nodes (edges).

The first term is to minimize the differences of textural features between the matched nodes. The second term aims to minimize the geometric distances between the transformed matched nodes. The physical meaning of the third term is to minimize the geometric distances between the transformed matched edges.

1) Textural Matching Cost: The first term in (2) measures the difference of descriptors for the matched keypoints. We obtain the textual matching cost by summing up the Euclidean distances of the matched descriptors:

$$K_t = \sum_i^{n_p} \sum_j^{n_g} \mathbf{T}_{ij} \mathbf{X}_{ij} = tr(\mathbf{T}^{\mathrm{T}} \mathbf{X}), \tag{3}$$

where

$$\mathbf{T}_{ij} = \sqrt{((\mathbf{t}_i^P - \mathbf{t}_j^G)^{\mathrm{T}}(\mathbf{t}_i^P - \mathbf{t}_j^G))}. \tag{4}$$

2) Node-Wise Matching Cost: The second term in (2) measures the geometric differences between the transformed probe keypoints and the corresponding gallery keypoints.

In order to incorporate the non-rigid transformation, we apply the thin plate spline (TPS) model, and the bending energy of TPS is formulated as follows:

$$\begin{aligned} K_p = &\sum_i^{n_g} \|f^p(\mathbf{p}_i^P) - h^p(\mathbf{p}_i^P)\|_2^2 \\ &+ \frac{\lambda}{\lambda_p} \iint \left[\left(\frac{\partial^2 f^p}{\partial x^2} \right)^2 + \left(\frac{\partial^2 f^p}{\partial x \partial y} \right)^2 \right. \\ &\left. + \left(\frac{\partial^2 f^p}{\partial y^2} \right)^2 \right] \mathrm{d}x \mathrm{d}y, \end{aligned} \tag{5}$$

where $f^p(\mathbf{p}_i^P)$ is computed as

$$f^p(\mathbf{p}_i^P) = \mathbf{A}_{2\times 2}\mathbf{p}_i^P + \mathbf{b}_{2\times 1} + \mathbf{W}\Phi(i)_{n_p\times 1}. \tag{6}$$

$$\Phi(i)_{n_p\times 1} = \begin{bmatrix} \|\mathbf{p}_i^P - \mathbf{p}_1^P\|_2^2 \log(\|\mathbf{p}_i^P - \mathbf{p}_1^P\|_2^2) \\ \|\mathbf{p}_i^P - \mathbf{p}_2^P\|_2^2 \log(\|\mathbf{p}_i^P - \mathbf{p}_2^P\|_2^2) \\ \vdots \\ \|\mathbf{p}_i^P - \mathbf{p}_{n_p}^P\|_2^2 \log(\|\mathbf{p}_i^P - \mathbf{p}_{n_p}^P\|_2^2) \end{bmatrix}, \tag{7}$$

where $\mathbf{A}$ and $\mathbf{b}$ are the affine transformation matrix and the translation vector. $\Phi(i)$ is an $n_p \times 1$ vector, representing the internal geometry structure with the TPS kernel of $\|\mathbf{p}_i^P - \mathbf{p}_j^P\|_2^2 \log(\|\mathbf{p}_i^P - \mathbf{p}_j^P\|_2^2)$. Φ is an $n_p \times n_p$ symmetric matrix with $\Phi(i)$ as its ith column. $\mathbf{W}$ is the weight matrix associated with Φ.

We substitute L1 norm for the L2 norm for linearization and rewrite the non-rigid transformation term as follows:

$$\begin{aligned} K_p &= \|\mathbf{A}\mathbf{P}_{2\times n_p}^P + \mathbf{b}\mathbf{1}_{n_p}^{\mathrm{T}} + \mathbf{W}\Phi_{n_p\times n_p} - \mathbf{P}_{2\times n_g}^G \mathbf{X}_{n_g\times n_p}^{\mathrm{T}}\|_1 \\ &\quad + \frac{\lambda}{\lambda_p}\|\mathbf{W}\Phi\|_1 \\ &= K_{p1} + \frac{\lambda}{\lambda_p} K_{p2}, \end{aligned} \tag{8}$$

where $\|\mathbf{W}\Phi\|_1$ represents the extent of affinity of the transformation. The transformation $f(\cdot)$ will be full affine if $\|\mathbf{W}\Phi\|_1$ is equal to zero.

3) Edge-Wise Matching Cost: The third term in (2) measures the geometric differences between transformed probe edges and corresponding gallery edges.

We calculate the geometric distance between a pair of edges with the squared L2 norm, where the difference between $\mathbf{q}_i^P$ and $\mathbf{q}_i^G$ is $\|\mathbf{q}_i^P - \mathbf{q}_i^G\|_2^2$. Let $\mathbf{q}_i^P = \mathbf{p}_{i1}^P - \mathbf{p}_{i2}^P$ and $\mathbf{q}_i^G = \mathbf{p}_{i1}^G - \mathbf{p}_{i2}^G$, and we obtain the following edge-wise objective function with the non-rigid transformation:

$$\begin{aligned} K_q &= \sum_i^{m_p} \|f^q(\mathbf{q}_i^P) - h^q(\mathbf{q}_i^P)\|_2^2 \\ &= \sum_i^{m_p} \|(f^p(\mathbf{p}_{i1}^P) - f^p(\mathbf{p}_{i2}^P)) \\ &\quad - (h^p(\mathbf{p}_{i1}^P) - h^p(\mathbf{p}_{i2}^P))\|_2^2. \end{aligned} \tag{9}$$

We reformulate and linearize the above equation into the matrix form:

$$\begin{aligned} K_q = \|&\mathbf{A}\mathbf{Q}_{2\times m_p}^P + \mathbf{W}_{2\times n_p}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j)_{n_p\times m_p} \\ &- \mathbf{Q}_{2\times m_g}^G \mathbf{Y}_{m_g\times m_p}^{\mathrm{T}}\|_1. \end{aligned} \tag{10}$$

Therefore, the objective function (2) can be rewritten as follows:

$$\begin{aligned} \min J = &tr(\mathbf{T}^{\mathrm{T}}\mathbf{X}) + \lambda_p\|\mathbf{A}\mathbf{P}_{2\times n_p}^P + \mathbf{b}\mathbf{1}_{n_p}^{\mathrm{T}} + \mathbf{W}\Phi_{n_p\times n_p} \\ &- \mathbf{P}_{2\times n_g}^G \mathbf{X}_{n_g\times n_p}^{\mathrm{T}}\|_1 + \lambda\|\mathbf{W}\Phi\|_1 \\ &+ \lambda_q\|\mathbf{A}\mathbf{Q}_{2\times m_p}^P + \mathbf{W}_{2\times n_p}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j)_{n_p\times m_p} \\ &- \mathbf{Q}_{2\times m_g}^G \mathbf{Y}_{m_g\times m_p}^{\mathrm{T}}\|_1. \end{aligned} \tag{11}$$

subject to

$$\begin{aligned} &\sum_i \mathbf{X}_{ij} \le 1, \quad \sum_j \mathbf{X}_{ij} \le 1, \ \mathbf{X}_{ij} \in \{0,1\}^{n_p\times n_g}, \\ &\mathbf{A}_{1,1} = \mathbf{A}_{2,2}, \quad \mathbf{A}_{1,2} = -\mathbf{A}_{2,1}. \end{aligned} \tag{12}$$

C. Optimization Details

As the objective function is NP-hard with discrete constraints which cannot be efficiently solved, we perform a linearization and rewrite the objective function into a linear programming (LP) form:

$$\begin{aligned} \min J = &tr(\mathbf{T}^{\mathrm{T}}X) + \lambda_p \mathbf{1}_2^{\mathrm{T}}\mathbf{U}\mathbf{1}_{n_p} + \lambda \mathbf{1}_2^{\mathrm{T}}\mathbf{V}\mathbf{1}_{n_p} \\ &+ \lambda_q \mathbf{1}_2^{\mathrm{T}}\mathbf{M}\mathbf{1}_{m_p} - \lambda_x \mathbf{1}_{n_p}^{\mathrm{T}}\mathbf{X}\mathbf{1}_{m_p}, \end{aligned} \tag{13}$$

subject to

$$\begin{aligned} &-\mathbf{U} \le \mathbf{A}\mathbf{P}^P + \mathbf{b}\mathbf{1}_{n_p}^{\mathrm{T}} + \mathbf{W}\Phi - \mathbf{P}^G\mathbf{X}^{\mathrm{T}} \le \mathbf{U}, \\ &-\mathbf{V} \le \mathbf{W}\Phi \le \mathbf{V}, \\ &-\mathbf{M} \le \mathbf{A}\mathbf{Q}^P + \mathbf{W}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j) - \mathbf{Q}^G\mathbf{Y}^{\mathrm{T}} \le \mathbf{M}, \\ &\mathbf{U} \ge 0, \ \mathbf{V} \ge 0, \ \mathbf{M} \ge 0, \\ &\sum_i \mathbf{X}_{ij} \le 1, \ \sum_j \mathbf{X}_{ij} \le 1, \ \mathbf{X}_{ij} \ge 0, \\ &\mathbf{A}_{1,1} = \mathbf{A}_{2,2}, \ \mathbf{A}_{1,2} = -\mathbf{A}_{2,1}. \end{aligned} \tag{14}$$

