Face Recognition with Occlusion Using Dynamic Image-to-Class Warping (DICW)

Xingjie Wei, Chang-Tsun Li and Yongjian Hu
Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK
{x.wei,c-t.li}@warwick.ac.uk, yongjian.hu@dcs.warwick.ac.uk

***Abstract*— A novel approach Dynamic Image-to-Class Warping (DICW) is proposed to deal with partially occluded face recognition in this work. An image is partitioned into sub-patches, which are then concatenated in the raster scan order to form a sequence. A face consists of forehead, eyes, nose, mouth and chin in a natural order and this order does not change despite occlusion or small rotation. Thus, in this work, a face is represented by the aforementioned sequence which contains the order of facial features. Taking the *order information* into account, DICW computes the distance between a query face and an enrolled person by finding the optimal alignment between the query sequence and all sequences of that person along both *time* dimension and *within-class* dimension. Extensive experiments on public face databases with various types of occlusion have verified the effectiveness of the proposed method. In addition, our method, which considers the inherent structure of the face, performs with greater consistency than current methods when the number of enrolled images per person is limited. Our method does not require any training process and has a low computational cost, which makes it applicable for real-world FR applications.**

I. Introduction

In recent years, face recognition (FR) under unconstrained conditions has attracted wide research interests. FR systems in unconstrained conditions need to deal with variations in illumination, pose, expression, etc. In this paper, we focus on the recognition of partially occluded faces, which often occurs in real-world environments but has not attracted much attention yet. The sources of occlusion include accessories (e.g., sunglasses, scarves), shadows or other objects in front of a face. The difficulty due to occlusion is twofold. Firstly, occlusion increases the distance in the feature space between two face images of the same person. The intra-class variations are usually larger than the inter-class variations, which results in poorer recognition performance. Holistic methods such as Eigenface[2] are not applicable since all the extracted features would be distorted due to occlusion. Secondly, occlusion causes imprecise registration of faces. For FR systems, face registration by aligning facial landmarks (e.g., the centres of the eyes, the tip of the nose) is needed before performing feature extraction and recognition. However, since facial landmarks are occluded, a large registration error may occur and make the recognition fail.

If the occlusion is well detected, better recognition performance can be achieved by using only the unoccluded areas. Unfortunately, the sources of occlusion are unpredictable in real-world scenarios. The location, size and shape of occlusion are unknown, hence increasing the difficulty in segmenting the occluded areas from a face image.

There are two categories of approaches to handle occlusion related problems. The first one treats occluded FR as a reconstruction problem[6], [18], [19], [20]. A clean image is reconstructed from an occluded image by a linearly combination of gallery images (i.e., enrolled faces). Then the occluded image is assigned to the class with the minimal reconstruction error. These approaches require sufficient samples per person to represent the probe images (i.e., query faces), which is not always available in the real environment. The other category is patch-matching based approach[9], [16]. An occluded image is first partitioned into sub-patches wherein features are extracted. Then matching is performed based on the similarity of each patch-pair. In order to minimize the large matching error due to the occluded patches, different weights generated through training are given to different patches. However, the training process is usually data-dependent.

We propose a patch-matching method Dynamic Image-to-Class Warping (DICW) to deal with occlusion. DICW is based on Dynamic Time Warping (DTW) which is widely used in speech recognition[15]. In our work, an image is partitioned into sub-patches, which are then concatenated in the raster scan order to form a sequence. In this way, a face is represented by a patch sequence which contains *order information* of facial features. DICW calculates the *Image-to-Class* distance between a query face and an enrolled person by finding the optimal alignment between the query sequence and all the enrolled sequences of that person. The advantages of DICW are as follows:

1) It can directly deal with occluded images without occlusion detection. DICW is a patch based method where features are locally extracted, so it is less likely to be affected by occlusion compared with the holistic approaches.
2) It does not need any training process. On one hand, different from the learning based approaches such as PCA[2] and SVM[7], DICW does not need a training process to find the projection vector or hyperplane for classification. Such a training process usually requires sufficient and representative training samples. On the other hand, compared with other patch-matching based methods, weights training is not required. In DICW, the optimal alignment between the probe and the gallery can be automatically found by *Dynamic Programming (DP)*

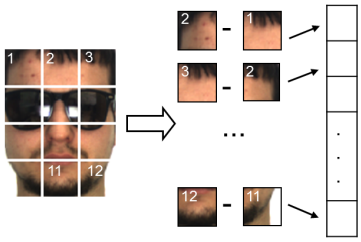


Fig. 1. Image representation of DICW.

and matching is mainly based on the unoccluded parts of the face.

