Partial Face Recognition: An Alignment Free Approach

Shengcai Liao and Anil K. Jain*
Department of Computer Science and Engineering
Michigan State University, East Lansing, MI 48824, U.S.A.
{scliao,jain}@cse.msu.edu

Abstract

Many approaches have been developed for holistic face recognition with impressive performance. However, few studies have addressed the question of how to recognize an arbitrary image patch of a holistic face. In this paper we address this problem of partial face recognition. Partial faces frequently appear in unconstrained image capture environments, particularly when faces are captured by surveillance cameras or handheld devices (e.g., mobile phones). The proposed approach adopts a variable-size description which represents each face with a set of keypoint descriptors. In this way, we argue that a probe face image, holistic or partial, can be sparsely represented by a large dictionary of gallery descriptors. The proposed method is alignment free and we address large-scale face recognition problems by a fast filtering strategy. Experimental results on three public domain face databases (FRGCv2.0, AR, and LFW) show that the proposed method achieves promising results in recognizing both holistic and partial faces.

1. Introduction

Face recognition deals with verifying or identifying a face from its image. It has received substantial attention and its performance has advanced significantly over the last three decades due to its value both in understanding how the face recognition process works in humans as well as in addressing many applications, including access control and video surveillance. While face recognition in controlled conditions (frontal face and uniform illumination) has already achieved satisfactory performance, there still exist many challenges in uncontrolled scenarios, such as non-frontal pose, facial occlusion, and illumination variations.

Typical applications of face recognition in uncontrolled environments include recognition of individuals in video surveillance frames and images captured by handheld devices (e.g., a mobile phone), where a face may be captured in arbitrary pose without user cooperation. In such scenarios, it is quite likely that the captured face image is not a holistic face. These scenarios raise the following important questions: (i) is it possible to recognize a person from a partial image of his face? and (ii) which portion of the face and what size of the partial face are critical for accurate recognition? We call the resulting problem as Partial Face Recognition (PFR), so as to differentiate it from the holistic face recognition problem. State-of-the-art face recognition systems [1, 2] have difficulties with the general PFR problem.

Partial face images result due to many factors, which are categorized in Table 1. Fig. 1(a) shows some examples of partial faces which include a non-frontal face, a face occluded by another face, and a face that is not completely visible in the image frame. Images or videos captured in unconstrained environments often contain partial faces that are difficult to recognize by state-of-the-art face recognition systems [1, 2].

Table 1. Categories of partial faces

<table>
<thead>
<tr>
<th>Category</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>External occlusion</td>
<td>occlusion by other objects</td>
</tr>
<tr>
<td>Self occlusion</td>
<td>non-frontal pose</td>
</tr>
<tr>
<td>Facial accessories</td>
<td>hat, sunglasses, scarf, mask</td>
</tr>
<tr>
<td>Limited field of view</td>
<td>not completely in camera’s FOV</td>
</tr>
<tr>
<td>Extreme illumination</td>
<td>gloomy or highlighted facial area</td>
</tr>
<tr>
<td>Sensor saturation</td>
<td>underexposure or overexposure</td>
</tr>
</tbody>
</table>

*Anil K. Jain is also with the Dept. of Brain and Cognitive Engineering, Korea University, Anam-dong, Seongbuk-gu, Seoul 136-713, Korea.

2http://www.howtovanish.com/2010/01/avoid-nosy-surveillance-cameras/
systems. Therefore, research on partial face recognition is important to advance the state of the art. As an example, a PFR system will enable law enforcement agencies to identify a suspect in a crowd from his partial face captured by a mobile phone and match it to a watch list through a wireless link. Figs. 1(b) and (c) show two scenarios in uncontrolled environments, where (b) many occluded faces exist in a crowd; and (c) two persons are trying to hide their identities by wearing hooded sweatshirt and sunglasses.

To our knowledge, almost all of the current face recognition systems require face alignment. To align a face, the most popular approach is to first detect the two eyes and then normalize the face image geometrically. However, this approach will fail as long as one eye is invisible in a partial face image. Other face alignment methods include Active Shape Model (ASM) [9] and Active Appearance Model (AAM) [8]. They depend on localization of a certain fixed number of landmarks in the holistic face image.