In (13), we relax the binary constraint of $\mathbf{X}_{ij} \in \{0,1\}^{n_p\times n_g}$ to a continuous domain $0 \le \mathbf{X}_{ij} \le 1$, and $\mathbf{U} \in R^{2\times n_p}$, $\mathbf{V} \in R^{2\times n_p}$ and $\mathbf{M} \in R^{2\times m_p}$ are auxiliary matrices representing the upper bounds of K_{p1}, K_{p2} and K_q, respectively. In order to avoid the degenerate cases when the optimal solution of $\{\mathbf{A}, \mathbf{W}, \mathbf{X}, \mathbf{b}\}$ are all zeros, we add a penalty term of $\mathbf{X}$ in the formulation.

1) Parameter Analysis: λ controls the effect of affine transformation and non-rigid transformation, where a large λ performs as a penalty term for the affine transformation, while a small λ would lead to non-rigid transformation. λ_x balances the number of matched pairs. A large λ_x preserves more matching pairs while a small λ_x is more strict with less matching pairs.

2) Region Shrinkage: We apply the successive candidate region shrinkage [27], [41] to solve the LP model. For each keypoint $\mathbf{p}_i^P$ in the probe image, we set a candidate region D_i in the gallery image, which is a circle with the center point of $f^p(\mathbf{p}^P)$ and the radius of r. We select the keypoints in the candidate region as the candidates, while the others are excluded in the optimization process by setting $\mathbf{X}_{ij} = 0$. We initialize the candidate regions as the holistic gallery image, and gradually shrinks the candidate regions to refine the matching candidates during the iteration. We formulate the candidate region shrinkage as $r^{(n+1)} = \alpha_1 \times r^{(n)}$, where $0 < \alpha_1 < 1$.

3) Parameter Shrinkage: As face images are globally rigid and locally non-rigid, we apply relatively rigid transformation in the global exploration period, and gradually convert to non-affine transformation in the local exploitation period. As λ controls the degree of the non-affine transformation, we initialize λ to a relatively large value, which is gradually decreased during iterations. More specifically, we apply $\lambda^{(n+1)} = \alpha_2 \times \lambda^{(n+1)}$, where $0 < \alpha_2 < 1$.

4) Outlier Removal: The Lowe's matching approach matches the keypoint pairs only based their descriptors, which is coarse and may exist many incorrect matchings. Such imposter matching pairs would harm the estimation of transformation in the process of graph matching, and increase the computational costs as well. Therefore, we remove the outliers during the iteration by calculating the summation of each row of $\mathbf{X}$. For an outlier probe $\mathbf{p}_j^P$, all elements in the jth row of $\mathbf{X}$ should be close to 0. Given a threshold $0 < \tau < 1$, we consider the $\mathbf{p}_j^P$ as an outlier and remove it from the candidate keypoints if $\sum_{k=1}^{n_p} \mathbf{X}_{jk} < \tau$.

Algorithm 1 summarizes the detailed procedure of the proposed TPGM method. We followed [76] by setting $\lambda_p = 0.01$, $\lambda_x = min(C)$ and $\lambda = 5$ for simplicity and a fair comparison. λ_q is fixed to 0.05 through cross validation for all the experiments.

D. Topology Preserving Structural Matching

The proposed TPGM only exploits second order graph to estimate robust correspondence, which ignore the higher order structural information. In order to exploit higher order topological information, we propose a topology preserving structural matching (TPSM) method by modelling a structure of higher order and estimate the transformation through the graph matching.

Similar to TPGM, we apply Delaunay triangulation to construct the graph where the point p_j is a neighbour point (NP) of p_i only when p_i links to p_j through a single neighbour edge (NE) q_c. In TPGM, we consider the point p_i as the first order information and its NE q_c as the second order. In order to exploit higher order structural information, we extend the NPs and NEs of a central point into more general nth order ones. More specifically, we define the order of a NP as the smallest number of passing edges to the central point, and the order of a NE is the sum of its two linking points' orders. Fig. 4 illustrates the high order NEs of a keypoint. As the 1st order NEs exploit 2nd order information of a node, nth order NEs represent the $(n+1)$th order information. When n is large, these edges express the configuration of the global graph structure. When n is small, these sets of edges represent the local structure, which is critical in partial identification.

Let $S(p_i)$ be the set of neighbour edges of p_i and E_n be the number of nth order neighbour edges of p_i, where $S^1(p_i) = \{q_{i,1}^1, q_{i,2}^1, \ldots, q_{i,E_1}^1\}$ is the 1st order NEs of p_i and $S^n(p_i) = \{q_{i,1}^n, q_{i,2}^n, \ldots, q_{i,E_n}^n\}$ is the nth order NEs. We formulate the structural matching cost $K_s(f^s(S(p^P)), h^s(S(p^P)))$ as

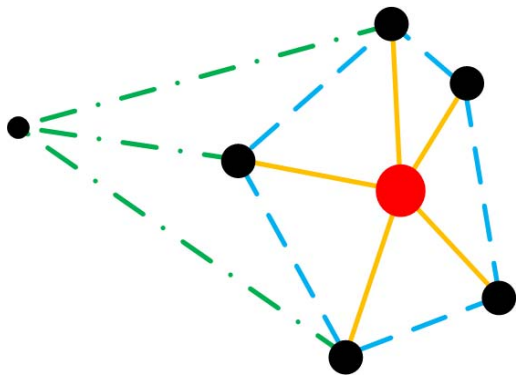


Fig. 4. An illustration of the high order NPs and NEs. For the central big point, the five neighbouring points are its 1st order NPs, and the small point on the top is the 2nd order as the shortest distance to the central point is two edges. With the orders of the NPs, the solid lines are 1st order NEs, the dashed lines are 2nd order NEs and the chain lines are 3rd order NEs.

Algorithm 1 TPGM

Input: Probe image P, Gallery image G, iteration number T and parameters λ_p, λ_x, λ, λ_q, τ, α_1 and α_2
Output: $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$

Detect keypoints and extract SiftSurfSILBP features.
Using Lowe's method to filter the keypoints and construct the graph g^P and g^G.
Initialize $r = r_{init}$ and set the constraint set $\Phi = \varnothing$.
for $t = 1, 2, \cdots, T$ **do**
  Construct Φ.
  Obtain $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$ using (13).
  Clear the constraint set Φ to $\varnothing$.
  Find the outsiders of the candidate region and construct the constraint set Φ.
  Shrinkage: $r^{(n+1)} = \alpha_1 \times r^{(n)}$ and $\lambda_2^{(n+1)} = \alpha_2 \times \lambda_2^{(n+1)}$.
  Remove $\mathbf{p}_i^P, \mathbf{X}_{i:}$, if $\sum_j \mathbf{X}_{ij} < \tau$.
  Remove $\mathbf{p}_j^G, \mathbf{X}_{:j}$, if $\sum_i \mathbf{X}_{ij} < \tau$.
end for
Binarize $\mathbf{X}$.
return $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$

follows:

$$K_s = \sum_{i,k} K_s(f^s(S^k(\mathbf{p}_i^P)), h^s(S^k(\mathbf{p}_i^P))) = \sum_{i,j,k} w_k \| f^q(q_{i,j}^{k,P}) - h^q(q_{i,j}^{k,G}) \|_2^2 \quad (15)$$

where w_k is the weight of structures under different orders. In general, w_k decreases with the increasing order of k, because the structural matching costs should be more sensitive for the nearby local structure rather than structures in distance.

