Fixation and Saccade based Face Recognition from Single Image per Person with Various Occlusions and Expressions

Xingjie Wei and Chang-Tsun Li
Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK
{x.wei, c-t.li}@warwick.ac.uk http://warwick.ac.uk/xwei

Abstract

Face recognition technique is widely used in the real-world applications over the past decade. Different from other biometric traits such as fingerprint and iris, face is the biological nature for humans to recognise a person even met just once. In this paper, we propose a novel method, which simulates the mechanism of fixations and saccades in human visual perception, to handle the face recognition from single image per person problem. Our method is robust to the local deformations of the face (i.e., expression changes and occlusions). Especially for the occlusion related problems, which have not received enough attentions compared with other challenging variations of illumination, expression and pose, our method significantly outperforms the state-of-the-art approaches despite various types of occlusions. Experimental results on the FRGC and the AR databases confirm the effectiveness of our method.

1. Introduction and related work

1.1. Face recognition by computers

Face recognition technique is widely used in biometrics applications in the areas of access control, law enforcement, surveillance and so on. In the unconstrained environment, the appearance of a face is prone to be distorted by the variations such as partial occlusions and expression changes, which could result in poor recognition performance. Different from other challenging conditions such as illumination changes, which affect the facial appearance globally, occlusions and expressions usually distort the face locally. The sources of occlusions (e.g., sunglasses, scarf, hand) are variable and unpredictable, and the distortion due to expressions is non-linear. In the real-world environment, it is difficult to explicitly model these local deformations since no prior knowledge is available. A large number of works have been proposed to deal with the occlusion and expression variations over the past decade[18, 7, 10, 26, 14, 15]. One promising example is the sparse representation-based classification (SRC) method[29], which represents a test image as a linear combination of training images aided with an occlusion dictionary. This kind of method requires sufficient training samples per person to reconstruct the occluded images.

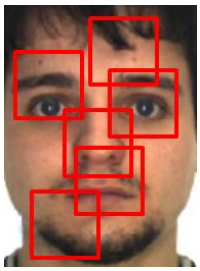


Figure 1. Illustration of fixations and saccades in human visual perception for a face. The red boxes indicate the fixations.

However, in the real-world application scenarios, usually only one image is available per person. The performance of traditional learning based methods[25, 3] will suffer because the training samples are limited. Many approaches are proposed to handle this *single sample per person* (SSPP) problem[24]. SOM[23] creates a suitable self-organising map from images for representing each person. Partial distance (PD)[22] uses non-metric partial similarity measure for matching with a similarity threshold which is learned from the training set. Discriminative multi-manifold analysis (DMMA)[16] formulates the SSPP face recognition as a manifold-to-manifold matching problem by learning multiple feature spaces to maximize the manifold margins of different persons. These methods are more or less model-based. The thresholds or parameters in the model are trained on a representative data set.

1.2. Face recognition by humans

Compared with other biometric traits such as fingerprint and iris, face is the biological nature for humans to recognise each other. Face recognition receives research interests from not only computer scientists, but also neuroscientists and psychologists. When observing an object (e.g.,

face), humans will focus on a small part (i.e., *fixation*) of the whole object and quickly jump between them (the rapid movement of eyes is called *saccade*)[2]. This is shown in Fig. 1. In addition, when humans recognising a face, the features are locally sampled by fixations but the whole facial structure is also considered[21]. Face recognition by humans is dependent on the process involving both featural and structural information.

Recently we proposed a Dynamic *Image-to-Class* Warping (DICW) method in our previous works[27, 28]. It has demonstrated good performance in robust face recognition. A face consists of the forehead, eyes, nose, mouth and chin in a natural order and this order does not change despite occlusions or expressions. DICW represents a face image as a patch sequence which contains the facial order. It employs both the local (i.e., patch-based features) and the global (i.e. facial order) information, which is compatible with the process of face recognition by humans.

In this paper, motivated by the human visual perception, we propose a scheme to improve DICW by simulating the mechanism of fixations and saccades. Like DICW, our method does not require a training phase and can be applied to face recognition from single image per person. In addition, our method is more robust than DICW to occlusions and expression changes.

