ROBUST PARTIAL FACE RECOGNITION USING INSTANCE-TO-CLASS DISTANCE

Junlin Hu[1], *Jiwen Lu*[2], *and Yap-Peng Tan*[1]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Advanced Digital Sciences Center, Singapore

ABSTRACT

We present a new face recognition approach from partial face patches by using an instance-to-class distance. While numerous face recognition methods have been proposed over the past two decades, most of them recognize persons from whole face images. In many real world applications, partial faces usually occur in unconstrained scenarios such as visual surveillance systems. Hence, it is very important to recognize an arbitrary facial patch to enhance the intelligence of such systems. In this paper, we develop a robust partial face recognition approach based on local feature representation, where the similarity between each probe patch and gallery face is computed by using the instance-to-class distance with the sparse constraint. Experiments on two popular face datasets are presented to show the efficacy of our proposed method.

Index Terms— Partial face recognition, occluded face, instance-to-class distance.

1. INTRODUCTION

Face recognition has been widely investigated over the past two decades, and a large number of face recognition algorithms have been proposed in the literature [1–7]. State-of-the-art face recognition methods can achieve reasonably good performance under controlled imaging conditions, such as frontal faces, manually aligned images, normal expressions and consistent illuminations. However, these requirements are usually not held in many practical applications because of the ineffectual control over the subjects and the environments, posing great challenges to face recognition in real world scenarios, such as partial occlusions, and large pose variations.

Recently, there have been some efforts [8–11] to address the problem of face recognition in uncontrolled scenarios, especially after the release of the Labeled Face in the Wild (LFW) dataset [8]. Face images in the LFW dataset were taken from Yahoo! News, and show large appearance variations as they were captured in uncontrolled conditions, such as pose, expression, scale, lighting, background, focus, and so on. While state-of-the-art face recognition methods have achieved encouraging performance on this dataset, most of them would degrade or fail to work when face images are not well-aligned or partially occluded. In many real world applications, partial faces or occluded faces usually occurs in unconstrained scenarios such as visual surveillance systems. Hence, it is important to recognize an arbitrary facial patch or occluded face sample (as shown in Fig. 1) to enhance the intelligence of such systems. In this paper, we address the problem of partial face recognition and propose using the instance-to-class distance as the similarity measure for robust face recognition.

Fig. 1. Several partial and occluded face samples. (a) Three partial face patches in the red ellipse are from the LFW database [8]; (b) Holistic faces with scarf and sunglasses occlusion in the AR dataset [12]. The objective of our study in this paper aims to identify people from such partial or occluded face images.

1.1. Related Work

There have been several attempts to address the partial face recognition in recent years. For example, Sato *et al.* [13] introduced utilizing facial patches such as the eye and nose parts in each face image for face representation. Gutta *et al.* [14] used a half face (left or right) for face recognition. While some encouraging results were obtained, facial components used in these methods need to be first manually located and pre-defined, which may require more prior information before the recognition task. More recently, Liao *et al.* [11] proposed an alignment-free face representation method for partial face recognition, where multiple keypoint descriptors were extracted for feature representation and sparse representation based classification (SRC) [15] was used for classification. However, the computational cost of their method is extensive, which may limit its practical applications. Different from their work on face recognition, we extract simple SIFT descriptors for face representation, and present an instance-to-class distance as the similarity measure for partial face recognition task.