Among various types of partial faces, in practice the occluded and non-frontal ones are the most frequently encountered. While several papers have dealt with occlusion [17, 24, 14, 16], they still require well aligned face images to compensate for the occlusion effect. Ekenel and Stiefelhagen showed that face alignment plays a key role in recognition performance in case of occlusion [10].

Wright et al. introduced sparse representation classification (SRC) scheme for face recognition [34], which was further improved in [32] to compensate for some possible registration errors. Good performance under small amounts of pose and occlusion variations was reported. However, SRC is not applicable to partial face images with large pose variations or limited field of view (see Fig. 1(a)).

Non-frontal face recognition has also attracted significant attention, including multi-view [4, 26] and cross-view face recognition [30, 11, 13, 33]. A critical step in these approaches is to localize a certain fixed number of representative facial landmarks and establish landmark correspondences between the input image and the target image in different views. As a result, both images are required to have visible anchor points. This requirement would not be satisfied in case of external occlusion or limited field of view.

Yi et al. developed a recognition method [35] for frontal, near infrared images with limited field of view. While this method achieved promising matching accuracy, it requires high resolution images with good skin texture, and it has difficulty with pose variations.

Some research on partial face recognition that only require face sub-images as input has been reported. Sato et al. [29] showed that certain facial sub-images (such as eye, nose, and ear) could be used for recognition. Gutta et al. [12] showed that a half (left or right) of the face is sufficient for the recognition task. Park et al. showed that the peripheral region is useful for identification [25]. Nevertheless, these approaches require the presence of expected and predefined facial components in the face image for recognition.

Instead of holistic representation, some face recognition approaches adopted part-based fusion to deal with occlusion and pose variation [6, 26, 13, 22]. These approaches either manually divide the face image into several sub-regions, or automatically detect several predefined components (e.g. eye, nose, and mouth), and fuse results from individual parts for recognition. However, their performance deteriorates when some of the facial portions are invisible and cannot be localized.

For images that are difficult to align (e.g. flowers), bag-of-words (BoW) representation is frequently used for visual object categorization, which applies clustering techniques to build “visual words”. However, histogram of visual words of a partial face may be quite different from that of the whole face. In [20], the BoW representation was applied for face recognition, but it required pre-alignment so that face images could be divided into regular blocks.

In this paper, we present a general formulation of the partial face recognition problem. We do not know a priori whether the input face is holistic or partial. Further, we do not assume the presence of the eyes or any other facial component in the image, and we do not assume any prior or information about face alignment. We provide a general matching solution to all types of partial faces listed in Table 1. Our goal is to recognize an arbitrary partial face without alignment. This is achieved by first deriving a multi keypoint descriptor based representation and constructing a gallery dictionary accordingly. Then multi-task sparse representation is learned from features of each probe face, and recognition is done by the returned sparse coefficients. We call the proposed method as multi keypoint descriptor based sparse representation classification (MKD-SRC).

The various approaches we have reviewed along with the proposed one are summarized in Table 2. The novelties of the proposed approach are: (i) it addresses the general partial face recognition problem without an alignment stage; (ii) by applying multi keypoint descriptors in SRC, it is a unified face recognition framework for both holistic and partial faces; (iii) a fast filtering strategy is proposed to address large-scale face recognition.

2. Proposed Approach

2.1. Gallery Dictionary Construction

To allow for general partial face recognition capability, we construct a large gallery dictionary of multi keypoint descriptors (MKD) as follows. First, we detect multiple SIFT keypoints [21] in each image, and compute the corresponding descriptors. For each class (subject) c in the gallery, suppose k_c keypoints, say, p_{c1}, p_{c2}, \ldots, p_{ck_c}, are detected. Note that if class c has multiple images, we sim-
Table 2. A comparison of partial face recognition approaches along with the proposed one