The rest of this paper is organized as follows. Sec. 2 briefly introduces the Dynamic *Image-to-Class* Warping (DICW) algorithm. Sec. 3 explains our method in details. To evaluate the effectiveness of our method, extensive experiments are conducted on the FRGC and the AR databases and the results are reported in Sec. 4. Finally, Sec. 5 concludes the paper.

2. Dynamic image-to-class warping

We first briefly introduce the Dynamic *Image-to-Class* Warping (DICW) algorithm. A face image is first partitioned into non-overlapping sub-patches which are then concatenated in the raster scan order to form a sequence. In this way, the facial order (i.e., from the forehead, eyes, nose and mouth to the chin) information is encoded in each sequence. DICW calculates the distance between a query sequence and a set of reference sequences of an enrolled class (person). Then the nearest neighbour (NN) classifier is used for classification.

A probe image which is partitioned into M patches is denoted by $\boldsymbol{P} = \{\boldsymbol{p}_1, ..., \boldsymbol{p}_m, ..., \boldsymbol{p}_M\}$ where $\boldsymbol{p}_m$ is the feature vector extracted from the m-th patch. The gallery set of a certain class containing K images is denoted by $\boldsymbol{G} = \{\boldsymbol{G}_1, ..., \boldsymbol{G}_k, ..., \boldsymbol{G}_K\}$. Each gallery image is also represented as a sequence of M patches as $\boldsymbol{G}_k = \{\boldsymbol{g}_{k1}, ..., \boldsymbol{g}_{km}, ..., \boldsymbol{g}_{kM}\}$. A warping path $\boldsymbol{W}$ which indicates the matching correspondence of patches between $\boldsymbol{P}$ and $\boldsymbol{G}$ by T steps is defined as $\boldsymbol{W} = \{w_1, ..., w_t, ..., w_T\}$. The t-th element $w_t = (m_t, m'_t, k_t) \in \{1, 2, ..., M\} \times \{1, 2, ..., M\} \times \{1, 2, ..., K\}$ indicates that patch $\boldsymbol{p}_m$ is matched to patch $\boldsymbol{g}_{km'}$ at step t where $\times$ indicates the Cartesian product operation. $\boldsymbol{W}$ satisfies the following three constraints[27] which maintain the order information. It can be found that the number of matching steps T satisfies $M \leqslant T \leqslant 2M - 1$.

1. **Boundary**: $m_1 = m'_1 = 1$ and $m_T = m'_T = M$.
2. **Continuity**: $m_t - m_{t-1} \leqslant 1$ and $m'_t - m'_{t-1} \leqslant 1$.
3. **Monotonicity**: $m_{t-1} \leqslant m_t$ and $m'_{t-1} \leqslant m'_t$.

Let $C_{w_t} = C_{m_t, m'_t, k_t} = \|\boldsymbol{p}_m - \boldsymbol{g}_{km'}\|_2$ be the local cost between two patches $\boldsymbol{p}_m$ and $\boldsymbol{g}_{km'}$. The *Image-to-Class* distance is the cost of the optimal warping path which has the minimal overall cost:

$$DICW(\boldsymbol{P}, \boldsymbol{G}) = \min_{w_t \in \boldsymbol{W}} \sum_{t=1}^{T} C_{w_t} \tag{1}$$

Eq.(1) can be solved through the *Dynamic Programming* (DP) by creating a 3-D cumulative distance matrix $\boldsymbol{D}$. Each element $D_{m,m',k}$ is recursively calculated based on the results of predecessors (i.e., sub-problems):

$$D_{m,m',k} = \min \begin{pmatrix} \boldsymbol{D}_{\{(m-1,m'-1)\} \times \{1,2,...,K\}} \\ \boldsymbol{D}_{\{(m-1,m')\} \times \{1,2,...,K\}} \\ \boldsymbol{D}_{\{(m,m'-1)\} \times \{1,2,...,K\}} \end{pmatrix} + C_{m,m',k} \tag{2}$$

where the initial state is $\boldsymbol{D}_{0,0,\cdot} = 0, \boldsymbol{D}_{0,m',\cdot} = \boldsymbol{D}_{m,0,\cdot} = \infty$. The final *Image-to-Class* distance is the minimal element of vector $\boldsymbol{D}_{M,M,\cdot}$ as:

$$DICW(\boldsymbol{P}, \boldsymbol{G}) = \min_{k \in \{1,2,...,K\}} \boldsymbol{D}_{M,M,k} \tag{3}$$