If we set $k = 1$, the 1st order structures cover all the edges in this special case:

$$\{S^1(\mathbf{p}_1^P), \ldots, S^1(\mathbf{p}_{(n_p)}{}^P)\} = \{\mathbf{q}_1^P, \ldots, \mathbf{q}_{m_p}^P\}. \quad (16)$$

If we further set $w_1 = 1$, we obtain:

$$K_s = \sum_{i}^{n_p} \sum_{k}^{E_1} \| f^q(q_{i,j}^P) - h^q(q_{i,j}^G) \|_2^2 = \sum_{i}^{m_p} \| f^q(\mathbf{q}_i^P) - h^q(\mathbf{q}_i^G) \|_2^2 = K_q(f^q(\mathbf{q}^P), h^q(\mathbf{q}^P)), \quad (17)$$

where the structural matching cost in TPSM degenerates into the edge-wise matching cost in TPGM with both n and w_1 equal to 1.

Similarly, we formulate (15) into a matrix form:

$$K_s = \sum_{i}^{n_p} \sum_{k}^{n} K_s(f^s(S^k(\mathbf{p}_i^P)), h^s(S^k(\mathbf{p}_i^P))) = \sum_{k}^{n} w_k \| \mathbf{A}\hat{\mathbf{Q}}_{2\times E_k^P}^P + \mathbf{W}_{2\times n_p}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j)_{n_p \times E_k^P} - \hat{\mathbf{Q}}_{2\times E_k^G}^G \hat{\mathbf{Y}}_{E_k^G \times E_k^P}^T \|_1, \quad (18)$$

where $\hat{\mathbf{Q}}_{2\times E_k^P}^P$ is the list of all kth order neighbour edges of all keypoints in the probe set, and $\hat{\mathbf{Q}}_{2\times E_k^G}^G$ is the list of all kth order neighbour edges in the gallery set. r is the rank of $\mathbf{X}$ to show the exact number of matching pairs, so that $E_k^P = r \times m_p$ and $E_k^G = r \times m_g$. We set $\hat{\mathbf{Q}}_{2\times E_k^P}^P = [\hat{\mathbf{Q}}_{2\times m_p}^{1,k,P}, \hat{\mathbf{Q}}_{2\times m_p}^{2,k,P}, \ldots, \hat{\mathbf{Q}}_{2\times m_p}^{n_p,k,P}]$ where each $\hat{\mathbf{Q}}_{2\times m_p}^{i,k,P} = \{q_{i,1}^{k,P}, q_{i,2}^{k,P}, \ldots, q_{i,m_p}^{k,P}\}$ is the kth order neighbour edges of ith probe node. $\hat{\mathbf{Y}}_{E_k^G \times E_k^P}^T$ is a block diagonal matrix with r identical $\mathbf{Y}_{m_g \times m_p}^T$ on its diagonal. The predefined adjacency matrix $\mathbf{M}$ represents the 1st order node connection. We further define the power of a matrix. Let $\mathbf{M}^0 = I$ and $\mathbf{M}^k = \mathbf{M} \times \mathbf{M}^{k-1}$ for $k \neq 1$, and $\mathbf{M}^k$ represents the kth order adjacency, where $M_{ij}^k = 1$ means that node i is one of the kth order neighbour points of node j. $\hat{\mathbf{Q}}_{2\times E_k^P}^P$ can be pre-computed using $\mathbf{M}^k$.

Similarly, with the extension of the structure geometric matching cost, we obtain the objective function of TPSM:

$$\begin{aligned} \min J &= K_t + \lambda_p K_p + \lambda_s K_s \\ &= tr(\mathbf{T}^T\mathbf{X}) + \lambda_p \| \mathbf{A}\mathbf{P}_{2\times n_p}^P + \mathbf{b}\mathbf{1}_{n_p}^T \\ &\quad + \mathbf{W}\Phi_{n_p \times n_p} - \mathbf{P}_{2\times n_g}^G \mathbf{X}_{n_g \times n_p}^T \|_1 + \lambda \| \mathbf{W}\Phi \|_1 \\ &\quad + \lambda_s \sum_{k}^{n} w_k \| \mathbf{A}\hat{\mathbf{Q}}_{2\times E_k^P}^P + \mathbf{W}_{2\times n_p}(\Phi P_i - \Phi P_j) \\ &\quad - \hat{\mathbf{Q}}_{2\times E_p^G}^G \hat{\mathbf{Y}}_{E_p^G \times E_p^P}^T \|_1 \end{aligned} \quad (19)$$

subject to

$$\sum_i \mathbf{X}_{ij} \leq 1, \quad \sum_j \mathbf{X}_{ij} \leq 1, \ \mathbf{X}_{ij} \in \{0,1\}^{n_p \times n_g}, \quad \mathbf{A}_{1,1} = \mathbf{A}_{2,2}, \quad \mathbf{A}_{1,2} = -\mathbf{A}_{2,1}. \quad (20)$$

The objective function is NP-hard with binary constraints and cannot be efficiently solved. We regulate the affine transformation matrix $\mathbf{A}$ as a rigid transformation to prevent unrealistic image warping, and relax the objective function by linearization to convert the original problem into a linear programming (LP) form:

$$\min J = tr(\mathbf{T}^T\mathbf{X}) + \lambda_p \mathbf{1}_2^T \mathbf{U} \mathbf{1}_{n_p} + \lambda \mathbf{1}_2^T \mathbf{V} \mathbf{1}_{n_p} + \lambda_s \sum_{k=1}^{n} w_k \mathbf{1}_2^T \mathbf{M}_k \mathbf{1}_{E_k^P} \quad (21)$$

subject to

$$\begin{aligned} -\mathbf{U} &\leq \mathbf{A}\mathbf{P}^P + \mathbf{b}\mathbf{1}_{n_p}^{\mathrm{T}} + \mathbf{W}\Phi - \mathbf{P}^G\mathbf{X}^{\mathrm{T}} \leq \mathbf{U}, \\ -\mathbf{V} &\leq \mathbf{W}\Phi \leq \mathbf{V}, \\ -\mathbf{M} &\leq \mathbf{A}\hat{\mathbf{Q}} + \mathbf{W}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j) - \hat{\mathbf{Q}}\mathbf{Y}^{\mathrm{T}} \leq \mathbf{M}, \\ \mathbf{U} &\geq 0, \quad \mathbf{V} \geq 0, \ \mathbf{M} \geq 0, \\ \sum_i \mathbf{X}_{ij} &\leq 1, \quad \sum_j \mathbf{X}_{ij} \leq 1, \ \mathbf{X}_{ij} \geq 0, \\ \mathbf{A}_{1,1} &= \mathbf{A}_{2,2}, \quad \mathbf{A}_{1,2} = -\mathbf{A}_{2,1}. \end{aligned} \quad (22)$$

In the above formulation, we relax the binary constraint of $\mathbf{X}_{ij} \in \{0,1\}^{n_p \times n_g}$ to a continuous domain $0 \leq \mathbf{X}_{ij} \leq 1$. $\mathbf{U} \in R^{2\times n_p}$, $\mathbf{V} \in R^{2\times n_p}$ and $\mathbf{M}_k \in R^{2\times E_k^P}$ ($k = 1, 2, \ldots, n$) are auxiliary matrices representing upper bounds of $\| \mathbf{A}\mathbf{P}_{2\times n_p}^P + \mathbf{b}\mathbf{1}_{n_p}^T + \mathbf{W}\Phi_{n_p \times n_p} - \mathbf{P}_{2\times n_g}^G \mathbf{X}_{n_g \times n_p}^T \|_1$, $\| \mathbf{W}\Phi \|_1$ and $\| \mathbf{A}\mathbf{Q}_{2\times E_k^P}^P + \mathbf{W}_{2\times n_p}(\Phi\mathbf{P}_i - \Phi\mathbf{P}_j)_{n_p \times E_k^P} - \mathbf{Q}_{2\times E_k^G}^G \mathbf{Y}_{E_k^G \times E_k^P}^T \|_1$ respectively. We summarize the detailed procedure of TPSM in Algorithm 2.

E. Partial Face Recognition Using TPGM and TPSM

We obtain the average matching cost $\bar{d} = J_{min}/(\sum_{i,j} \mathbf{X}_{ij})$ by dividing the matching cost with the number of matching keypoints, and we compute the distance between the probe and the gallery as follows:

$$d = \frac{\bar{d}}{\sum_{i,j} \mathbf{X}_{ij}} = \frac{J_{min}}{(\sum_{i,j} \mathbf{X}_{ij})^2} = \frac{K_t + \lambda_p K_p + \lambda_q K_q}{(\sum_{i,j} \mathbf{X}_{ij})^2}. \quad (23)$$

In (23), we separate the distance into two parts, the texture matching distance and the graph matching distance.

