Face Recognition with Occlusions in the Training and Testing Sets

Hongjun Jia and Aleix M. Martinez
The Department of Electrical and Computer Engineering
The Ohio State University, Columbus, OH 43210, USA
jia.22@osu.edu aleix@ece.osu.edu

Abstract

Partial occlusions in face images pose a great problem for most face recognition algorithms. Several solutions to this problem have been proposed over the years – ranging from dividing the face image into a set of local regions to sophisticated statistical methods. In the present paper, we pose the problem as a reconstruction one. In this approach, each test image is described as a linear combination of the training samples in each class. The class samples providing the best reconstruction determine the class label. Here, "best reconstruction" means that reconstruction providing the smallest matching error when using an appropriate metric to compare the reconstructed and test images. A key point in our formulation is to base this reconstruction solely on the visible data in the training and testing sets. This allows to have partial occlusions in both the training and testing samples, while previous methods only dealt with occlusions in the testing set. We show extensive experimental results using a large variety of comparative studies, demonstrating the superiority of the proposed approach over the state of the art.

1. Introduction

To date, many algorithms have been defined to do recognition of faces under a large variety of image conditions [11]. One problem that has received considerable attention in recent years is that of partially occluded faces. Since the work of Martinez [4], a variety of methods have been proposed for matching non-occluded training samples to partially occluded test images [1, 2, 6, 8, 9, 10]. The goal is to define a matching technique that omits the large matching errors due to occlusions while concentrating with those of the non-occluded parts.

In this paper, we take a different view. We redefine the face recognition problem as a reconstruction one. In this approach, the training samples of a class are linearly combined to create a new image that is as close as possible to the test image. The hypothesis is that the most accurate reconstruction will be given when one uses the samples of the correct class. This is a grounded hypothesis, since the image reconstruction of a frontal face image will generally be most accurately obtained when combining face images of the person it represents rather than with images of other individuals.

Under this view, the major problem is to define an appropriate mechanism to do the reconstruction. If the training and testing images had no occlusions, one of the simplest approaches would be to try to represent the test image $\mathbf{t}$ as a linear combination of the n_i training samples in class i, $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\}$,

$$\mathbf{t} \approx \sum_{j=1}^{n_i} w_{ij}\mathbf{x}_{ij},$$

where $w_{ij} \in \mathbb{R}$ are the weights describing the contribution of each image, $\mathbf{t}$ and $\mathbf{x}_j$ are the p-dimensional vectorized images, $p = ab$, and $a \times b$ defines the image size.

The problem of estimating the weights w_{ij} can be stated as,

$$\mathbf{w}_i = \arg \min_{(w_{i1},\ldots,w_{in_i})^T} \left\| \mathbf{t} - \sum_{j=1}^{n_i} w_{ij}\mathbf{x}_{ij} \right\|_r. \qquad (1)$$

The selection of the metric above, which is given by the r-norm $\|\cdot\|_r$, regulates how each feature (dimension) can be used to reconstruct the test image. The selection of the norm also defines the space ℓ_r.

The most commonly used norms in minimization problems such as that defined in (1) are the 2-norm, which provides the least-squares (LS) solution, and the 1-norm, which generates sparse representations. A recent result [10] shows that ℓ_1 is preferred because it can handle sparse occlusions in the test image. In this approach, the sample images of all classes are used to estimate an occluded test image. Since ℓ_1 emphasizes sparseness, only a very small number of sample images will be used to linearly reconstruct the test image. The samples which will generally best reconstruct the test instance, are those associated to the same class label (*i.e.*, identity), facilitating the recognition process. However, a concern of this approach is that the ℓ_1-minimization is computationally expensive.

978-1-4244-2154-1/08/$25.00

The major problem with the reconstruction approach defined above, is that one is only allowed to use a single weight per training image. That is, all the pixels in each sample image are weighted equally. In most instances, it would be useful to be able to weight distinct areas differently. For example, assume that we have two sample images per class in our database – one with close eyes and mouth and another with open eyes and mouth. We now wish to reconstruct a test image showing open eyes and a close mouth. This can be readily achieved if we allow different weights for the top and bottom parts of each image. Hence, our goal is to define a reliable and fast mechanism to do this weight assignment.

In the present paper, we rework the general reconstruction framework defined in (1) to efficiently address this problem. By taking advantage of the natural partitions provided by the occluders, we derive a formulation that allows different weighting factors in distinct parts of the image. We demonstrate that this framework outperforms the general definition given in (1), and that the derived solution can efficiently work with a ℓ_2-minimization, resulting in very fast processing times. We present this approach in Section 2. Experimental results are in Section 3.