Then the probe image $\boldsymbol{P}$ is classified as the class with the shortest distance. Different from patch-wise matching, DICW tries every possible warping path by DP then selects the one with minimal overall cost. So the warping path which involves large distance errors will not be selected.

The time complexity of DICW is $\boldsymbol{O}(M^2K)$[27]. When only one sample is available per person ($K = 1$), the *Image-to-Class* distance degenerates into the *Image-to-Image* distance and the time complexity is $\boldsymbol{O}(M^2)$.

3. Fixations and saccades based face recognition

Our new method relies on the three key observations:

- The structure of the face (i.e., the spatial relationship between facial features) is very important in recognition by humans[21].

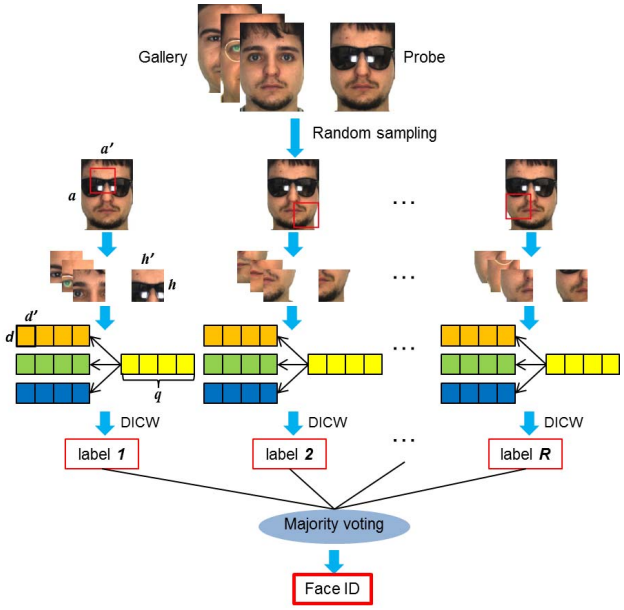


Figure 2. The framework of the proposed method.

- Humans scan a series of *fixations* instead of the whole face when performing recognition[2]. In addition, considering the local deformations due to occlusions and expressions, these affected areas are not helpful for recognition. Since the locations of deformations are unpredictable, random sampling is a good choice. The similar idea is also employed in other pattern recognition applications[9, 13].

- One fixation may not be sufficient for recognition, but a large field of random selections will be likely to produce a good output. This is what is called the *law of large numbers* (LLN).

Inspired by the mechanism of fixations and saccades in human visual perception, we propose a novel method based on DICW. Firstly, a number of R fixations, $\{\boldsymbol{x}^1, ..., \boldsymbol{x}^r, ..., \boldsymbol{x}^R\}$, are randomly sampled from a face. Each fixation (size of $e = h \times h'$ pixels) is also partitioned into q (i.e., set $M = q$ in Sec. 2) patches (size of $s = d \times d'$ pixels) and then forms a sequence which maintains the facial order. Then each fixation sequence is compared with the fixation sequences from the corresponding area of enrolled face images by DICW. We define a binary function $f(r, l)$ for recording the voting result for each fixation as:

$$f(r,l) = \begin{cases} 1 & if \quad class(\boldsymbol{x}^r) = l \\ 0 & otherwise \end{cases} \tag{4}$$

where $l \in [1, 2, ..., L]$ and L is the number of classes. $class(\boldsymbol{x}^r)$ is the label of fixation $\boldsymbol{x}^r$ which is obtained by the NN classifier according to the DICW distance. Even just only one image is available per person, the final classification decision can be made by the majority voting of a large number of fixations:

$$assign \quad \boldsymbol{P} \rightarrow class\, l \quad if \quad \sum_{r=1}^{R} f(r,l) = \max_{i=1}^{L} \sum_{r=1}^{R} f(r,i) \tag{5}$$

where $\boldsymbol{P}$ is the probe image. Here each fixation has the possibilities to classify the face correctly or wrongly. The correct recognition rate is the probability of the consensus being correct. The combined decision is wrong only if a majority of the fixations votes are wrong and they all make the *same* misclassification. But this does not often happen due to the large number of different possible misclassifications.

The framework of our method is shown in Fig. 2. As mentioned in Sec. 2, the time complexity of DICW with single sample per person is $\boldsymbol{O}(M^2)$ (here $M = q$) so the time complexity of our method is $\boldsymbol{O}(R(\frac{e}{s})^2)$ where $q = \frac{e}{s}$.

4. Experimental results

In this section, two well-known face databases (FRGC[19] and AR[17]) are used to evaluate the effectiveness of the proposed method. The FRGC database contains 44,832 still images of 688 subjects with different illuminations and facial expressions. The AR database contains more than 4000 frontal view images of 126 subjects with different expressions, illumination conditions and occlusions. It is one of the very few databases which contain real disguises.

Experiments of face recognition with different occlusions (e.g., randomly located squares, sunglasses, scarves) and expressions (e.g., smile, anger, scream) are conducted on these public databases. Images are cropped and re-sized to 80×65 pixels in the FRGC database and 83×60 pixels in the AR database. For feature extraction, we use LBP ($LBP^{u2}_{8,2}$) descriptor[1], which is insensitive to illumination changes and robust to small misalignment.

We quantitatively compare our method with the methods mentioned in Sec. 1 as well as some methods based on similar ideas as ours. We set the fixation size e to 3% of an image, and use 300 fixations ($R = 300$) which will be about 1.5 minutes of viewing time for a face assuming 3 fixations per second[13]. We follow the settings in [27] and use patch size of $s = 6 \times 5$ pixels in the FRGC database and $s = 5 \times 5$ pixels in the AR database for each fixation. In all experiments, we run our method 10 times and report the average recognition rate.

4.1. Face recognition with randomly located occlusions

We first evaluate the proposed method using the FRGC database. Similar as the work in [12], a set of images of 100

Figure 3. a) Images from the FRGC database with randomly located occlusions. b) Images from the AR database with occlusions. c) Images from the AR database with different expressions.

subjects, which are taken in two sessions, are used in our experiments. For each subject, we choose 1 image as the gallery set and 4 images as the probe set. In order to simulate the contiguous occlusion, we replace a randomly located square (size from 0% to 50% of the image) from each test image with a black patch. Notice that the location of occlusion is randomly chosen and unknown to the algorithm. As shown in Fig. 3a, in some cases most salient facial features are occluded, which is very challenging for recognition. Our aim here is to recognise an occluded face based on the unaffected parts from single image rather than to reconstruct the face from occlusions, so the face with more than 50% occluded area is not discussed in this work. In that case, the unaffected parts are too small to recognise even for humans.

Table 1. Recog. rates (%) on the FRGC database

Occlusion	0%	10%	20%	30%	40%	50%
LSVM[4]	69.5	65.8	57.3	36.8	36.8	22.3
SRC-block[29]	65.8	55.8	47.8	39.8	32.5	22.8
DICW[27]	79.3	77.8	77.3	72.8	70.8	64.5
Ours	**84.2**	**82.4**	**80.3**	**78.1**	**73.2**	**69.6**

There are 6 probe sets, each corresponding to a different level of occlusion (i.e., 0%, 10%, 20%, 30%, 40% 50%). The recognition results are shown in Tab. 1. As expected, the recognition rates decrease when the level of occlusion increases. Our new method outperforms DICW[27] by about 5% and other methods such as the reconstruction based SRC[29] (using 4×2 block partitioning for performance improvement) and the baseline linear support vector machine(LSVM)[4]. Even half of the face is occluded, our method still archives nearly 70% accuracy, which is much better than other methods.