Instance to instance distance has been widely used in face recognition, especially for the nearest neighbor (NN) based face matching. However, the performance of this distance degrades heavily if the training samples are not enough. To address this problem, nearest feature line (NFL) [3, 16], nearest feature plane (NFP) [6, 17], and nearest feature space (NFS) [6, 17] methods were proposed to measure the distance between a probe face image and two or more gallery images from the same person for face recognition. These methods represent each face image as a feature vector holistically, and they are usually sensitive to face misalignment. Hence, they can not be directly used for partial face recognition. Unlike these works, Huang *et al.* [18] used an image-to-class distance for face and human gait recognition, where a set of local descriptors are extracted for each image and a spatial neighborhood matching constraint is enforced for feature matching. Similar to NFL, NFP and NFS, however, the performance of their methods decreases when there are spatial misalignments on facial images. Moreover, this distance computation is time consuming. Maturana *et al.* [19] applied the naive bayes nearest neighbor (NBNN) [20] for face recognition against misalignment. One common shortcoming of these methods is that they assume that each image patch of the probe face corresponds to the same part of the face images from the gallery set, therefore they only achieve good performance for frontal and well-aligned face images, and they are not suitable for partial face recognition.

We present an instance-to-class distance with SIFT descriptors for partial face recognition in this paper, where each SIFT descriptor is treated as an instance. The instance-to-class distance computes the distance between a set of local descriptors of one probe face image and their nearest neighbor descriptors of all gallery images belonging to the same class. Our approach is not sensitive to face misalignments. Also the proposed method can perform well on both holistic and partial faces, and it doesn't require any prior knowledge such as the locations of facial components.

The rest of the paper is organized as follows. We details our proposed approach in Section 2, and Section 3 presents experimental results. Finally, Section 4 concludes the paper.

2. PROPOSED APPROACH

In this section, we present our approach to partial face recognition, including partial face representation based on SIFT descriptors and recognition using the instance-to-class distance. The flow chart of our approach is shown in Fig. 2.

2.1. Feature Extraction

Face representation plays an important role on face recognition, and many face representation methods have been proposed over the past decades. Generally, these approaches can be divided into two classes: global methods and local methods. Eigenfaces [21] and Fisherfaces [22] are two representative global methods to holistically represent each face image as a fixed length vector. Since some facial portions and misalignments occurs on partial faces, global features are not suitable to our partial face recognition problem. Local methods such as local binary patterns (LBP) [23] and scale invariant feature transform (SIFT) [24] are more robust to spatial misalignments and more free on the feature length, therefore they can be used to represent partial faces. The reason we choose SIFT features here is that SIFT is more robust to face misalignment and even unaligned face. Fig. 3 shows the SIFT descriptors extracted on both partial and holistic faces. It can be observed that SIFT features can well preserve the similarity between the whole face and the partial patch.

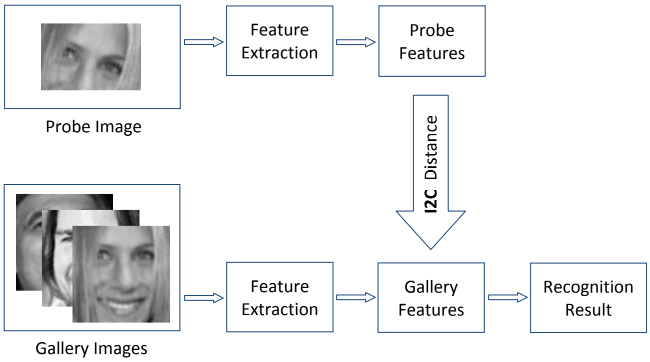

Fig. 2. Flow chart of the proposed partial face recognition using instance-to-class (I2C) distance.

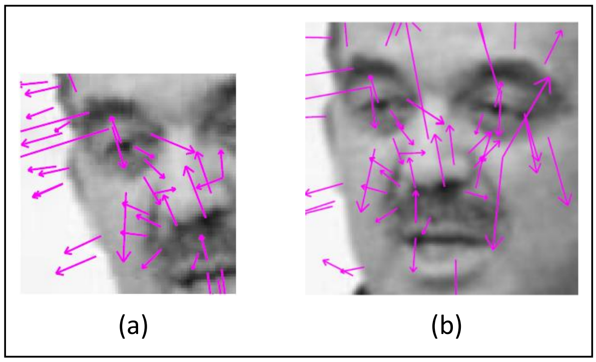

Fig. 3. SIFT descriptors on the two images from the same subject. (a) 36 SIFT descriptors extracted on one partial face; (b) 49 SIFT descriptors extracted on a holistic face.