<table>
<thead>
<tr>
<th>Problem addressed</th>
<th>Approach</th>
<th>Face image requirement</th>
<th>Gallery size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Occlusion</td>
<td>Subspace [17, 24]</td>
<td>Frontal, aligned and cropped</td>
<td>≤ 1, 196</td>
</tr>
<tr>
<td></td>
<td>SRC [34, 32]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>SVM [14, 16]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Part-based fusion [6, 26, 13, 22]</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Single component [29, 12, 25]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pose</td>
<td>Multi-view [4, 26]</td>
<td>Alignment via landmarks</td>
<td>≤ 250</td>
</tr>
<tr>
<td></td>
<td>Cross-view [30, 11, 13, 33]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limited FOV</td>
<td>Skin texture [35]</td>
<td>Frontal, partial face alignment</td>
<td>114</td>
</tr>
<tr>
<td>Occlusion, pose, limited FOV</td>
<td>MKD-SRC (proposed)</td>
<td>Alignment free</td>
<td>&gt; 10,000</td>
</tr>
</tbody>
</table>

We aim to solve a multi-task $l_1$-minimization problem

$$\hat{X} = \arg \min_{X} \sum_{i=1}^{k} \|x_i\|_1, \text{s.t. } Y = DX,$$

where $X = (x_1, x_2, \cdots, x_k) \in \mathbb{R}^{K \times k}$ is the sparse coefficient matrix, and $\|\cdot\|_1$ denotes the $l_1$ norm of a vector. This is equivalent to solving the following $k$ $l_1$-minimization problems for each probe descriptor $y_i$

$$\hat{x}_i = \arg \min_{x_i} \|x_i\|_1, \text{s.t. } y_i = Dx_i, \quad i = 1, 2, \cdots, k.$$

We adopt the following multi-task sparse representation based classification to determine the identity of the probe image, which is inspired from [34]

$$\min_c r_c(Y) = \frac{1}{k} \sum_{i=1}^{k} ||y_i - D_c \delta_c(\hat{x}_i)||^2_2.$$
D, solving Eq. (5) will be computationally challenging in practice. Therefore, we adopt a fast approximate solution. For each probe descriptor $y_i$, we first compute the following linear correlation coefficients between $y_i$ and all the descriptors in the dictionary $D$

$$c_i = D^T y_i, \; i = 1, 2, \cdots, k. \tag{7}$$

Then for each $y_i$, we filter out only $L$ ($L \ll K$) descriptors according to the top $L$ largest values in $c_i$, resulting in a small sub-dictionary $D^{(i)}_{M \times L}$. Next, $D$ is replaced by $D^{(i)}$ in Eq. (5), and Eq. (6) is adjusted accordingly.

As we will see later, this approximate solution is much faster, with no significant degradation in performance. From Eq. (7) and the selection of top $L$ elements (note that this can be done in $O(K\log_2 L)$ by the partial quick sort algorithm) we know that the computation time of the filtering step scales linearly with respect to $K$ (the number of gallery keypoints) and $k$ (the number of probe keypoints), and hence scales almost linearly with respect to the gallery size for each probe image (considering an average number of keypoints per image).

### 2.4. Differences with Related Methods

The differences between the proposed method and three related methods for face recognition are listed in Table 3. Both SIFT [21] and MKD-SRC use variable-size description; they are alignment free and can be used for both holistic and partial face recognition. In contrast, both SRC [34] and LBP-SRC [7] (which uses LBP histograms instead of pixel values in SRC) use a fixed-length description; they require face alignment and cannot be applied to general partial face recognition problem. All SRC based methods exploit gallery class information for face recognition, but SIFT matches each pair of images separately.

The difference among SRC, LBP-SRC and MKD-SRC lies in the feature representation. Since both SRC and LBP-SRC require face alignment and use a single fixed-size feature vector to represent an image, each column of their corresponding dictionaries is related to one gallery image. However, in such a scheme a partial face might have difficulty in alignment and representation due to some unknown missing facial regions. In contrast, MKD-SRC uses a variable-size description; each image is represented by a set of descriptors. The MKD dictionary is composed of a large number of gallery descriptors, making it possible to sparsely represent descriptors from a probe face, regardless of being holistic or partial. Note that although LBP-SRC also extracts local features (LBP histograms) from the image, they are extracted at a certain fixed number of predefined locations after alignment.