4.2. Face recognition with real disguise

Next we investigate the robustness of the proposed method using partially occluded faces. Similar as the works in [29, 5, 23, 22, 12, 11], in the experiments we choose a subset (50 male and 50 female subjects) of the AR database. For each subject, the neural expression face from session 1 is selected as the gallery set. The faces with sunglasses and scarf from both sessions are used as the probe sets (Fig. 3b).

The comparison of the recognition results between our method and other state-of-the-art methods are provided in Tab. 2. We use our own implementations of SRC and DICW, and the recognition rates of other methods are cited from their papers following the same experimental settings. The performance of DICW is improved by our scheme, especially for the scarf set of session 2 (the most difficult set for other approaches), from 81.0% to 94.9%. Stringface[5] here represents a face as a string of line segments, which also maintains the structural information of a face as DICW and the proposed method. FARO is also a patch-based method as ours but is based on the partitioned iterated function system (PIFSs)[6]. SRC-block, PD and SOM were introduced earlier. In $\text{PWCM}_{0.5}$[11], an occlusion mask is trained through the use of the skin colour. Our method clearly outperforms these approaches without any data-dependent training. CTSDP is a 2D warping method which is also model-free, like ours. Its performance is improved by learning a suitable occlusion handling threshold on occluded images. The overall recognition rate of our method (96.6%) without occlusion pre-processing is very close to that of CTSDP (98.5%) with the occlusion threshold.

Table 2. Rec. rates (%) on the AR database (occlusion)

Method	Session 1		Session 2		Avg.	M/H[1]
	Sung.	Scarf	Sung.	Scarf		
Stringface[5]	88.0	96.0	76.0	88.0	87.0	No
FARO[6]	90.0	85.0	-	-	87.5	No
SRC-block[29]	86.0	87.0	49.0	70.0	73.0	No
PD[22]	98.0	90.0	-	-	94.0	No
SOM[23]	97.0	95.0	60.0	52.0	76.0	No
CTSDP[8]	-	-	-	-	90.6	No
DICW[27]	**99.0**	97.0	93.0	81.0	92.5	No
$\text{PWCM}_{0.5}$[11]	97.0	94.0	72.0	71.0	83.5	Yes
CTSDP[8]	-	-	-	-	**98.5**	Yes
Ours	**99.0**	**98.7**	**93.7**	**94.9**	96.6	No

[1] Occlusion mask/threshold training required

4.3. Face recognition with various expressions

We also evaluate the effectiveness of our method using images with expression changes in the AR database. We use the same gallery set as in Sec. 4.2 and the images from both sessions with smile, anger and scream expressions as the probe sets (Fig. 3c). The Recognition results are shown in Tab. 3. We use our own implementations of SRC and DICW, and the recognition rates of other methods are cited from their papers following the same experimental settings.

Our method outperforms the 7 listed approaches in most cases. The scream expression causes large deformations of

the face. The overall performance of all approaches on the scream sets (especially from session 2) is relatively low due to its challenging nature. On the other hand, our method achieves comparable recognition rates with the 2D warping based method CTSDP. Notice that the time complexity of CTSDP is $\boldsymbol{O}(i^2)$[20] where $i = a \times a'$ (pixels) is the size of the image, compared with ours is just $\boldsymbol{O}(R(\frac{e}{s})^2)$. Here R, the number of fixations, can be viewed as a constant. e is the fixation size and s is the size of the patch in a fixation sequence. Generally $e \ll i$ and $s > 1$. So our method achieves comparable performance in most cases but is more efficient than CTSDP.