2. Reconstruction and Classification

Our first goal is to present the basic formulation to represent the visible pixels of a test image as a linear combination of the visible pixels in the training set. The use of different norms within this approach is then discussed.

2.1. Within-class approximation with occlusions

Let the training set for each class be, as above, $\{\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i}\}$, with $i = 1, \ldots, C$, C the number of classes. For each $\mathbf{x}_{ij} \in \mathbb{R}^p$, we define its occlusion mask $\mathbf{m}_{ij} \in \mathbb{R}^p$ as

$$m_{ijk} = \begin{cases} 1 & \text{if the pixel } x_{ijk} \text{ is not occluded} \\ 0 & \text{otherwise,} \end{cases}$$

where x_{ijk} and m_{ijk} are the k^{th} element of the p-dimensional vectors $\mathbf{x}_{ij}$ and $\mathbf{m}_{ij}$, respectively. Similarly, we define $\hat{\mathbf{m}} \in \mathbb{R}^p$ as the occlusion mask of the test vector $\mathbf{t} \in \mathbb{R}^p$.

The above notation, allows us to rewrite (1) as

$$\mathbf{w}_i = \arg \min_{(w_{i1}, \ldots, w_{in_i})^T} \left\| \mathbf{t} \odot \hat{\mathbf{m}} - \sum_{j=1}^{n_i} \mathbf{m}_{ij} \odot (w_{ij} \mathbf{x}_{ij}) \right\|_r, \tag{2}$$

where $\odot$ is the Hadamard product (*i.e.*, the element to element multiplication of two vectors, $c_k = a_k b_k$, $\mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbf{R}^p$), and $\| \cdot \|_r$ defines an appropriate metric.

As mentioned earlier, the major problem with the above equation is that it only allows for a single weight for all the pixels in the same sample image $\mathbf{x}_{ij}$.

Our solution to this problem, is to redefine the minimization procedure as a fitting process where the occlusions have been eliminated from the equations. To do this, let

$$\mathbf{M}_i = (\mathbf{m}_{i1}, \ldots, \mathbf{m}_{in_i})$$

be a $p \times n_i$ matrix whose columns are the occlusion masks of each sample in class i. Let $\mathbf{M}_{ij}$ denote the j^{th} row of this matrix. $\mathbf{M}_{ij}$ defines the sample images that can be used to reconstruct the j^{th} image pixel of $\mathbf{t}$, t_j.

Note that since each $\mathbf{M}_{ij}$ has n_i values, there are 2^{n_i} possible patterns of useful pixels to reconstruct t_j. That is, each visible pixel in the test image can be approximated by either zero pixels from the sample images (when all are occluded), a subset of them, or all of them (when there are no occlusions). Let these options be denoted by l, with $l = 1, \ldots, 2^{n_i}$.

Now consider all the non-occluded pixels in $\mathbf{t}$ that can be reconstructed using the same pattern l, and denote these pixels $\mathbf{t}^l$. Also, let $\mathbf{X}_i = (\mathbf{x}_{i1}, \ldots, \mathbf{x}_{in_i})$ be a $p \times n_i$ matrix whose columns are the samples in class i, and $\mathbf{X}_{ij}$ be the j^{th} row of $\mathbf{X}_i$. Then, the pixels in $\mathbf{t}^l$ can be reconstructed with the matrix $\mathbf{X}_i^l$, where $\mathbf{X}_{ij}$ is a row of $\mathbf{X}_i^l$ if the corresponding $\mathbf{M}_{ij}$ has the l^{th} pattern. Using this notation, we can define our reconstruction problem as

$$\mathbf{t}^l \approx \mathbf{X}_i^l \mathbf{w}_i^l, \tag{3}$$

where the weights $\mathbf{w}_i^l = (w_{i1}^l, \ldots, w_{in_i}^l)^T$ are given by

$$\arg \min_{(w_{i1}^l, \ldots, w_{in_i}^l)^T} \left\| \mathbf{t}^l - \mathbf{X}_i^l \mathbf{w}_i^l \right\|_r. \tag{4}$$

This result allows us to generate the reconstruction

$$\hat{\mathbf{t}}^l(i) = \mathbf{X}_i^l \mathbf{w}_i^l.$$

The final reconstruction of $\mathbf{t}$, obtained with the samples in class i and denoted $\hat{\mathbf{t}}(i)$, is given by the combination of all the rows in $\hat{\mathbf{t}}^l(i)$, $l = 1, \ldots, 2^{n_i}$.