3) It works well with limited gallery images per person. DICW calculates the *Image-to-Class* distance, instead of individual *Image-to-Image* distance. It is capable of exploiting the intra-class information from the limited gallery images of each class.
4) It is appropriate for real FR applications. The distances between a query face and each enrolled person are calculated separately thus can be generated in parallel. The enrolled database can be updated incrementally in FR systems.

Most of the existing works that simply treat occluded FR as a signal recovery problem or just employ the framework for general object classification, do not consider the inherent structure of the face. In this paper, we proposed a novel approach that takes the *order information* of facial features into account when dealing with partially occluded faces. In uncontrolled and uncooperative environments, occlusion preprocessing and collection of sufficient and representative training samples are generally very difficult. The contribution of our work is that we provide an effective candidacy for realistic FR applications. Our model is compatible with various image descriptors and feature extraction methods for real-world FR applications.

II. Proposed Approach

A. Image representation

An image is partitioned into M non-overlapping sub-patches of $d \times d'$ pixels. Then those sub-patches are concatenated in the raster scan order (i.e., left to right and top to bottom) to form a single sequence. A face consists of forehead, eyes, nose, mouth and chin in a natural order. Despite the occlusion or imprecise registration, this order does not change. A face image is represented as a patch sequence where the order of patch positions (i.e., the order of facial features) can be viewed as the temporal information. Thus, time series analysis techniques can be used to handle the FR problem.

Denoting i-th patch by F_i, a difference patch ∇F_i is computed by applying subtraction between the values of corresponding pixels in two neighbouring patches as:

$$\nabla F_i(x, y) = F_{i+1}(x, y) - F_i(x, y) \quad (1)$$

where (x, y) is the coordinates of a pixel.

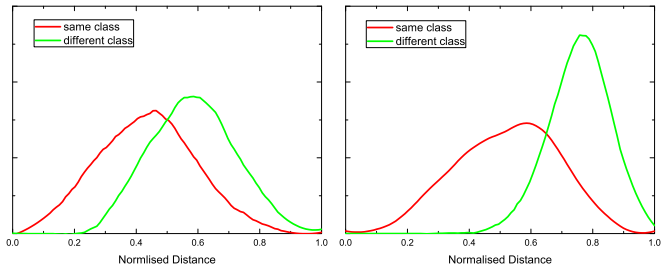


Fig. 2. Distributions of face image distance of the same and different classes. Using the difference patch (right), the distance distribution of the same class and that of the different class are separated more widely compared with those using the original patch (left).

Those difference patches generated by the spatially continuous patches enhance the *order information* within the face sequence, which is compatible with our model for image distance calculation (see Section II-B). In addition, from a macroscopic perspective, when the size of each patch is very small, each patch can be viewed as a *pixel*. A difference patch ∇F_i actually can be viewed as the approximation of the first-order derivative of adjacent *pixels* F_{i+1} and F_i. The salient facial features which represent detailed textured regions such as eyes, nose and mouth can be enhanced since the first-order derivative operator is sensitive to edge. Fig.1 shows the image representation framework of DICW.

In our DICW, we use the grey value of difference patches as feature. Fig.2 shows the face distance distributions of the same and different classes (persons) computed using the original patches and the difference patches. As can be seen, the distance distributions of the same and different classes are separated more widely after using the difference patches (Fig.2 right).