Algorithm 2 TPSM

Input: Probe image P, Gallery image G, iteration number T and parameters λ_p, λ_x, λ, λ_s, τ, α_1 and α_2

Output: $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$

Detect keypoints and extract SiftSurfSILBP features.
Using Lowe's method to filter the keypoints and construct the graph g^P and g^G.
Initialize $r = r_{init}$ and set the constraint set $\Phi = \varnothing$.
for $t = 1, 2, \cdots, T$ **do**
  Construct Φ
  Add the constraint set Φ to (21) and solve LP to obtain $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$
  Clear the constraint set Φ to $\varnothing$
  Find the outsiders of the candidate region and construct the constraint set Φ
  Shrinkage: $r^{(n+1)} = \alpha_1 \times r^{(n)}$ and $\lambda_2^{(n+1)} = \alpha_2 \times \lambda_2^{(n+1)}$
  Remove $\mathbf{p}_i^P, \mathbf{X}_{i:}$, if $\sum_j \mathbf{X}_{ij} < \tau$
  Remove $\mathbf{p}_j^G, \mathbf{X}_{:j}$, if $\sum_i \mathbf{X}_{ij} < \tau$
end for
Binarize $\mathbf{X}$
return $\mathbf{A}, \mathbf{b}, \mathbf{W}, \mathbf{X}, \mathbf{Y}$

The distance is proportional to the average matching cost which indicates the matching difference, while it is inversely proportional to the number of matching pairs which shows the similarity of local structures in image pairs.

F. Discussion

1) Handling 3D Face Poses: In essence, a face is a 3D object and the yaw change on a 3D face causes the self-occlusion, so that a 3D model is theoretically a more accurate way to represent a face especially in dealing with pose changes. However, the 3D spatial character of a face is not really a major impact in most existing 2D face recognition systems. A variety of 2D face alignment methods have achieved promising results over the past three decades, where a 2D non-rigid transformation was used to approximate 3D pose changes and obtained acceptable results. We also believe that extending this algorithm to 3D space will make more promising results, which is one interesting future direction.

2) Influence of Initial Imposter Matching: The first step of the proposed methods is feature extraction and matching, which may suffer from imposter matching despite a relatively strict threshold of Lowe's method. To deal with this problem, we gradually remove the outliers during iterations, and most imposter matches are removed within only a few iterations in our experiments. Therefore, the outlier removal step guarantees the convergence under initial mismatches.

IV. Experiments

We evaluated the performance of our proposed TPGM and TPSM on four benchmark databases: LFW [23], PubFig [32], AR [52] and EYB [36], where LFW, PubFig and EYB are holistic face datasets and AR is a partial face dataset. We compared the proposed TPGM and TPSM to several state-of-the-art methods on both holistic and partial face recognition tasks, which shows the effectiveness and robustness of the proposed approach.

For LFW and PubFig, we first added random transformation (e.g. random occlusion, rotation) to evaluate the proposed methods on partial face recognition tasks, and then tested on the original holistic facial images. For AR, we first evaluated on the cropped and aligned datasets to make a fair comparison with existing holistic face recognition methods, and then tested on the original AR datasets with variant image sizes and disalignment. For EYB, we followed the experimental settings of [76] by adding arbitrary occlusion ranging from 0% (the original dataset) to 50%.

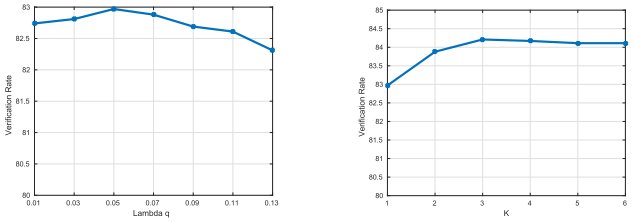
A. Baseline Methods

In the experiments, we compared with 5 baseline algorithms, which include Locally Affine Invariant Robust Point set Matching (LAIRPM) [76], Robust Point Set Matching (RPSM) [76], Metric Learned Extended Robust Point Matching (MLERPM) [75], Multi-Keypoint Descriptors-Sparse Representation-based Classification-Gabor Ternary Pattern (MKD-SRC-GTP) [44] and Coherent Point Drift (CPD) [57]. Among the five baselines, three of them are alignment-free matching methods and the others are the state-of-the-art partial face recognition approaches.

1) Locally Affine Invariant Robust Point Set Matching (LAIRPM) [76]: LAIRPM is based on [41], which is a linear programming framework for feature matching. The neighbourhood is built by k-nearest-neighbour (KNN) with $k = 5$ to achieve the best performance.
2) Robust Point Set Matching (RPSM) [76]: In RPSM, both texture similarity and node affinity are used for alignment-free matching. We set the parameters as $\lambda_p = 0.001, \lambda_x = \min C, \lambda = 5$ to obtain the best performances.
3) Metric Learned Extended Robust Point Matching (MLERPM) [75]: MLERPM utilizes feature set matching to register the extracted local features. The correspondence matrix and transformation parameters are updated during the iteration process. Moreover, it applies the RBF kernel for non-rigid transformation.
4) Multi-Keypoint Descriptors-Sparse Representation-based Classification-Gabor Ternary Pattern (MKD-SRC-GTP) [44]: MKD-SRC-GTP is designed for well-aligned partial face recognition. It applies local MKD-GTP features to describe partial faces and utilize gallery feature sets to represent the images sparsely.
5) Coherent Point Drift (CPD) [57]: CPD considers the alignment of two point sets as a probability density estimation problem. It can be applied to both non-rigid and rigid transformations. In the experiment, we utilized the non-rigid RBF kernel.

B. Results on LFW

The Labeled Face in the Wild (LFW) database [23] consists of 13233 labeled faces of 5749 subjects. The images were



(a) (b)

Fig. 5. Comparison of verification rates (%) under the unsupervised setting of the LFW dataset versus varying (a) λ_q and (b) k.

obtained under unconstrained circumstances with large variations in scale, rotation, illumination and occlusion.

1) Parameter Analysis: We first tested the mean verification rate under different parameters of the proposed TPGM with unsupervised setting on "View 1" dataset in LFW, and then applied these parameters on other experiments including "View 2" of LFW, PubFig, AR and EYB.

We followed [76] by setting $\lambda_p = 0.01$, $\lambda_x = min(C)$ and $\lambda = 5$ for simplicity and a fair comparison. We tested the mean verification rate with different λ_q fixing other parameters, and Fig. 5 (a) shows that the best result was achieved when $\lambda_q = 0.05$. Then, we evaluated the proper highest order k of NEs in TPSM. TPSM would degenerate to TPGM for $k = 1$, and the computational cost would increase with a too large k. Also, as there are only a few candidate keypoints detected on a partial face image, NEs with very high order may not exist. Fig. 5 (b) shows that we should set the highest order k as 3.

2) Partial Face Recognition: We evaluated the proposed method on the "View 2" of the LFW dataset, which included 6000 pairs in total with half of them matched and the other mismatched. They are separated into 10 folds with 600 pairs for each fold. In order to obtain partial face images, we first applied the Viola-Jones face detector to detect and crop the facial regions of all images. Then, we randomly transformed those detected holistic facial images to obtain the arbitrary face patches, where both gallery and probe images were partial faces with some facial components cropped out, making it extremely difficult to match. We added high dimensional LBP (HDLBP) method [7] with face alignment by CFAN facial landmark detector for comparison. In order to conduct full comparisons with the state-of-the-art face recognition methods, we have further evaluated the performance of VGG [67], FaceNet [64], Center Loss [74], Light CNN [79] and SphereFace [46] on the partial LFW dataset.

Table II shows the mean verification rate of our method, HDLBP, CNN methods and baseline methods. Our TPGM outperforms other baseline methods which proves its effectiveness. HDLBP presents the poorest performance on the partial face recognition task, although it achieves average accuracy 84% in LFW with the unsupervised setting of LFW. HDLBP applies CFAN to detect 25 landmarks. However, it does not work on partial faces, especially when some facial components are occluded. Therefore, HDLBP suffers from severe misestimation of facial landmarks, which is the main reason of the performance gap. The proposed methods also outperform the state-of-the-art holistic face recognition methods on the partial faces, because the CNN models are trained on the aligned holistic faces which have limitations to address the variations of the unaligned partial faces.