Table 3. Rec. rate (%) on the AR database (expression)

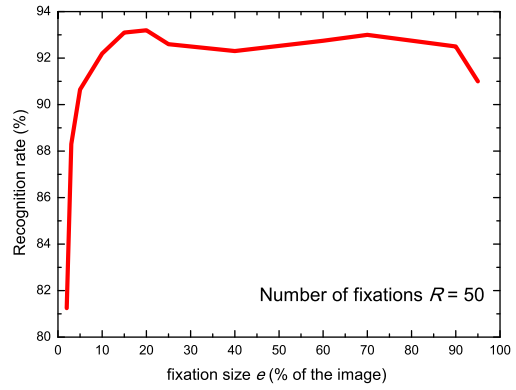
Method	Session 1			Session 2			Avg.
	Sm.	An.	Sc.	Sm.	An.	Sc.	
Stringface[5]	87.5	87.5	25.9	-	-	-	67.0
FARO[6]	96.0	-	60.0	-	-	-	78.0
SRC[29]	98.0	89.0	55.0	79.0	78.0	31.0	71.7
PD[22]	**100.0**	97.0	93.0	88.0	86.0	63.0	87.8
SOM[23]	**100.0**	98.0	88.0	88.0	90.0	64.0	88.0
DMMA[16]	99.0	93.0	69.0	85.0	79.0	45.0	78.3
DICW[27]	**100.0**	99.0	84.0	91.0	92.0	44.0	85.0
CTSDP[8]	**100.0**	**100.0**	**95.5**	**98.2**	**99.1**	**86.4**	**96.5**
Ours	**100.0**	**100.0**	91.4	94.5	98.0	58.6	90.4

4.4. Discussion of parameters

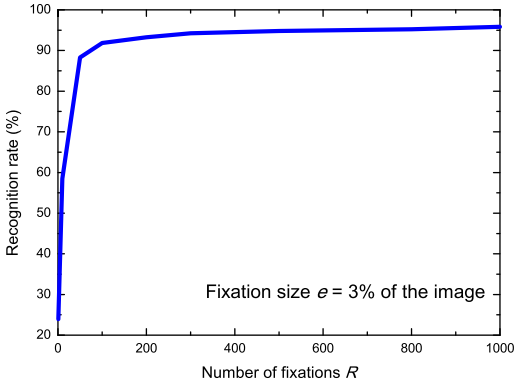
The effect of patch size s on the recognition performance is discussed in our previous work[27]. Here we fix s according to the settings in [27] and study the influence of the fixation size e and the number of fixations R. We conduct experiments on the AR database using images with sunglasses and scarves. The recognition rates as a function of e and R when one parameter is fixed are shown in Fig. 4. Intuitively, if e is too small (< 3% of the image), the order information contained in the fixation sequence will be very limited and not suitable for recognition. On the other hand, as mentioned in Sec. 3, since the time complexity of our method is $\boldsymbol{O}(R(\frac{e}{s})^2)$, the increase of e leads to higher computational cost (s is fixed). It can be seen from Fig. 4b, the recognition rate is monotonically increasing with respect to the increasing R. Considering the computational efficiency, in our experiments, we set $e = 3\%$ of the image and empirically increase the number of fixations to $R = 300$ in order to gain higher recognition accuracy.

5. Conclusion

We proposed a novel method for face recognition from single image per person, which is inspired by the mechanism of fixations and saccades in human visual perception. On the two well-known face databases (FRGC and AR), our method clearly outperforms the current approaches when dealing with the occlusions and expression changes. In some extreme cases where the uncontrolled variations cause large deformations of the face, our method achieves comparable performance with the 2D warping based method at a much lower computational cost.



(a)



(b)

Figure 4. a) Recognition rate (%) as a function of the fixation size e and b) recognition rate (%) as a function of the number of fixations R.

References

[1] T. Ahonen, A. Hadid, and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2037–2041, 2006.

[2] L. Barrington, T. K. Marks, J. Hui-wen Hsiao, and G. W. Cottrell. Nimble: A kernel density model of saccade-based visual memory. *Journal of Vision*, 8(14), 2008.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[4] C.-C. Chang and C.-J. Lin. LIBSVM: A library for support vector machines. *ACM Trans. Intelligent Systems*

and Technology, 2:27:1–27:27, 2011. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[5] W. Chen and Y. Gao. Recognizing partially occluded faces from a single sample per class using string-based matching. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision - ECCV 2012*, volume 6313 of *Lecture Notes in Computer Science*, pages 496–509. Springer Berlin Heidelberg, 2010.