B. Dynamic Image-to-Class Warping

Matching faces is implemented by defining a distance measure between sequences. Generally, a small distance is expected if two sequences are similar to each other. Our DICW is based on the DTW algorithm[15] which is used to compute the distance between two temporal sequences. For example, for two sequences $A = (3, 1, 10, 5, 6)$ and $B = (3, 2, 1, 10, 5)$, the Euclidean distance between them is $\sqrt{0 + 1 + 81 + 25 + 1} \approx 10.39$ which is a bit large for these two similar sequences. However, if we ignore the "2" in B and match "1,10,5" in A with "1,10,5" in B rather than using the bit-wise matching, the distance between A and B will be largely reduced. DTW which is based on this idea calculates the distance between two sequences by finding the optimal alignment between them with the minimal overall cost. The *order information* is considered during matching, thus cross-matching is not allowed (e.g., matching "3" in A with "2" in B while matching "1" in A with "3" in B). For FR, this is reasonable since the order of facial features should not be reversed.

Borrowing this idea to FR, we want to find the optimal alignment between face sequences while reducing the effect

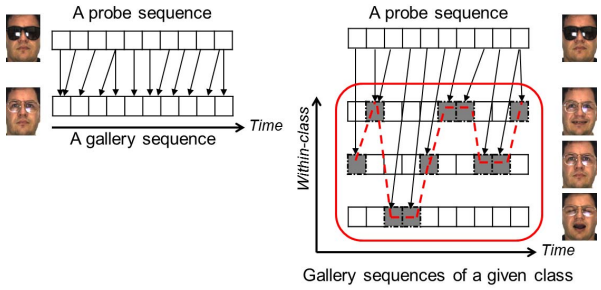


Fig. 3. An illustration of alignment by DTW (left) and DICW (right). The arrows indicate the alignment for each patch. The dashed line marks the optimal warping path between an image and a gallery set.

of occluded patches. In addition, in this work we extend the *alignment* between two sequences into between a sequence and the sequence set of a given class. A probe image consisting of M patches is denoted by $\boldsymbol{P} = \{\boldsymbol{p}_1, ..., \boldsymbol{p}_m, ..., \boldsymbol{p}_M\}$ where $\boldsymbol{p}_m$ is a difference patch as mentioned in Section II-A. The gallery set of a given class containing K images is denoted by $\boldsymbol{G} = \{\boldsymbol{G}_1, ..., \boldsymbol{G}_k, ..., \boldsymbol{G}_K\}$ where each $\boldsymbol{G}_k = \{\boldsymbol{g}_{k1}, ..., \boldsymbol{g}_{kn}, ..., \boldsymbol{g}_{kN}\}$ is a gallery image consisting of N patches and $\boldsymbol{g}_{kn}$ indicates a patch vector like $\boldsymbol{p}_m$. An illustration of the alignment by Dynamic Image-to-Class Warping is shown in Fig.3. Here *warping* is performed in two directions: 1) a probe sequence $\boldsymbol{P}$ is aligned to a set of gallery sequences $\boldsymbol{G}$ of a given class according to the *time* dimension (maintaining the *order information*) and 2) simultaneously, in each temporal step each patch in $\boldsymbol{P}$ is matched to the most similar patch among all gallery sequences along the *within-class* dimension.

A warping path $\boldsymbol{W}$ indicating the alignment between $\boldsymbol{P}$ and $\boldsymbol{G}$ of T temporal steps is defined as $\boldsymbol{W} = \{w_1, w_2, ..., w_T\}$. The t-th element w_t is an index triplets (m_t, n_t, k_t) which indicates that patch $\boldsymbol{p}_{m_t}$ is matched to patch $\boldsymbol{g}_{k_t n_t}$ where $m_t \in \{1, 2, ..., M\}, n_t \in \{1, 2, ..., N\}$ and $k_t \in \{1, 2, ..., K\}$. Considering the context of FR, $\boldsymbol{W}$ satisfies the following four constraints:

1) Boundary constraint: $m_1 = 1, n_1 = 1$ and $m_T = M, n_T = N$. The path starts at matching $\boldsymbol{p}_1$ to $\boldsymbol{g}_{k_1 1}$ and ends at matching $\boldsymbol{p}_M$ to $\boldsymbol{g}_{k_T N}$. From 1 to T, k can be any value from 1 to K since the probe patch can be matched with patches from any K gallery images.
2) Continuity constraint: $m_t - m_{t-1} \leqslant 1$ and $n_t - n_{t-1} \leqslant 1$. The indexes of the path increase by 1 in each step. All patches in probe and gallery images will be processed.
3) Monotonicity constraint: $m_{t-1} \leqslant m_t$ and $n_{t-1} \leqslant n_t$. The path preserves the *temporal order* and increase monotonically.
4) Window constraint: $|m_t - n_t| \leqslant l$ where $l \in \mathbb{N}^+$ is the window length[15]. A probe patch should not match to a patch too far away (e.g., an eye should not be matched to the mouth). The window with a length l is able to constrain the warping path within an appropriate range.