TABLE II

COMPARISON OF MEAN VERIFICATION ACCURACY AND STANDARD DEVIATIONS (%) ON THE LFW DATASET

Method	$u \pm S_E$
HDLBP	49.32 ± 1.09
CPD-SiftSurfSILBP	61.62 ± 1.19
MKD-SRC-GTP	68.18 ± 1.77
MLERPM-SiftSurf	65.55 ± 1.53
MLERPM-SiftSurfLBP	67.22 ± 1.83
LAIRPM-SiftSurf	70.40 ± 1.02
LAIRPM-SiftSurfSILBP	70.73 ± 1.68
RPSM-SiftSurf	70.81 ± 1.46
RPSM-SiftSurfSILBP	71.65 ± 1.57
VGG	71.27 ± 1.38
FaceNet	71.05 ± 1.90
Center Loss	74.27 ± 1.84
Light CNN	66.77 ± 3.08
SphereFace	69.88 ± 2.71
TPGM-SiftSurfSILBP	72.22 ± 1.05
TPGM-SiftSurfSILBPCNN	73.78 ± 1.19
TPSM-SiftSurfSILBP	73.17 ± 0.94
TPSM-SiftSurfSILBPCNN	$\mathbf{74.33 \pm 1.02}$

TABLE III

COMPARISON OF MEAN VERIFICATION RATE (VR) (%) WITH THE STATE-OF-THE-ART FACE DESCRIPTORS UNDER THE UNSUPERVISED SETTING OF THE STANDARD LFW PROTOCOL

Method	VR
LBP [71]	69.45
SIFT [71]	64.10
LARK [65]	72.23
POEM [72]	75.22
LHS [66]	73.40
MRF-MLBP [2]	80.08
PEM (LBP) [40]	81.10
PEM (SIFT) [40]	81.38
DFD [37]	84.02
HDLBP [7]	84.08
TPGM	82.95
TPSM	**84.40**

3) Holistic Face Recognition: We then evaluated the proposed methods with the unsupervised setting of LFW, following the standard evaluation protocol on the "View 2" dataset Each face image is aligned and cropped into 128 × 128 to remove background information.

Table III tabulates the mean verification rates and Fig. 6 shows the ROC curves of our TPGM and TPSM compared with the state-of-the-art face descriptors with the unsupervised setting of LFW. We see that both TPGM and TPSM achieve comparable results on this popular holistic face recognition database. As keypoint-based partial face recognition methods, although they sacrifice the whole structure of faces to some extent for the robustness reasons, its structural graph still has strong ability to describe the holistic face. Moreover, the proposed methods match two face images directly, while

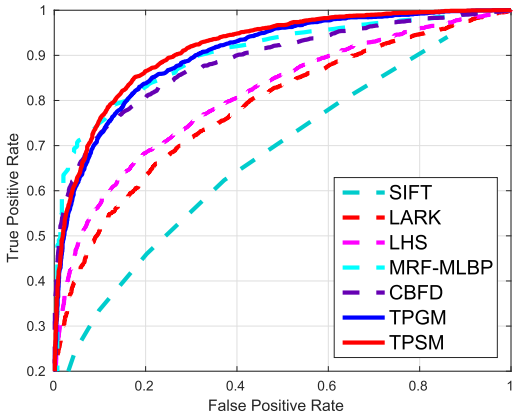


Fig. 6. Comparison of ROC Curves With Different Methods Under the Unsupervised Setting of the LFW Dataset.

TABLE IV
ITERATION NUMBER AND COMPUTATIONAL TIME (S) PER IMAGE PAIR OF RPSM AND THE PROPOSED TPGM

Method	Iteration Number	Computational Time
RPSM	5.8	2.2
TPGM	**1.4**	**0.9**

some learning-based methods need training procedure with extra images. Compared with TPGM, TPSM exploits higher order topological information of holistic faces, which achieves better performance. The proposed methods obtain comparable results on holistic face recognition task and achieve the highest accuracy on partial face recognition, which proves their effectiveness and robustness.

4) Iteration Number and Computational Time: We compared the average iteration number and computational time of RPSM [76] and the proposed TPGM. Our hardware configuration comprises of a 2.8-GHz CPU and a 15G RAM. In each experiment, the time complexity varies with the number of initial matching keypoints, and Table IV shows the mean iteration number and the computational time of the proposed TPGM as well as RPSM on LFW. We observe that TPGM converge faster than RPSM and costs less computational time, because the geometric information accelerates the convergence.

5) Partial Face Detection: The proposed TPGM and TPSM should work on a detected and cropped single face to avoid doubts. In our experiments, we directly processed the cropped holistic faces to obtain partial faces, where we verified whether these processed faces can still be detected by existing face detectors in this subsection.

We randomly selected 16 processed partial faces from the LFW dataset and grouped them into one image for face detection. Figure 7 shows the detection results by the widely-used Fast R-CNN method [17]. We observe that all the partial faces can be correctly detected. In order to further evaluate the performance of face detector under partial faces, we tested Fast R-CNN on two real examples with severe occlusions on faces. Figure 8 shows that existing face detectors successfully detect partial faces.

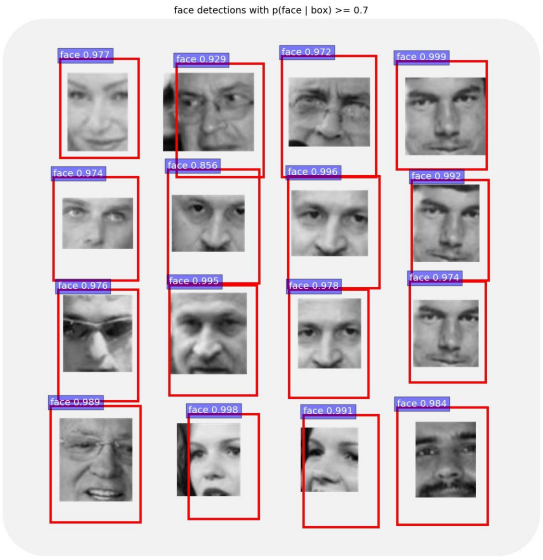


Fig. 7. The Results of Face Detection for the Partial Faces From the LFW Dataset.

C. Results on PubFig

PubFig dataset [32] contains 58797 images of 200 people under unconstrained conditions, with variations in poses, illuminations, expressions and background. We utilized the evaluation sets of 140 people as the test set and the remaining 60 people as the development sets. In order to address the problems of duplication, unavailability and false labels, we applied a selected dataset [76] in which each subject has 5 images. The selected dataset well preserves the variation in poses, illuminations, expressions and background.

1) Partial Face Recognition: We tested on the modified PubFig dataset for partial face recognition to evaluate the effectiveness of the proposed methods. We first randomly transformed the original images to obtain partial face patches. In the transformation process, We first randomly rotated each image with an angle uniformly distributed in [-20,20]. Then, we cropped the rotated image into $h \times w$, where h and w could be as small as 0.8 times of original height and width respectively. Finally, we randomly scaled the cropped image ranging from 0.8 to 1.2. After the transformation, the patches were randomly split into five subsets, and we conducted five-fold testing scheme to test. We added three baseline methods for comparison, including Housdorff distance (HausDist) [25], earth mover's distance (EMD) [63] and Lowe's matching method, which only apply the textural features (SiftSurfSILBP) for image matching.

Table V shows that the proposed TPGM and TPSM outperform other state-of-the-art partial face recognition methods. We observe that HousDist, EMD and Lowe's matching method obtain poor results as they rely on feature descriptors without geometric information. However, the textural hand-crafted features are not robust enough for the unconstrained wild condition.

2) Holistic Face Recognition: We tested our method on holistic face recognition by directly using the selected

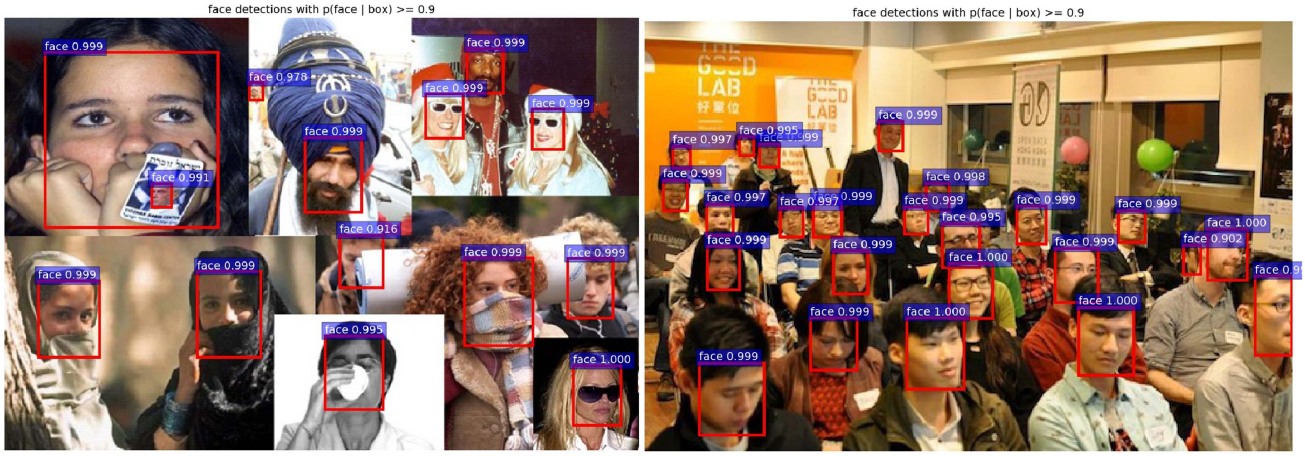


Fig. 8. The Results of Face Detection for the Partial Faces From the Real Pictures, Where the Faces Under Severe Occlusions Can Still be Detected by the Widely-Used Face Detector.