Some other existing face recognition approaches based on SIFT include [5, 18, 20, 19], however, all of them depend on pre-aligned face images.

<table>
<thead>
<tr>
<th>Database</th>
<th>FRGCv2.0</th>
<th>AR</th>
<th>LFW</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scenario</td>
<td>partial patch</td>
<td>occlusion</td>
<td>pose &amp; occlusion</td>
</tr>
<tr>
<td>#Gallery</td>
<td>11,398</td>
<td>1,331</td>
<td>20,489</td>
</tr>
<tr>
<td>#Probe</td>
<td>14,630</td>
<td>1,530</td>
<td>2,744</td>
</tr>
<tr>
<td>#Subjects</td>
<td>10,466</td>
<td>1,331</td>
<td>15,749</td>
</tr>
</tbody>
</table>

### 3. Experiments

In this section, we provide experimental results on three public domain databases: FRGCv2.0, AR, and LFW, which are summarized in Table 4. With these three databases we focus on three scenarios of partial face recognition: arbitrary patch of holistic face, occluded face, and face images with arbitrary occlusion and pose variations, respectively. Since in surveillance scenarios, identification is a more reasonable mode than verification, we performed identification experiments on these three databases and used the Cumulative Matching Characteristic (CMC) curve for performance evaluation. We compared the proposed MKD-SRC method to the SIFT matching approach [21], as well as a commercial face recognition software FaceVACS [1] and the original SRC algorithm [34] if they are applicable.

#### 3.1. Partial Face Recognition with Arbitrary Patch

##### 3.1.1 Database

To evaluate the proposed method, we synthetically generated a large database of partial faces. We used 16,028 frontal face images of 466 subjects from the Face Recognition Grand Challenge Ver2.0 (FRGCv2.0) database [28]. The selected images were randomly divided into gallery and probe subsets. The gallery set consisted of 1,398 images of 466 subjects, with 3 images per subject. The remaining 14,630 images composed the probe set. To make the identification problem more challenging, we included an additional 10,000 frontal face images (1 image per subject) from a private database to enlarge the gallery set. Therefore, the gallery set contained 11,398 images in total.

All gallery face images were cropped to $128 \times 128$ pixels according to the two eye coordinates provided. Fig. 3(a) shows some examples of cropped face images. The probe images were first cropped in the same manner as the gallery images. Next, for each cropped probe image, a patch at random position of a random size $h \times w$ was cropped to represent a partial face, where both $h$ and $w$ were uniformly distributed in [50,100]. Fig. 3(b) shows some instances of such randomly cropped face patches. Our objective is to recognize these arbitrary patches (we assume they have been manually cropped, as commonly made in forensic applications), which can be viewed as partial faces with external occlusions or limited FOV. Note that the position information of how these patches were cropped was not known.
Table 3. Differences between the proposed and three related face recognition algorithms

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size of descriptor per image</strong></td>
<td>variable</td>
<td>fixed</td>
<td>fixed</td>
<td>variable</td>
</tr>
<tr>
<td><strong>Face image requirement</strong></td>
<td>alignment free</td>
<td>aligned and cropped</td>
<td>aligned and cropped</td>
<td>alignment free</td>
</tr>
<tr>
<td><strong>Gallery class information</strong></td>
<td>not used</td>
<td>utilized</td>
<td>utilized</td>
<td>utilized</td>
</tr>
<tr>
<td><strong>Holistic face</strong></td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td><strong>Arbitrary partial face</strong></td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>yes</td>
</tr>
</tbody>
</table>

Figure 3. Example face images: (a) gallery images from the FRGCv2.0 database (upper row) and the extended gallery set (bottom row); (b) partial face images cropped from FRGCv2.0.