We create a local distance matrix $\boldsymbol{C} \in \mathbb{R}^{M \times N \times K}$ where each element $C_{m,n,k}$ stores the Euclidean distance, namely the *local cost*, between $\boldsymbol{p}_m$ and $\boldsymbol{g}_{kn}$: $C_{m,n,k} = \|\boldsymbol{p}_m - \boldsymbol{g}_{kn}\|_2$. The overall cost of $\boldsymbol{W}$ is defined as:

$$S(\boldsymbol{W}) = \sum_{t=1}^{T} C_{w_t} \tag{2}$$

The optimal alignment (i.e., the optimal warping path) $\boldsymbol{W}^*$ is the path with minimal $S(\boldsymbol{W})$. The *Image-to-Class* distance between $\boldsymbol{P}$ and $\boldsymbol{G}$ is simply the overall cost of $\boldsymbol{W}^*$:

$$DICW(\boldsymbol{P}, \boldsymbol{G}) = \min_{\boldsymbol{W}} S(\boldsymbol{W}) \tag{3}$$

To compute $DICW(\boldsymbol{P}, \boldsymbol{G})$, one could test every possible warping path but with a high computational cost. Fortunately, equation (3) can be solved efficiently by *Dynamic Programming (DP)*. A three-dimensional cumulative matrix $\boldsymbol{D} \in \mathbb{R}^{M \times N \times K}$ is created. The element $D_{m,n,k}$ stores the cost of the sub-problem that assigning a m-patch sequence to a set of n-patch sequences and matching the m-th patch $\boldsymbol{p}_m$ to the patch from the k-th gallery image. The calculation of $DICW(\boldsymbol{P}, \boldsymbol{G})$ is based on the results of a series of sub-problems. $\boldsymbol{D}$ can be computed recursively as:

$$D_{m,n,k} = \min \begin{pmatrix} \boldsymbol{D}_{\{(m-1,n-1)\} \times \{1,2,...,K\}} \\ \boldsymbol{D}_{\{(m-1,n)\} \times \{1,2,...,K\}} \\ \boldsymbol{D}_{\{(m,n-1)\} \times \{1,2,...,K\}} \end{pmatrix} + C_{m,n,k} \tag{4}$$

where the initialization is by extending $\boldsymbol{D}$ as an $(M+1)$-by-$(N+1)$-by-K matrix and setting $\boldsymbol{D}_{0,0,k} = 0, \boldsymbol{D}_{0,n,k} = \boldsymbol{D}_{m,0,k} = \infty$. $\times$ indicates the Cartesian product operation. Thus, $DICW(\boldsymbol{P}, \boldsymbol{G})$ can be obtained by:

$$DICW(\boldsymbol{P}, \boldsymbol{G}) = \min_{k \in \{1,2,...,K\}} \boldsymbol{D}_{M,N,k} \tag{5}$$

Then the probe image $\boldsymbol{P}$ is classified as the class with the shortest distance. Different from the restrictions often used in video retrieval as in [1], in DICW no restriction is placed on k in the first term of equation(4). Because in each temporal step, the probe patch can be matched with patches from any K gallery images.

DICW is able to find the alignment with minimal overall cost from all possible warping combinations. Thus, the overall distance mainly relies on the most similar parts of faces. This is compatible with the visual perception since human being recognizes an occluded face based on the unoccluded areas. Fig.4 shows an illustration of the *Image-to-Class* warping. In Fig.4a, the occluded face belongs to class 2 but is misclassified when using the *Image-to-Image* distance (the nearest neighbour is from class 1). Although the *Image-to-Image* distance is large to each individual gallery image of class 2, the *Image-to-Class* distance is small and leads to the correct classification. As shown in Fig.4b, in each step, each patch in the probe image is matched to the most similar patch from one gallery image. The *Image-to-Class* distance is calculated mainly based on these most similar patch pairs (matched patches are indicated by the same colour).