2.2. Instance-to-Class Distance for Recognition

Let $\mathbf{P} = \{\mathbf{p}_i\}_{i=1}^{N}$ be a set of SIFT feature descriptors extracted from one probe face image $\mathbf{P}$, where $\mathbf{p}_i \in \mathbb{R}^d$, N is the total number of the SIFT descriptors, and d is dimension

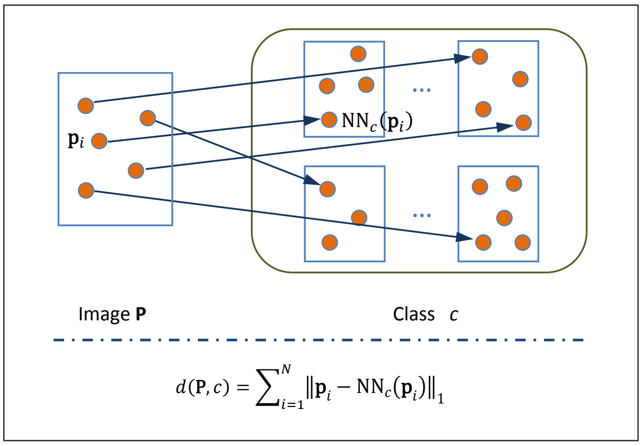

Fig. 4. Toy example of how to compute instance-to-class distance from one probe image to a certain gallery class.

Algorithm 1: Instance-to-Class Distance for Partial Face Recognition

---

Input: A probe partial face image $\mathbf{P}$; The gallery SIFT descriptors set $\mathbf{G} = \{\mathbf{G}_c\}_{c=1}^{C}$; The total number of the class C.

Output: The optimal gallery class c^*.

Extracting SIFT descriptors $\{\mathbf{p}_i\}_{i=1}^{N}$ of the image $\mathbf{P}$.

for *each gallery class* c **do**

  for *each descriptor* $\mathbf{p}_i$ **do**

    $\mathrm{NN}_c(\mathbf{p}_i)$ is the nearest neighbor of descriptor $\mathbf{p}_i$ among the gallery set $\mathbf{G}_c$.

  end

  I2C: $d(\mathbf{P}, c) = \sum_{i=1}^{N} \|\mathbf{p}_i - \mathrm{NN}_c(\mathbf{p}_i)\|_1$.

end

Return: $c^* = \arg\min_c d(\mathbf{P}, c)$.

---

of $\mathbf{p}_i$. Let $\mathbf{G} = \{\mathbf{G}_c\}_{c=1}^{C}$ be gallery SIFT descriptors set from all the gallery classes, where $\mathbf{G}_c = \{\mathbf{g}_j^c\}_{j=1}^{N_c}$ denotes N_c SIFT descriptors extracted from the cth class, and C is the total number of the class.

Our instance-to-class (I2C) distance computes and summarizes all the distances from a set of local feature descriptors of the probe image $\mathbf{P}$ to the corresponding nearest neighbor descriptors of the gallery set $\mathbf{G}_c$. Having computed the I2C distance to all the gallery classes, the nearest neighbor classifier is used to perform the recognition task. Fig. 4 shows a toy example to illustrate how to compute the instance-to-class distance from a probe face image $\mathbf{P}$ to the gallery samples of the cth class, where $\mathrm{NN}_c(\mathbf{p}_i)$ is the nearest neighbor of the descriptor $\mathbf{p}_i$ in the gallery set $\mathbf{G}_c$. The detailed procedure of our method is listed in Algorithm 1, where $d(\mathbf{P}, c) = \sum_{i=1}^{N} \|\mathbf{p}_i - \mathrm{NN}_c(\mathbf{p}_i)\|_1$ is our instance-to-class distance with the sparse constraint used for partial face recognition, and $\|\mathbf{x}\|_1$ means L_1 norm of vector $\mathbf{x}$.