[6] M. De Marsico, M. Nappi, and D. Riccio. Faro: Face recognition against occlusions and expression variations. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 40(1):121–132, 2010.

[7] H. K. Ekenel and R. Stiefelhagen. Why is facial occlusion a challenging problem? In *Proceedings of the Third International Conference on Advances in Biometrics*, ICB '09, pages 299–308, Berlin, Heidelberg, 2009. Springer-Verlag.

[8] T. Gass, L. Pishchulin, P. Dreuw, and H. Ney. Warp that smile on your face: Optimal and smooth deformations for face recognition. In *Automatic Face Gesture Recognition, 2011. FG '11. 9th IEEE International Conference on*, pages 456–463, 2011.

[9] Y. Guan and C.-T. Li. A robust speed-invariant gait recognition system for walker and runner identification. In *Biometrics (ICB), 2013 6th IAPR International Conference on*, 2013.

[10] C.-K. Hsieh, S.-H. Lai, and Y.-C. Chen. An optical flow-based approach to robust face recognition under expression variations. *Image Processing, IEEE Transactions on*, 19(1):233–240, 2010.

[11] H. Jia and A. Martinez. Face recognition with occlusions in the training and testing sets. In *Automatic Face Gesture Recognition, 2008. FG '08. 8th IEEE International Conference on*, pages 1 –6, sept. 2008.

[12] H. Jia and A. Martinez. Support vector machines in face recognition with occlusions. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 136–141, 2009.

[13] C. Kanan and G. Cottrell. Robust classification of objects, faces, and flowers using natural image statistics. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2472–2479, 2010.

[14] X.-X. Li, D.-Q. Dai, X.-F. Zhang, and C.-X. Ren. Structured sparse error coding for face recognition with occlusion. *Image Processing, IEEE Transactions on*, 22(5):1889–1900, 2013.

[15] S. Liao, A. K. Jain, and S. Z. Li. Partial face recognition: Alignment-free approach. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(5):1193–1205, 2013.

[16] J. Lu, Y.-P. Tan, and G. Wang. Discriminative multimanifold analysis for face recognition from a single training sample per person. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):39–51, 2013.

[17] A. Martınez and R. Benavente. The ar face database. *Computer Vision Center, Technical Report*, 24, 1998.

[18] A. M. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(6):748–763, 2002.

[19] P. Phillips, P. Flynn, T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 947 – 954 vol. 1, june 2005.

[20] L. Pishchulin, T. Gass, P. Dreuw, and H. Ney. Image warping for face recognition: From local optimality towards global optimization. *Pattern Recognition*, 45(9):3131 – 3140, 2012.

[21] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE*, 94(11):1948–1962, 2006.

[22] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Face recognition under occlusions and variant expressions with partial similarity. *IEEE Trans. Information Forensics and Security*, 4(2):217 – 230, june 2009.

[23] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Recognizing partially occluded, expression variant faces from single training image per person with som and soft k-nn ensemble. *Neural Networks, IEEE Transactions on*, 16(4):875–886, 2005.

[24] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang. Face recognition from a single image per person: A survey. *Pattern Recognition*, 39(9):1725 – 1745, 2006.

[25] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*, 3(1):71–86, Jan. 1991.

[26] X. Wei, C.-T. Li, and Y. Hu. Robust face recognition under varying illumination and occlusion considering structured sparsity. In *Digital Image Computing Techniques and Applications (DICTA), 2012 International Conference on*, pages 1–7, 2012.

[27] X. Wei, C.-T. Li, and Y. Hu. Face recognition with occlusion using dynamic image-to-class warping (dicw). In *Automatic Face Gesture Recognition, 2013. FG '13. 10th IEEE International Conference on*, 2013.

[28] X. Wei, C.-T. Li, and Y. Hu. Robust face recognition with occlusions in both reference and query images. In *International Workshop on Biometrics and Forensics (IWBF'13)*, 2013.

[29] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(2):210 –227, feb. 2009.