Algorithm 1 summarizes the procedure of computing the *Image-to-Class* distance between a probe image and a class. The time complexity is $\boldsymbol{O}(l \max(M, N)K)$ where $l \ll M$

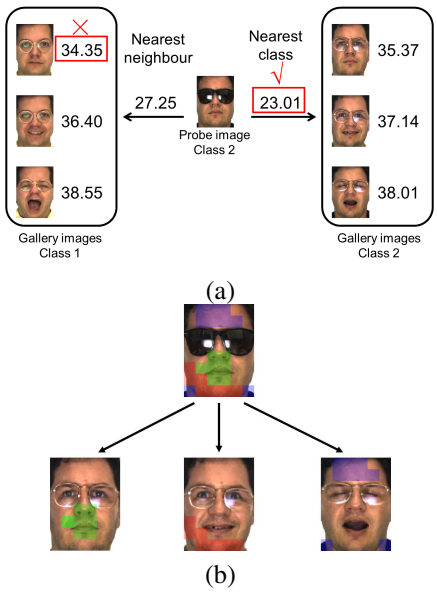


Fig. 4. (a): Comparison of the *Image-to-Image* distance and the *Image-to-Class* distance. The occluded face belongs to class 2 but is misclassified when using the *Image-to-Image* distance (the nearest neighbour is from class 1). Using the *Image-to-Class* distance leads to the correct classification. (b): The distance between a probe image and a class is calculated mainly based on the most similar patches from different gallery images. Matched patches are indicated by the same colour (best viewed in colour version).

(l is usually set to 10% of $\max(M, N)$[15]). When the number of gallery images per class is limited, K is relatively very small. So the warping distance can be obtained very efficiently. Our Matlab implementation takes average 0.05 seconds for computing the warping distance between one image and a class using about 200-patch sequences (with 3.10GHz CPU and 8 GB memory). Compared with the reconstruction based approaches[6], [19], [20] which represent a test image using all enrolled images, DICW computes the distance between the probe image and each enrolled class independently. So in the real FR applications, the distance matrix can be generated in parallel and the enrolled database can be updated incrementally.

III. Experimental Analysis

In this section, we use the proposed method to deal with different types of occlusion: synthetic occlusion, disguises occlusion and realistic occlusion. A set of large-scale experiments are conducted using two public databases (the FRGC2.0 (Face Recognition Grand Challenge) database[14] and the AR database[10]) and one outdoor environment database[12]. In the FRGC database, to simulate different levels (from 0% to 50% of an image) of contiguous occlusion, we replace a randomly located square patch from each test image with a black block. Notice that the location of occlusion is randomly chosen and unknown to the algorithm. The AR database is used since it is one of the very few databases that contain real disguise. We also evaluate our method using a database with realistic images collected by Dexter Miranda[12]. These images contain frontal view faces of strangers on the streets with uncontrolled lighting. The

Algorithm 1 Dynamic Image-to-Class Warping distance DICW($\boldsymbol{P}, \boldsymbol{G}, l$)

Input:
The probe image: $\boldsymbol{P}$;
The set of gallery images of a given class: $\boldsymbol{G}$;
The window length: l;
Output:
The *Image-to-Class* distance between $\boldsymbol{P}$ and $\boldsymbol{G}$: S;

Compute the local distance matrix $\boldsymbol{C}$;
Initialize: each element in $\boldsymbol{D}$ is set to ∞;
$\boldsymbol{D}[0, 0, 1:K] = 0$;
for $m = 1$ to M **do**
  for $n = \max(1, m-l)$ to $\min(N, m+l)$ **do**
    minNeighbour=$\min \begin{pmatrix} \boldsymbol{D}[m-1, n-1, 1:K] \\ \boldsymbol{D}[m-1, n, 1:K] \\ \boldsymbol{D}[m, n-1, 1:K] \end{pmatrix}$;
    for $k = 1$ to K **do**
      $D[m, n, k]$ =minNeighbour+$C[m, n, k]$;
    end for
  end for
end for
$S = \min(\boldsymbol{D}[M, N, 1:K])$;
return S;

Fig. 5. Examples of the test images from the realistic image database.

sources of occlusion include glasses, hat, hair and hand on the face. Besides occlusion, these images also contains expression variations and small rotation (Fig.5).