TABLE V
COMPARISON OF AVERAGE RECOGNITION RATES (%) ON THE PUBFIG DATASET

Method	rank = 1	rank = 10	rank = 20
HousDist	0.71	10.71	19.29
EMD	3.57	20.00	30.71
Lowe	25.00	49.29	57.86
CPD	28.36	51.93	62.29
MLERPM	27.86	52.86	64.29
MKD-SRC-GTP	38.57	62.14	72.14
LAIRPM	37.14	64.29	72.86
RPSM	42.86	65.00	74.29
TPGM	44.46	66.84	76.38
TPSM	**46.79**	**67.93**	**78.79**

TABLE VI
COMPARISON OF MEAN VERIFICATION RATES AND STANDARD DEVIATIONS (%) ON THE PUBFIG DATASET

Method	$u \pm S_E$
CPD-SiftSurfSILBP	61.46 ± 3.00
MKD-SRC-GTP	69.82 ± 1.56
MLERPM-SiftSurf	66.18 ± 2.02
MLERPM-SiftSurfLBP	68.64 ± 2.04
LAIRPM-SiftSurf	69.32 ± 1.70
LAIRPM-SiftSurfSILBP	70.00 ± 2.75
RPSM-SiftSurf	73.35 ± 2.40
RPSM-SiftSurfSILBP	78.89 ± 1.75
TPGM-SiftSurfSILBP	80.23 ± 1.16
TPGM-SiftSurfSILBPCNN	81.76 ± 1.43
TPSM-SiftSurfSILBP	82.17 ± 1.03
TPSM-SiftSurfSILBPCNN	$\mathbf{82.98 \pm 1.27}$

evaluation set of PubFig database in our experiments, without alignment or size normalization. We paired up every two images from the five images for each subject to construct 1400 genuine matching pairs. Then, we randomly selected two images from different subjects to construct 1400 imposter matching pairs.

Table VI tabulates the mean verification rates and the corresponding standard deviations of the proposed methods, compared with several alignment-free and partial face recognition baseline methods. We observe that the proposed TPGM and TPSM methods achieve the best results in the PubFig dataset. For the feature matching methods, TPGM and TPSM outperform CPD and MKD-SRC-GTP as they exploit both the textural and geometric information of the facial images, which are robust to illumination and local changes. However, CPD and MKD-SRC-GTP only consider the feature set similarity, which may be affected by large variations. Compared with MLERPM, RPSM and LAIRPM, our TPGM and TPSM exploit higher order geometric information which are more robust and discriminative. Moreover, the SiftSurfSILBPCNN-based approaches have obtained higher verification rates than with SiftSurfSILBP, which shows the effectiveness of the our descriptor.

D. Results on AR

AR database [52] consists of 126 identities including 70 males and 56 females. It contains two sessions and each session has 13 face images for each subject, where 3 of them are taken under various illumination conditions, 4 of them with different expressions, 3 of them wearing sunglasses and 3 wearing scarves.

In order to test the performance of the proposed methods under occlusion scenarios, we conducted several experiments on the AR database. We followed [76], [78], [87] by randomly selecting 50 male subjects and 50 female subjects from the original AR database. We denoted S1-G and S1-S as the images with sunglasses and scarves in session 1, respectively, and S2-G and S2-S in session 2. For each identity, we used a single holistic face image without occlusion as the gallery image, and 12 partial faces as the probe, where half of them are with sunglasses and the other with scarves.

1) Aligned Partial Face Recognition: In order to make a fair comparison with holistic face recognition methods, we cropped all the images into 128×128 and aligned properly. Table VII shows the identification accuracy, where our TPGM and TPSM achieve better performance than other state-of-the-art methods on S1-G and S2-G, and perform comparably to MKD-SRC-GTP on S1-S and S2-S. In addition, the proposed

TABLE VII

COMPARISON OF RECOGNITION ACCURACY (%) ON THE ALIGNED AR DATASET

Method	S1-G	S1-S	S2-G	S2-S
HSO-EBGM [34]	88.20	91.80	**80.00**	83.30
Stringfaces	88.00	96.00	76.00	88.00
CPD	79.33	81.67	51.67	66.67
MLERPM	82.00	85.33	59.67	71.33
LAIRPM	91.00	92.33	72.67	85.33
MKD-SRC-GTP	88.67	**97.33**	70.00	**93.33**
RPSM	93.00	94.33	76.67	86.00
TPGM	93.67	96.33	79.33	89.67
TPSM	**94.33**	96.33	**80.00**	90.00

TABLE VIII

COMPARISON OF RECOGNITION ACCURACY (%) WITH BASELINE ALGORITHMS ON THE ORIGINAL AR DATASET

Method	S1-G	S1-S	S2-G	S2-S
VGG	84.67	85.67	63.33	83.00
CPD	71.00	75.67	49.33	61.00
MLERPM	75.00	78.33	53.33	66.67
LAIRPM	87.33	88.33	56.33	81.33
MKD-SRC-GTP	82.33	83.33	57.67	76.33
RPSM	88.67	90.33	63.67	85.67
TPGM	89.67	91.00	66.00	86.67
TPSM	**90.33**	**91.67**	**67.33**	**87.00**

approaches have a consistently superior performance over RPSM, MLERPM and LAIRPM, showing the effectiveness of the higher order structural information. Through the experiments, we found a key character of our TPGM and TPSM is to robustly detect the outlier keypoints located in occlusion parts and remove them during the iterations, which enhanced the robustness of the methods.

2) Unaligned Partial Face Recognition: We further evaluated the performance of our proposed approaches on the face recognition task where the original AR datasets were directly used as probe images without cropped and alignment. It is ubiquitous in real-life applications that probe images and gallery images are of different size and they may not be properly aligned. Table VIII shows the recognition accuracy of different methods. We observe that our methods obtained the best results, showing the robustness against various scales. The performance of MKD-SRC-GTP decreased severely comparing with the recognition results shown in Table VII. MKD-SRC-GTP utilizes Harris-Laplacian keypoint detector [44] which is sensitive to corners, and a great number of improper keypoints are detected for the probe images containing the hair region full of corner points. However, keypoints in the hair regions are not stable and discriminative enough in the face recognition problem. We also tested the performance of CNN using the pre-trained VGG-16 network [60] where the proposed methods outperforms VGG on this dataset. The reason is that CNN approaches describe the holistic facial images for partial faces, leading to large intra-class variance.

TABLE IX

COMPARISON OF RECOGNITION ACCURACY (%) UNDER DIFFERENT OCCLUSION RATES ON THE EYB DATASET

Method	0%	10%	20%	30%	40%	50%
SRC [78]	100	100	99.80	98.50	90.30	65.30
MKD-SRC	100	100	96.67	96.67	**93.33**	**76.67**
MLERPM	100	100	100	98.30	80.20	30.20
LAIRPM	100	100	100	96.67	66.67	53.33
RPSM	100	100	100	100	73.33	56.67
TPGM	**100**	**100**	**100**	**100**	81.63	68.13
TPSM	**100**	**100**	**100**	**100**	84.87	71.37

E. Results on EYB

The Extended Yale B (EYB) database [36] contains 2414 face images of 38 subjects with various illumination conditions. For each subject, we randomly chose 32 images as training sets and the rest as testing sets. In our experiment, we directly used the cropped EYB database, where the image size of each facial image is normalized to 192×168.

We conducted the experiments of partial face recognition under arbitrary block occlusions on the EYB database. We synthesized images with occlusion rates varying from 0% to 50% by randomly attaching an unrelated image to each probe image, where 0% occlusion represented the original dataset. Table IX shows that our methods outperform other methods before the occlusion rate reaches 40%. However, their performance degrades when the occlusion rate reaches 40%. In spite of this, they perform consistently better than RPSM and LAIRPM at all occlusion rates, showing the benefits of exploiting higher order geometric information to enhance the robustness of the matching method.