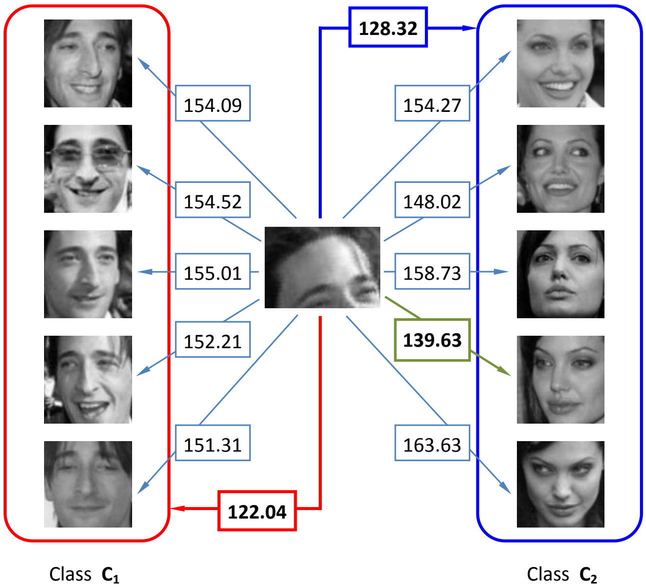

Fig. 5. The performance comparison of patrial face recognition using instance-to-class distance (in rounded rectangles) and instance-to-instance distance on the LFW dataset.

Fig. 5 compares the performance of the instance-to-instance distance and the instance-to-class distance for partial face matching. Given a partial probe face image $\mathbf{P}$ belonging to the class C_1, we find that the instance-to-instance distance between $\mathbf{P}$ and the nearest neighbor image of class C_2 (139.63), is smaller than that from the class C_1 (151.31). Hence, it is wrongly classified to the class C_2 if the instance-to-instance distance is used. However, the instance-to-class distance from $\mathbf{P}$ to the class C_1 (122.04) is smaller than that from $\mathbf{P}$ to the class C_2 (128.32), and we can correctly recognize this sample when using the instance-to-class distance.

3. EXPERIMENTS

3.1. Data Sets

To evaluate the effectiveness of our proposed method, we perform experiments on two widely used datasets: LFW dataset [8] and AR dataset [12].

LFW dataset. The Labeled Face in the Wild (LFW) dataset [8] contains 13233 labeled faces of 5749 people, in which 1680 people have two or more face images. Images in this dataset are taken from the Yahoo! News under the uncontrolled settings, and show large appearance variations such as pose, lighting, scale, background, expression, image resolution, focus, and so on, which poses a great challenge to our recognition task.

AR dataset. The AR dataset [12] contains 126 subjects, including 70 male and 56 female, respectively. For each sub-

Fig. 6. Example face images from the LFW dataset. Blue rectangle: Gallery images; Red rectangle: Partial face images cropped from the LFW dataset.

Fig. 7. Example images in the AR dataset. Top row: gallery images without occlusion, Bottom row: probe images occluded by sunglasses and scarf.

ject, there are 26 face pictures taken in two different sections (each section has 13 face images). Among these face images captured from each section, there are 3 images with different illumination conditions, 4 images with different expressions, and 6 images with different facial disguises (3 images wearing sunglasses and 3 images wearing scarf, respectively).

3.2. Experimental Settings

For the subjects in the LFW dataset, we chose the subjects which have no less than ten images, therefore the 158 subjects are selected in this dataset. For the subjects who have greater than ten face images, the first 10 images were taken in our experiments. These images were converted to gray-scale in our experiments. For each subject, we randomly selected one image for synthetically producing a probe partial face set, while the other 9 images for generating gallery set. For the gallery set, each image was cropped to 128×128 pixels according to the manually labeled eyes positions. Fig. 6 lists several cropped face samples in the gallery set (in Blue rectangle).