We quantitatively compare the proposed method DICW with four representative methods in the literature: the linear SVM[4] using Eigenface[17] as features (LSVM), the reconstruction based Sparse Representation (SRC)[19], the patch-matching based Naive Bayes Nearest Neighbor (NBNN)[3] and the baseline Nearest Neighbor (NN)[5]. For NBNN[1] which is also a patch-matching based method, we use the original patches and difference patches for testing respectively and report the best results.

A. The effect of sub-patch size

We first investigate the effect of varying sub-patch sizes on the recognition performance. 400 unoccluded images from the FRGC database are used as the gallery set. Six probe sets, which contain synthetic occlusion from 0% level to 50% level respectively, are used for testing. All images are cropped and resized to 80×65 pixels. We test DICW with the sub-patch sizes from 3×3 pixels to 10×10 pixels.

[1]We have tested different values of location distance parameter for NBNN and report the best results.

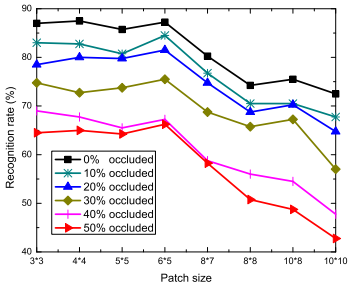


Fig. 6. Recognition rates as a function of patch size.

For simplicity, the probe images and the gallery images are partitioned into the same number of sub-patches, i.e.,$M = N$. The recognition rate as a function of the sub-patch size are shown as Fig.6. There is no sharp fluctuation in each of the recognition rate curve when the patch size is equal to or smaller than 6×5 pixels. Our method is robust to different sub-patch size in an appropriate range despite the ratio of occlusion. We also test DICW on the AR database and the results are very similar. The relatively smaller patches lead to better recognition rate since they provide more flexibility to use spatial information than the larger ones. However, if the size is too small, it would result in high computational and memory cost. As a result, the sizes from 4×4 to 6×5 pixels are recommended. In our experiments, according to the image size, patch size of 6×5 pixels is used in the FRGC database and 5×5 pixels is used in the AR database and the realistic database.

B. Recognition with synthetic occlusion

We first evaluate the proposed method using the FRGC database with synthetic occlusion. We choose 100 subjects and 8 images for each subject. For each subject, we select $K = 1, 2, 3$ and 4 unoccluded images as gallery sets respectively and the other 4 images with synthetic occlusion as the probe set. Notice that in each testing the gallery set does not intersect the probe set. All images are cropped and resized to 80×65 pixels.

Fig.7 shows the recognition rates of all five methods with different number of gallery images per person. The performance of SRC is slightly better than DICW at the beginning (occlusion level $\leqslant 10\%$), however, drops sharply when occlusion increases. The patch-matching based methods NBNN and DICW perform better than other methods on the whole. Our DICW behaves consistently and outperforms other methods on all tests. When only a single gallery image per person is available ($K = 1$), the *Image-to-class* distance is equal to the *Image-to-Image* distance in DICW. However, our method still achieves the best recognition rates on all levels of occlusion since it is able to find the optimal alignment between gallery and probe images by warping considering the order of facial features.