F. Discussion

The above experiments suggest the following four observations:

1) Our TPGM and TPSM outperform other state-of-the-art face descriptors on partial face recognition. Compared to conventional face descriptors extracted from holistic faces which are severely affected by occlusions and misalignment, the proposed methods extract and match on robust keypoints to remove the occluded facial parts. Compared to existing keypoint-based methods, the proposed methods exploit higher order topological information to obtain more robust transformation and faster convergence.
2) Our TPGM and TPSM obtain competitive results on holistic face recognition. Although they only utilize the textural and geometric information on the detected keypoints and sacrifice the holistic face information to some extent, its structural graph still has strong ability to describe the holistic faces.
3) The proposed TPSM method further improves the performance of TPGM. By modeling general high order NPs and NEs, TPSM exploits higher order structural information and delivers stronger effectiveness and robustness.

4) Our TPGM and TPSM directly match the face pairs without training procedure and have achieved the state-of-the-art verification rates on partial face datasets, which shows their strong efficiency and generalization ability.

V. Conclusion

In this paper, we have proposed a topology preserving graph matching (TPGM) method for partial face recognition, where a key character is to embrace the affinity of both nodes and edges in the problem of matching. With the geometric graph structure, TPGM obtains more accurate and robust correspondence which leads to a stronger recognition ability. In order to exploit higher order structural information, we have presented a topology preserving structural matching (TPSM) method by modeling and utilizing general high order neighbour points and neighbour edges. The experimental results on four widely used datasets demonstrate the accuracy, robustness and efficiency of the proposed methods. As deep learning is only used to extract keypoint descriptors, it is interesting to apply deep learning to the graph matching procedure to further improve the matching ability.

References

[1] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, Dec. 2006.

[2] S. R. Arashloo and J. Kittler, "Efficient processing of MRFs for unconstrained-pose face recognition," in *Proc. BTAS*, 2013, pp. 1–8.

[3] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)," *Comput. Vis. Image Understand.*, vol. 110, no. 3, pp. 346–359, 2008.

[4] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[5] W. Brendel and S. Todorovic, "Learning spatiotemporal graphs of human activities," in *Proc. ICCV*, 2011, pp. 778–785.

[6] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct. 1993.

[7] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. CVPR*, 2013, pp. 3025–3032.

[8] M. Cho, J. Lee, and K. M. Lee, "Reweighted random walks for graph matching," in *Proc. ECCV*, 2010, pp. 492–505.

[9] T. Cour, P. Srinivasan, and J. Shi, "Balanced graph matching," in *Proc. NIPS*, vol. 19. 2007, p. 313.

[10] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Topology preserving graph matching for partial face recognition," in *Proc. ICME*, 2017, pp. 1494–1499.

[11] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Learning rotation-invariant local binary descriptor," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 3636–3651, Aug. 2017.

[12] Y. Duan, J. Lu, J. Feng, and J. Zhou, "Context-aware local binary feature learning for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.

[13] Y. Duan, J. Lu, Z. Wang, J. Feng, and J. Zhou, "Learning deep binary descriptor with multi-quantization," in *Proc. CVPR*, 2017, pp. 1183–1192.

[14] O. Duchenne, F. Bach, I. S. Kweon, and J. Ponce, "A tensor-based algorithm for high-order graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2383–2395, Dec. 2011.

[15] O. Duchenne, A. Joulin, and J. Ponce, "A graph-matching kernel for object categorization," in *Proc. ICCV*, 2011, pp. 1792–1799.

[16] H. K. Ekenel and R. Stiefelhagen, "Why is facial occlusion a challenging problem?" in *Proc. ICB*, 2009, pp. 299–308.

[17] R. B. Girshick, "Fast R-CNN," in *Proc. ICCV*, 2015, pp. 1440–1448.

[18] S. Gold and A. Rangarajan, "A graduated assignment algorithm for graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 4, pp. 377–388, Apr. 1996.

[19] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. Alvey Vis. Conf.*, vol. 15. 1988, p. 50.

[20] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using Laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, Mar. 2005.

[21] B. Heisele, P. Ho, J. Wu, and T. Poggio, "Face recognition: Component-based versus global approaches," *Comput. Vis. Image Understand.*, vol. 91, nos. 1–2, pp. 6–21, 2003.

[22] K. Hotta, "Robust face recognition under partial occlusion based on support vector machine with local Gaussian summation kernel," *Image Vis. Comput.*, vol. 26, no. 11, pp. 1490–1498, 2008.

[23] 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," Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. 07-49, 2007.

[24] S. U. Hussain, T. Napoléon, and F. Jurie, "Face recognition using local quantized patterns," in *Proc. BMVC*, 2012, pp. 1–11.

[25] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge, "Comparing images using the Hausdorff distance," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 15, no. 9, pp. 850–863, Sep. 1993.

[26] H. Jia and A. M. Martínez, "Support vector machines in face recognition with occlusions," in *Proc. CVPR*, 2009, pp. 136–141.

[27] H. Jiang, M. S. Drew, and Z.-N. Li, "Matching by linear programming and successive convexification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 6, pp. 959–975, Jun. 2007.

[28] H. Jiang, S. X. Yu, and D. R. Martin, "Linear scale and rotation invariant matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 7, pp. 1339–1355, Jul. 2011.

[29] L. Juan and O. Gwun, "A comparison of SIFT, PCA-SIFT and SURF," *Int. J. Image Process.*, vol. 3, no. 4, pp. 143–152, 2009.

[30] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proc. CVPR*, 2004, pp. 506–513.

[31] J. Kim, J. Choi, J. Yi, and M. Turk, "Effective representation using ICA for face recognition robust to local distortion and partial occlusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1977–1981, Dec. 2005.

[32] N. Kumar, A. C. Berg, P. N. Belhumeur, and S. K. Nayar, "Attribute and simile classifiers for face verification," in *Proc. ICCV*, 2009, pp. 365–372.

[33] B. Lahasan, S. L. Lutfi, and R. San-Segundo, "A survey on techniques to handle face recognition challenges: Occlusion, single sample per subject and expression," in *Artificial Intelligence Review*. Dordrecht, The Netherlands: Springer, 2017, pp. 1–31.

[34] B. M. Lahasan, I. Venkat, M. A. Al-Betar, S. L. Lutfi, and P. de Wilde, "Recognizing faces prone to occlusions and common variations using optimal face subgraphs," *Appl. Math. Comput.*, vol. 283, pp. 316–332, Jun. 2016.

[35] J. Lee, M. Cho, and K. M. Lee, "Hyper-graph matching via reweighted random walks," in *Proc. CVPR*, 2011, pp. 1633–1640.

[36] K.-C. Lee, J. Ho, and D. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[37] Z. Lei, M. Pietikäinen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 289–302, Feb. 2014.

[38] M. Leordeanu and M. Hebert, "A spectral technique for correspondence problems using pairwise constraints," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2. Oct. 2005, pp. 1482–1489.

[39] M. Leordeanu, M. Hebert, and R. Sukthankar, "An integer projected fixed point method for graph matching and MAP inference," in *Proc. NIPS*, 2009, pp. 1114–1122.

[40] H. Li, G. Hua, Z. Lin, J. Brandt, and J. Yang, "Probabilistic elastic matching for pose variant face verification," in *Proc. CVPR*, 2013, pp. 3499–3506.

[41] H. Li, X. Huang, and L. He, "Object matching using a locally affine invariant and linear programming techniques," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 2, pp. 411–424, Feb. 2013.

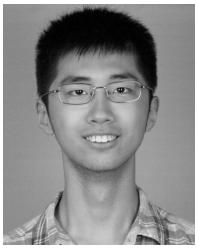
[42] S. Z. Li, X. W. Hou, H. Zhang, and Q. Cheng, "Learning spatially localized, parts-based representation," in *Proc. CVPR*, 2001, pp. 200–207.

[43] Z. Li, G. Liu, Y. Yang, and J. You, "Scale- and rotation-invariant local binary pattern using scale-adaptive texton and subuniform-based circular shift," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2130–2140, Apr. 2012.

[44] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 5, pp. 1193–1205, May 2013.

[45] K. Lin, J. Lu, C.-S. Chen, and J. Zhou, "Learning compact binary descriptors with unsupervised deep neural networks," in *Proc. CVPR*, 2016, pp. 1183–1192.