To generate partial face images as probe set, we follow the experimental settings in [11]. Firstly, one holistic image was rotated with random angles which follow the gaussian distribution with mean zero and standard deviation $10°$. Then, some partial face images were represented by the face patches with random sizes at random locations. Lastly, these face patches were randomly resized to $height \times width$ pixels, where both $height$ and $width$ fall into the uniformly distribution $[64, 128]$. In our experiments, the probe set contains 6320 partial face images of 158 subjects in all, where each subject has 40 partial images. Fig. 6 shows some partial face images from the probe set (in Red rectangle).

For the AR face database [12], we selected a subset which contains 50 female subjects and 50 male subjects from the first Session of the AR dataset as in [15,25]. For each subject, the first 7 face images (without occlusion) are utilized for training, while 3 images with sunglasses and 3 images with scarf are selected for testing. Thus there are 700 non-occluded face images in the gallery set, and there are 300 face images with sunglasses and 300 face images with scarf respectively in the probe set. The all images were converted to gray-scale and were finally cropped to 128×128 pixels. Fig. 7 shows several cropped images in the AR dataset.

3.3. Results and Analysis

Experiment 1: Partial Face Recognition for Arbitrary Patch. We conducted partial face recognition for arbitrary image patch on the LFW dataset. Table 1 lists the experimental results in this dataset. It can be observed from this table that our method obtains the best performance and outperforms the I2I method by a significant gain, which shows that I2C distance is very powerful for partial face recognition task. In addition, our experimental results also show that our sparse constraint L_1 norm is more better than L_2 norm over two percent in both I2I and I2C distances, this observation is consistent with some existing results on sparse representation framework [15]. Fig. 8 plots the correct recognition rates versus different ranks on the LFW dataset, we can see that partial face recognition is indeed a very challenging work.

Table 1. Partial face recognition accuracy (%) using SIFT descriptors on the LFW dataset.

Method	Accuracy
I2I (L_1)	19.37
I2I (L_2)	17.47
I2C (L_1)	**34.81**
I2C (L_2)	32.28

Experiment 2: Holistic Face Recognition with Occlusion. The AR dataset was selected for our holistic face recognition with occlusion. Table 2 records the recognition accuracy in

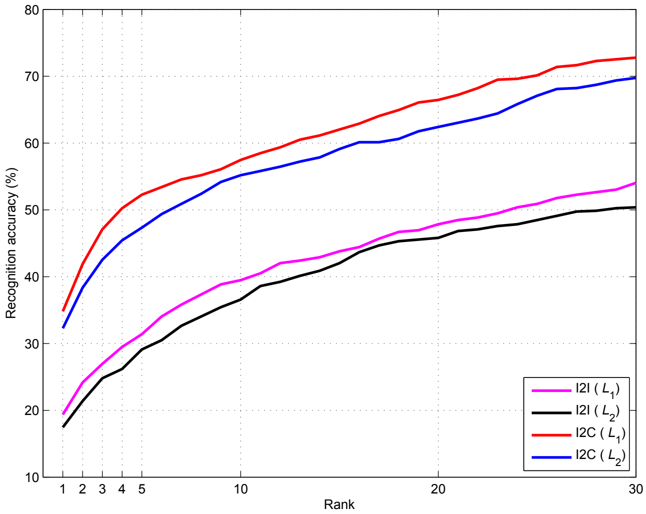

Fig. 8. Partial face recognition rank on the LFW dataset.

Table 2. Performance comparison (%) on the AR dataset with sunglasses and scarf.

Method	Sunglass	Scarf	Sunglass + Scarf
I2I (L_1)	79.67	96.67	88.17
I2I (L_2)	66.67	94.33	80.50
I2C (L_1)	**94.33**	**98.00**	**96.17**
I2C (L_2)	84.00	**98.00**	91.00

the AR dataset with sunglasses, scarf and both, respectively. Our proposed method shows superior performance than the other compared methods. We can see that our I2C approach is better I2I based method by a significant margin. Similarly, Fig. 9 plots the recognition rates versus different rank varying from 1 to 30 on the AR dataset.