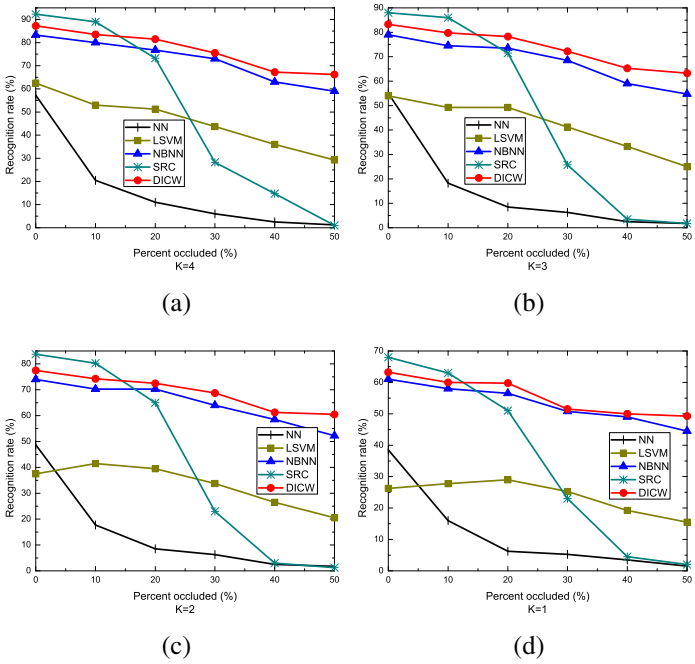


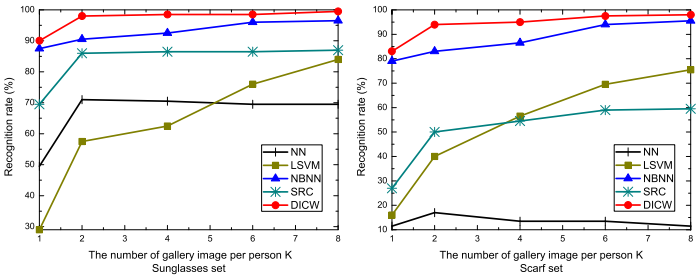
Fig. 7. Recognition rates on the FRGC database with different number of gallery images per person.

C. Recognition with disguises

We next test the proposed method on the AR database with real disguise. A subset[11] of the AR database (50 men and 50 women) containing varying illumination conditions, expressions and occlusion is used. The unoccluded frontal view images with various expressions are used as the gallery images (8 images per person). For each person, we select $K = 1, 2, 4, 6$ and 8 from those images as gallery sets respectively. Two separate sets (200 images each) of images containing sunglasses (cover about 30% of the image) and scarves (cover about 50% of the image) respectively are used as probe sets. All images are cropped and resized to 83×60 pixels.

Fig.8 shows the recognition results of the proposed DICW and other methods. The recognition rates drop when less gallery images are available. Our DICW has consistent performance and significantly outperforms other methods. Even at $K = 1$, DICW still achieves 90% and 83% on sunglasses set and scarf set respectively.

With the same experimental setting, we also compare DICW using 8 gallery image per person ($K = 8$) with those of other state-of-the-art algorithms in the FR literature. Only grey value features are used in all methods. The results are as shown in Table I. The recognition rates of other methods are cited directly from their papers. Notice that compared with these methods, our method does not require any training process. DICW achieves comparable or better recognition rates among these methods and about 15 times faster than the work in [19]. In the scarf set nearly half of the face is occluded, however, only 2% images are misclassified by DICW. To the best of our knowledge, this is the best results



(a) (b)

Fig. 8. Recognition rates on the AR database with sunglasses occlusion (a) and scarf occlusion (b).

achieved on the scarf set using only the grey value as feature.

TABLE I

RECOGNITION RATES (%) ON THE AR DATABASE WITH 8 GALLERY IMAGES PER PERSON

	Sunglasses	Scarf	Average
SRC-partition[19]	97.5	93.5	95.5
LRC[13]	96.0	26.0	61.0
CRC-RLS[20]	91.5	95.0	93.3
R-CRC[20]	92.0	94.5	93.3
Zhou et al[21]	99.0∼**100**	95.0∼ 97.5	97∼**98.8**
Jia et al[8]	99.5	87.5	93.5
Proposed DICW	99.5	**98.0**	**98.8**

D. Recognition with realistic images

Finally, we test our method on a face database captured under realistic conditions[12]. We select a set of images containing 55 subjects in our experiments. For each person, we choose $K = 1, 3, 5$ and 8 unoccluded images as gallery sets respectively. 110 images with different types of occlusion are used for testing. Each gallery set does not intersect the probe set. The testing examples are shown in Fig.5. The face area of each image is cropped from the background and resized to 80×60 pixels.