Table: Differences between the proposed and three related face recognition algorithms

- **Size of descriptor per image**: SIFT [21] is variable, SRC [34] is fixed, LBP-SRC [7] is fixed, and MKD-SRC is variable.
- **Face image requirement**: SIFT is alignment free, SRC is aligned and cropped, LBP-SRC is aligned and cropped, and MKD-SRC is alignment free.
- **Gallery class information**: SIFT is not used, SRC is utilized, LBP-SRC is utilized, and MKD-SRC is utilized.
- **Holistic face**: SIFT and SRC are yes, LBP-SRC is yes, and MKD-SRC is yes.
- **Arbitrary partial face**: SIFT and SRC are yes, LBP-SRC is no, and MKD-SRC is yes.

Next we built an MKD-based dictionary from the gallery images, and also extracted multi keypoint descriptors for each probe image. The MKD dictionary was of size $128 \times 1,181,514$, while the number of descriptors for all probe images was $280,838$. To solve Eq. (5), we utilized the L1 Homotopy MATLAB toolbox [3].

3.1.2 Parameter Selection and Computation Time

Our MATLAB implementation of MKD-SRC runs on a PC server with Intel Core i7 2.93GHz CPU and 16GB memory. There is one parameter $L$ in the proposed algorithm, as described in Section 2.3. We tried different values of $L$ and evaluated the resulting performance in both speed and accuracy, as shown in Fig. 4. We used a subset of the selected data, including 1,398 gallery images and 466 probe images of the 466 subjects. This resulted in $K=111,643$ for the dictionary. Using all the $K$ gallery descriptors (Eq. (7) was not computed), the rank-1 recognition rate is 83.05%, with an average recognition time of 50.4 seconds per probe. From Fig. 4 we can see that $L=100$ and $L=1000$ have a good balance between computation time and accuracy. We decided to set $L=100$ for all the following experiments.

Next we evaluated MKD-SRC on the large-scale dataset (11,398 gallery images and 14,630 probe images). The average number of keypoints per gallery image and per probe image were 104 and 19, respectively. A complete solution of the MKD-SRC framework consists of three steps: keypoint detection, gallery keypoint filtering (by Eq. (7)), and multi-task sparse representation and classification. For a probe image, the average computation times for the first and the last steps are 0.07 and 0.61 seconds, respectively, which are independent of the gallery size. For the second step, it takes 0.80 seconds per probe, on average. As analyzed in Section 2.3, the computation time of the second step scales approximately linearly with respect to the gallery size. In summary, the whole matching process takes 1.48 seconds per probe for a gallery of 11,398 images.

3.1.3 Performance and Comparison

The proposed MKD-SRC algorithm was compared to the SIFT matching approach with the same keypoints and descriptors [21]. The SRC algorithm was not compared because it is not applicable to partial face recognition without prior alignment. Fig. 5 shows the performance of the cumulative matching characteristic (CMC) curves of the two algorithms. It can be seen that the proposed MKD-SRC algorithm significantly improves partial face recognition performance compared to that of the SIFT matching. The rank-1 recognition rate of MKD-SRC is 81.31%, while that of SIFT matching is 58.70%.

Figs. 6 and 7 show examples of some correctly and incorrectly recognized partial faces, respectively, from the FRGCv2.0 database. It appears that the periocular region contains rich information for recognition, while face patch-
3.2. Holistic Face Recognition with occlusion

For a more convincing experimental evaluation and comparison, we performed two additional experiments on holistic faces with staged occlusions and real occlusions. The first experiment was done on the AR database [23] with staged occlusions. The AR database contains 135 subjects with 76 males and 59 females. We selected one image (neutral expression) per class as the gallery set, and 1,530 images, all with sunglasses or scarf, as the probe set. Each probe image may have left-side or right-side illumination. We added 1,196 frontal images from the FERET database [27] to increase the gallery size. All images were cropped to 128 × 128 pixels after face detection. No alignment was done between the probe and gallery sets. Fig. 9 shows some cropped face images from the AR database. It can be seen that they are not aligned very well.