Furthermore, for a fair comparison, Table 3 lists some results directly taken from several state-of-the art methods on the AR dataset. These approaches are mainly including collaborative representation based classification (CRC) [25], low-rank recovery (LR) [26], low-rank matrix recovery with structural incoherence [27], robust boltzmann machines (RoBM) [28], robust classification using structured sparse representation (P_{ℓ_2/ℓ_1}) [29], Stringfaces [30], and sparse representation using nonnegative curds and whey (NNCW) [31], and so on.

Experiment 3: Comparison of Different Feature Descriptors. Tables 4 and 5 report performance comparison using SIFT descriptor and LBP descriptor for partial face recognition on the LFW dataset and AR dataset, respectively. We can observe that SIFT descriptor with I2C distance outperforms LBP descriptor with I2C distance by a very significant gain both on the two datasets, these results show that SIFT is more robust to partial face recognition.

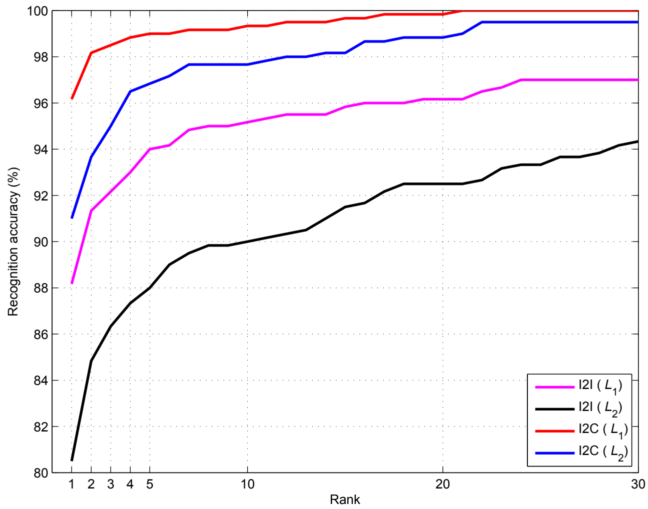

Fig. 9. Partial face recognition rank on the AR dataset with sunglasses and scarf.

Table 3. Performance comparison (%) with some state-of-the-art approaches on the AR dataset.

Method	Sunglass	Scarf	Sunglass + Scarf
SRC [25]	87.00	59.50	73.25
CRC [25]	68.50	90.50	79.50
LR [26]	84.58	77.00	78.92
Chen *et al.* [27]	85.42	84.36	81.62
RoBM [28]	84.50	80.70	82.60
P_{ℓ_2/ℓ_1} [29]	80.50	59.80	70.20
Stringfaces [30]	88.00	96.00	92.00
NNCW [31]	88.44	62.19	75.32
ℓ_1-ℓ_{struct} [32]	92.50	69.00	80.80
I2C (L_1)	**94.33**	**98.00**	**96.17**

4. CONCLUSION

In this paper, we propose a new face recognition method from partial face patches by using an instance-to-class distance. Partial faces usually occurs in unconstrained scenarios such as visual surveillance systems, hence, it is very important to recognize an arbitrary facial patch to enhance the intelligence of such systems. We develop a robust partial face recognition based on local feature representation, where the similarity between each probe patch and gallery face is computed by using the instance-to-class distance with the sparse constraint. Experimental results on two widely used face datasets are presented to show the efficacy of our proposed method.

5. ACKNOWLEDGEMENT

Jiwen Lu is partially supported by the research grant for the Human Sixth Sense Program at the Advanced Digital Sci-

Table 4. Performance comparison (%) using SIFT and LBP respectively on the LFW dataset.

Method	SIFT	LBP
I2C (L_1)	$\mathbf{34.81}$	1.52
I2C (L_2)	$\mathbf{32.28}$	1.39

Table 5. Performance comparison (%) using SIFT and LBP respectively on the AR dataset with sunglasses and scarf.