The recognition results are shown in Table.II. Notice that different from the previous testing images which captured under indoor conditions, these images are taken under totally uncontrolled conditions in real environment. The overall recognition rates of all methods are relatively low due to the challenging nature of the occlusion. Compared with other methods, DICW still achieves better performance.

TABLE II

RECOGNITION RATES (%) ON THE REALISTIC IMAGES

Gallery size per person	1	3	5	8
NN[5]	56.4	66.4	70.9	71.8
LSVM[4]	19.1	30.0	41.8	48.2
NBNN[3]	52.7	66.4	74.6	73.6
SRC[19]	60.0	67.3	70.9	70.9
Proposed DICW	**61.8**	**76.4**	**77.3**	**81.8**

IV. CONCLUSION

We have proposed a novel approach, DICW, to the recognition of occluded faces. Extensive experiments in three face databases show that DICW achieves much better performance than four representative methods in the literature. In the extreme case where only single gallery image is available for each person, DICW still performs consistently. Our DICW can be applied directly to raw data without performing occlusion detection in advance and does not require a training process. All of these make our approach more applicable in real-world scenarios.

In addition, our model is very flexible and can adopt other image descriptors such as LBP and Gabor. The investigation of combining our model with more sophisticated feature extraction approaches is an important direction for future work.

REFERENCES

[1] J. Alon, V. Athitsos, Q. Yuan, and S. Sclaroff. A unified framework for gesture recognition and spatiotemporal gesture segmentation. *IEEE Trans. PAMI*, 31(9):1685 –1699, sept. 2009.

[2] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans. PAMI*, 19(7):711 –720, jul 1997.

[3] O. Boiman, E. Shechtman, and M. Irani. In defense of nearest-neighbor based image classification. In *Proc. of CVPR 2008*, pages 1 –8, june 2008.

[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] T. Cover and P. Hart. Nearest neighbor pattern classification. *IEEE Trans. Information Theory*, 13(1):21 –27, january 1967.

[6] H. Jia and A. Martinez. Face recognition with occlusions in the training and testing sets. In *Proc. of FG 2008*, pages 1 –6, sept. 2008.

[7] H. Jia and A. Martinez. Support vector machines in face recognition with occlusions. In *Proc. of CVPR 2009*, pages 136 –141, june 2009.

[8] K. Jia, T.-H. Chan, and Y. Ma. Robust and practical face recognition via structured sparsity. In *Computer Vision C ECCV 2012*, Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2012.

[9] A. Martinez. Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class. *IEEE Trans.PAMI*, 24(6):748 –763, jun 2002.

[10] A. Martınez and R. Benavente. The ar face database. *CVC Tech. Rep.*, 24, 1998.

[11] A. Martinez and A. Kak. Pca versus lda. *IEEE Trans. PAMI*, 23(2):228 –233, feb 2001.

[12] D. Miranda. The face we make. `www.thefacewemake.org`.

[13] I. Naseem, R. Togneri, and M. Bennamoun. Linear regression for face recognition. *IEEE Trans. PAMI*, 32(11):2106 –2112, nov. 2010.

[14] 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 *Proc. of CVPR 2005*, volume 1, pages 947 – 954 vol. 1, june 2005.

[15] H. Sakoe and S. Chiba. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech and Signal Processing*, 26(1):43 – 49, feb 1978.

[16] 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.

[17] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proc. of CVPR 1991*, pages 586 –591, jun 1991.

[18] X. Wei, C.-T. Li, and Y. Hu. Robust face recognition under varying illumination and occlusion considering structured sparsity. In *Proc. of Int.l Conf. Digital Image Computing Techniques and Applications (DICTA) 2012*, dec. 2012.

[19] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *IEEE Trans. PAMI*, 31(2):210 –227, feb. 2009.

[20] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang. Collaborative Representation based Classification for Face Recognition. *eprint arXiv:1204.2358*, Apr. 2012.

[21] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma. Face recognition with contiguous occlusion using markov random fields. In *Proc. of ICCV 2009*, pages 1050 –1057, 29 2009-oct. 2 2009.