We compared the proposed method to SIFT matching [21], SRC [34], and FaceVACS [1]. The alignment stage was used only in FaceVACS. Fig. 10 shows the CMC curves of these methods on the AR database. These results indicate that MKD-SRC outperforms other algorithms with a notable margin, achieving rank-1 recognition rate of 76.01%. As expected, FaceVACS does not perform very well (49.41% rank-1 accuracy) on these occluded probe images. The rank-1 recognition rate for SIFT and SRC are 36.86% and 8.30%, respectively. Note that this experiment is really challenging because all the probe images have faces with sunglasses or scarf and also contain illumination variations; meanwhile, only one sample per class with neutral expression is available in the gallery set. While Wright et al. demonstrated that their SRC algorithm showed promising performance on well aligned AR faces with sunglasses and scarf, and with sufficient number of training samples [34], our experiments show that the performance of SRC is not good on unaligned faces with only one training sample per class.

3.3. Face Recognition on LFW

While the above experiments were done on synthesized or studio-controlled databases, we performed another experiment on more realistic face images captured in uncontrolled environments: Labeled Faces in the Wild (LFW) [15]. The LFW database was created in an effort to challenge face recognition algorithms in real scenarios. All images in this database were downloaded from the internet. The only condition for including an image in this database was that the face can be detected by the Viola-Jones face detector [31]. The LFW database includes 13,233 images of 5,749 subjects. Face images from LFW contain large variations in pose and illumination, and they might be arbitrarily occluded. Fig. 11 shows some examples from this database.

For our experiments on the LFW database, we select-
ed all the 5,749 subjects with up to 10 images per subject to compose the gallery set, which included 10,489 images. The remaining 2,744 images were selected as the probe set. We also added an additional 10,000 background images to enlarge the gallery set. Note that the original LFW benchmark protocol is for 1-vs-1 verification mode. Here, we conducted an identification experiment to simulate the surveillance scenario. Further, we used unaligned versions of LFW images in our experiments. These images contain extended windows (a ratio of 2.2) output by the Viola-Jones face detector, and re-scaled to \(250 \times 250\) pixels. We used the face detector output (the center part of \(114 \times 114\) pixels) directly for our experiments.

The CMC curves of three face recognition algorithms are depicted in Fig. 12. One can observe that the performance of the proposed MKD-SRC algorithm is similar to that of the state-of-the-art algorithm FaceVACS, and both of them achieve higher recognition rates than SIFT. The rank-1 recognition rates for MKD-SRC, FaceVACS, and SIFT are 56.23%, 54.96%, and 13.99%, respectively. Notice that although the recognition rates of both MKD-SRC and FaceVACS are not very satisfactory for this challenging database, it is promising that MKD-SRC could also deal with non-frontal faces. A sum score fusion of MKD-SRC and FaceVACS, also depicted in Fig. 12, shows an improved rank-1 accuracy of 62.79%. Fig. 13 shows examples of probe images that are correctly and incorrectly recognized by MKD-SRC. While MKD-SRC is able to recognize some non-frontal, occluded, or incorrectly localized faces, there still exist many challenging examples that are hard to recognize.

### 4. Summary and Future Work

We have addressed the problem of recognizing a face from its partial image. We have proposed a novel alignment-free approach, called MKD-SRC to deal with this problem in general. MKD-SRC represents each face image with a set of keypoint descriptors, and constructs a large dictionary from all the gallery descriptors. In this way descriptors of a probe image can be sparsely represented by the dictionary and the identity of the probe face image can be inferred accordingly. We have shown promising results on synthesized partial faces (from the FRGCv2.0 database), occluded holistic faces (in the AR database), and arbitrarily occluded or non-frontal faces (in the LFW database). Comparison with state-of-the-art approaches shows that the MKD-SRC is well suited for the general partial face recognition problem. Our future work would address to improve the description step for a better handling of illumination variations. Furthermore, due to the general framework of MKD-SRC, it may be interesting to also apply MKD-SRC to other image classification areas, such as recognition of biometric modalities other than face, and object categorization.

### Acknowledgement

The authors would like to thank Brendan Klare and Unsang Park for their valuable feedback. Anil K. Jain’s research was partially supported by the WCU (World Class University) program funded by the Ministry of Education, Science and Technology through the National Research Foundation of Korea (R31-10008).

### References