[46] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "SphereFace: Deep hypersphere embedding for face recognition," in *Proc. CVPR*, 2017, pp. 6738–6746.
[47] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.
[48] J. Lu, V. E. Liong, and J. Zhou, "Simultaneous local binary feature learning and encoding for homogeneous and heterogeneous face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
[49] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 10, pp. 2041–2056, Oct. 2015.
[50] J. Lu, G. Wang, and J. Zhou, "Simultaneous feature and dictionary learning for image set based face recognition," *IEEE Trans. Image Process.*, vol. 26, no. 8, pp. 4042–4054, Aug. 2017.
[51] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger, "Adaptive and generic corner detection based on the accelerated segment test," in *Proc. ECCV*, 2010, pp. 183–196.
[52] A. M. Martínez and R. Benavente, "The AR face database," Comput. Vis. Center (CVC), Univ. Autònama Barcelona, Barcelona, Spain, CVC Tech. Rep. #24, 1998.
[53] A. M. Martínez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.
[54] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.
[55] K. Mikolajczyk *et al.*, "A comparison of affine region detectors," *Int. J. Comput. Vis.*, vol. 65, no. 1, pp. 43–72, 2005.
[56] H. P. Moravec, "Rover visual obstacle avoidance," in *Proc. IJCAI*, 1981, pp. 785–790.
[57] A. Myronenko and X. Song, "Point set registration: Coherent point drift," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 12, pp. 2262–2275, Dec. 2010.
[58] H. J. Oh, K. M. Lee, and S. U. Lee, "Occlusion invariant face recognition using selective local non-negative matrix factorization basis images," *Image Vis. Comput.*, vol. 26, no. 11, pp. 1515–1523, 2008.
[59] K. Pan, S. Liao, Z. Zhang, S. Z. Li, and P. Zhang, "Part-based face recognition using near infrared images," in *Proc. CVPR*, 2007, pp. 1–6.
[60] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," in *Proc. Brit. Mach. Vis. Conf*, vol. 1. 2015, pp. 1–12.
[61] A. Pentland, B. Moghaddam, and T. Starner, "View-based and modular eigenspaces for face recognition," in *Proc. CVPR*, 1994, pp. 84–91.
[62] E. Rosten and T. Drummond, "Machine learning for high-speed corner detection," in *Proc. ECCV*, 2006, pp. 430–443.
[63] Y. Rubner, C. Tomasi, and L. J. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, Nov. 2000.
[64] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. CVPR*, 2015, pp. 815–823.
[65] H. J. Seo and P. Milanfar, "Face verification using the LARK representation," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 4, pp. 1275–1286, Dec. 2011.
[66] G. Sharma, S. ul Hussain, and F. Jurie, "Local higher-order statistics (LHS) for texture categorization and facial analysis," in *Proc. ECCV*, 2012, pp. 1–12.
[67] K. Simonyan and A. Zisserman. (Sep. 2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[68] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. CVPR*, 2014, pp. 1891–1898.
[69] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in *Proc. CVPR*, 2014, pp. 1701–1708.
[70] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. CVPR*, Jun. 1991, pp. 586–591.
[71] R. Verschae, J. Ruiz-Del-Solar, and M. Correa, "Face recognition in unconstrained environments: A comparative study," in *Proc. ECCVW*, 2008, pp. 1–12.
[72] N.-S. Vu and A. Caplier, "Face recognition with patterns of oriented edge magnitudes," in *Proc. ECCV*, 2010, pp. 313–326.
[73] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.
[74] Y. Wen, K. Zhang, Z. Li, and Y. Qiao, "A discriminative feature learning approach for deep face recognition," in *Proc. ECCV*, 2016, pp. 499–515.
[75] R. Weng, J. Lu, J. Hu, G. Yang, and Y.-P. Tan, "Robust feature set matching for partial face recognition," in *Proc. ICCV*, Dec. 2013, pp. 601–608.
[76] R. Weng, J. Lu, and Y.-P. Tan, "Robust point set matching for partial face recognition," *IEEE Trans. Image Process.*, vol. 25, no. 3, pp. 1163–1176, Mar. 2016.
[77] L. Wolf, T. Hassner, and I. Maoz, "Face recognition in unconstrained videos with matched background similarity," in *Proc. CVPR*, 2011, pp. 529–534.
[78] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[79] X. Wu, R. He, and Z. Sun. (Nov. 2015). "A light CNN for deep face representation with noisy labels." [Online]. Available: https://arxiv.org/abs/1511.02683
[80] H. Yan, "Collaborative discriminative multi-metric learning for facial expression recognition in video," *Pattern Recognit.*, vol. 75, pp. 33–40, Mar. 2018.
[81] H. Yan and J. Hu, "Video-based kinship verification using distance metric learning," *Pattern Recognit.*, vol. 75, pp. 15–24, Mar. 2018.
[82] H. Yan, J. Lu, and X. Zhou, "Prototype-based discriminative feature learning for kinship verification," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2535–2545, Nov. 2015.
[83] J. Yan, C. Zhang, H. Zha, W. Liu, X. Yang, and S. M. Chu, "Discrete hyper-graph matching," in *Proc. CVPR*, 2015, pp. 1520–1528.
[84] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.
[85] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.
[86] R. Zass and A. Shashua, "Probabilistic graph and hypergraph matching," in *Proc. CVPR*, 2008, pp. 1–8.
[87] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. ICCV*, 2011, pp. 471–478.
[88] C.-H. Zheng, Y.-F. Hou, and J. Zhang, "Improved sparse representation with low-rank representation for robust face recognition," *Neurocomputing*, vol. 198, pp. 114–124, Jul. 2016.
[89] F. Zhou and F. de la Torre, "Factorized graph matching," in *Proc. CVPR*, 2012, pp. 127–134.
[90] F. Zhou and F. Torre, "Deformable graph matching," in *Proc. CVPR*, 2013, pp. 2922–2929.



Yueqi Duan received the B.S. degree from the Department of Automation, Tsinghua University, China, in 2014, where he is currently pursuing the Ph.D. degree. His current research interests include deep learning, unsupervised learning, and binary representation learning. He has authored/co-authored 14 scientific papers in these areas, where five papers are published as the first author in top journals and conferences, including IEEE TPAMI, TIP, and CVPR. He has obtained the National Scholarship of Tsinghua in 2017.

Jiwen Lu (M'11–SM'15) received the B.Eng. degree in mechanical engineering and the M.Eng. degree in electrical engineering from the Xi'an University of Technology, Xi'an, China, in 2003 and 2006, respectively, and the Ph.D. degree in electrical engineering from Nanyang Technological University, Singapore, in 2012. He is currently an Associate Professor with the Department of Automation, Tsinghua University, Beijing, China. His current research interests include computer vision, pattern recognition, and machine learning. He has authored/co-authored over 180 scientific papers in these areas, where 53 of them are IEEE TRANSACTIONS papers. He is a member of the Information Forensics and Security Technical Committee and the Multimedia Signal Processing Technical Committee of the IEEE Signal Processing Society, and a member of the Multimedia Systems and Applications Technical Committee of the IEEE Circuits and Systems Society. He was a recipient of the National 1000 Young Talents Plan Program in 2015. He is/was a Workshop Chair/Special Session Chair/Area Chair for more than 20 international conferences. He serves/has served as an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, *Pattern Recognition*, *Pattern Recognition Letters*, the *Journal of Visual Communication and Image Representation*, *Neurocomputing*, and IEEE ACCESS.

Jianjiang Feng (M'10) received the B.S. and Ph.D. degrees from the School of Telecommunication Engineering, Beijing University of Posts and Telecommunications, China, in 2000 and 2007, respectively. From 2008 to 2009, he was a Post-Doctoral Researcher with the Pattern Recognition and Image Processing Laboratory, Michigan State University. He is currently an Associate Professor with the Department of Automation, Tsinghua University, Beijing. His research interests include fingerprint recognition and computer vision. He is an Associate Editor of *Image and Vision Computing*.

Jie Zhou (M'01–SM'04) received the B.S. and M.S. degrees from the Department of Mathematics, Nankai University, Tianjin, China, in 1990 and 1992, respectively, and the Ph.D. degree from the Institute of Pattern Recognition and Artificial Intelligence, Huazhong University of Science and Technology, Wuhan, China, in 1995. From 1995 to 1997, he served as a Post-Doctoral Fellow with the Department of Automation, Tsinghua University, Beijing, China, where he has been a Full Professor since 2003. In recent years, he has authored over 100 papers in peer-reviewed journals and conferences. Among them, more than 30 papers have been published in top journals and conferences, such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the IEEE TRANSACTIONS ON IMAGE PROCESSING, and Conference on Computer Vision and Pattern Recognition. His research interests include computer vision, pattern recognition, and image processing. He received the National Outstanding Youth Foundation of China Award. He is an Associate Editor for the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, the *International Journal of Robotics and Automation*, and two other journals.