Method	SIFT	LBP
I2C (L_1)	$\mathbf{96.17}$	16.50
I2C (L_2)	$\mathbf{91.00}$	12.00

ences Center (ADSC) from the Agency for Science, Technology and Research (A*STAR) of Singapore.

6. REFERENCES

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[2] J. Lu and Y.-P. Tan, "A doubly weighted approach for appearance-based subspace learning methods," *IEEE Transactions on Information Forensics and Security,*, vol. 5, no. 1, pp. 71–81, 2010.

[3] J. Lu and Y.-P. Tan, "Uncorrelated discriminant nearest feature line analysis for face recognition," *IEEE Signal Processing Letters*, vol. 17, no. 2, pp. 185–188, 2010.

[4] J. Lu and Y.-P. Tan, "Regularized locality preserving projections and its extensions for face recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 40, no. 3, pp. 958–963, 2010.

[5] J. Lu and Y.-P. Tan, "Cost-sensitive subspace learning for face recognition," in *IEEE International Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2661–2666.

[6] J. Lu and Y.-P. Tan, "Nearest feature space analysis for classification," *IEEE Signal Processing Letters,*, vol. 18, no. 1, pp. 55–58, 2011.

[7] J. Lu, Y.-P. Tan, and G. Wang, "Discriminative multimanifold analysis for face recognition from a single training sample per person," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 39–51, 2013.

[8] 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," Tech. Rep. 07-49, University of Massachusetts, Amherst, October 2007.

[9] L. Wolf, T. Hassner, and Y. Taigman, "Descriptor based methods in the wild," in *Real-Life Images workshop at the European Conference on Computer Vision (ECCV)*, October 2008.

[10] Z. Cao, Q. Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2707–2714.

[11] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. PrePrints, 2012.

[12] A. M. Martinez and R. Benavente, "The ar face database," Tech. Rep. 24, CVC Technical Report, June 1998.

[13] K. Sato, S. Shah, and J. K. Aggarwal, "Partial face recognition using radial basis function networks," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 288–293.

[14] S. Gutta, V. Philomin, and M. Trajkovic, "An investigation into the use of partial-faces for face recognition," in *IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 33–38.

[15] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[16] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439–443, 1999.

[17] J.-T. Chien and C.-C. Wu, "Discriminant waveletfaces and nearest feature classifiers for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 12, pp. 1644–1649, 2002.

[18] Y. Huang, D. Xu, and T. Cham, "Face and human gait recognition using image-to-class distance," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 20, no. 3, pp. 431–438, 2010.

[19] D. Maturana, D. Mery, and A. Soto, "Face recognition with local binary patterns, spatial pyramid histograms and naive bayes nearest neighbor classification," in *International Conference Of The Chilean Computer Science Society*, 2009, pp. 125–132.

[20] O. Boiman, E. Shechtman, and M. Irani, "In defense of nearest-neighbor based image classification," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008.

[21] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[22] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, July 1997.

[23] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face description with local binary patterns: Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[25] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *IEEE International Conference on Computer Vision*, 2011, pp. 471–478.

[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *Journal of ACM*, vol. 58, no. 3, pp. 11, 2011.

[27] C.-F. Chen, C.-P. Wei, and Y.-C. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2618–2625.

[28] Y. Tang, R. Salakhutdinov, and G. E. Hinton, "Robust boltzmann machines for recognition and denoising," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2264–2271.

[29] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1873–1879.

[30] W. Chen and Y. Gao, "Recognizing partially occluded faces from a single sample per class using string-based matching," in *European Conference on Computer Vision*, 2010, pp. 496–509.

[31] Y. Liu, F. Wu, Z. Zhang, Y. Zhuang, and S. Yan, "Sparse representation using nonnegative curds and whey," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3578–3585.

[32] K. Jia, T.-H. Chan, and Y. Ma, "Robust and practical face recognition via structured sparsity," in *European Conference on Computer Vision*, 2012, pp. 331–344.