The procedure described thus far provides C reconstructions of the test image, $\{\hat{\mathbf{t}}(1), \ldots, \hat{\mathbf{t}}(C)\}$, one per class. The next task is to determine which of these reconstructions is most similar to the original test image. We do this next.

2.2. Norms and quasi-norms for classification

The simplest mechanism to test how accurate the reconstruction is, would be to look at the reconstruction error, given by

$$\|\hat{\mathbf{t}}(i) - \mathbf{t}\|_r. \tag{5}$$

For example, in ℓ_2, this corresponds to the LS fitting error. In ℓ_1, it represents the error given by the sparse representation selected by the optimization mechanism.

However, the metric used to reconstruct the image is not always the most adequate for classification. For instance, if

we employ ℓ_1 to find a sparse set of training samples to describe the test image, it is generally adequate to compare the reconstructed and original images using the Euclidean distance (i.e., in ℓ_2). In this case, while a ℓ_1-minimization provides the advantages of a sparse representation, ℓ_2 is adequate for comparing images (or, equivalently, vectors). This is in fact one of the most used approaches.

In contrast, in our approach derived above, Eqs. (3)-(4), it makes sense to do the minimization in ℓ_2, because the goal is to use as much information from each sample of the *same class* as possible in an attempt to get more accurate reconstructions of the test image. However, once the reconstruction is obtained, it is generally prefer to compare the reconstruction and the original test image using a metric which emphasizes the overall similarity. Recall that the 2-norm is not a good choice for that, because it emphasizes those distances that are large while diminishing those that are small. This is a consequence of the quadratic term, which emphasizes large components and minimizes small ones. This is the same as saying that we would like to validate or invalidate a reconstruction based on the similarity of those areas that are most dissimilar, rather than those that are most similar. In fact, the 2-norm is well known to be sensitive to outliers (*i.e.*, the large distances), which are typically found in the types of reconstructions obtained with a linear fit. The area of robust statistics usually employs the 1-norm to resolve these issues.

In our application too, we can use the 1-norm to be robust to outliers and to base our judgment on the overall similarity. The advantage of this norm is that it does not emphasize the large or the small distances, since it is simply given by

$$\|\mathbf{a}\|_1 = |a_1| + \cdots + |a_p|,$$

where $\mathbf{a} \in \mathbb{R}^p$. Note that, in this norm, all components are treated equally – regardless of their size.

We can now go one step further and use a measure that deemphasizes large distances while emphasizing small ones. This would help put more emphasis on the similarity between the reconstructed and original test image, rather than on their dissimilarity as in ℓ_2. To achieve this we can use the .5-quasi-norm, given by

$$\|\mathbf{a}\|_{.5} = \left(a_1^{.5} + \cdots + a_p^{.5}\right)^2.$$

Recall that this is not a norm, because it does not satisfy the triangular inequality, which needs to be replaced by $\|\mathbf{a} + \mathbf{b}\|_r \leq K(\|\mathbf{a}\|_r + \|\mathbf{b}\|_r)$, for some $K > 1$. In our case, $r = .5$ and $K = 2$.

The important concept here, is that the .5-quasi-norm will deemphasize large distances (including the outliers) and emphasize the small ones (*i.e.*, the areas where the reconstruction was possible). This effect is due to the fact that the exponential term in the .5-quasi-norm is smaller than 1. One could use an $r < .5$, but this would only emphasize points of agreement and would no longer consider the overall similarity of the two images.

Figure 1. Two examples of face color detection and occlusion mask generation. In each row, from left to right: the full face, the face mask of the full face, the partially occluded face, the face mask of the occluded face and the final face occlusion mask.

These are thus the two measures that we use for classification, *i.e.*, the 1-norm and the .5-quasi-norm. And, the class label, c_t, of our testing image $\mathbf{t}$ is given by

$$c_t = \arg \min_{i=1,\ldots,C} \|\hat{\mathbf{t}}(i) - \mathbf{t}\|_r, \tag{6}$$

where r is either 1 or .5.

2.3. Computing the occlusion mask

As the reader will have noted in our formulation above, we require of the occlusion masks $\mathbf{m}_{ij}$ and $\hat{\mathbf{m}}$, which specify the pixels that are occluded in each image. We now present an algorithm to calculate $\mathbf{m}_{ij}$ and $\hat{\mathbf{m}}$.

In face detection and segmentation, color is a practical cue for robust detection, because human skin color can be reliably modeled using statistical methods [3, 7]. Detection of an occlusion is a bit trickier, because we need to distinguish between these and the background. Here, we adopt a variant of the approach presented in [4]. After a principal component analysis face detection step, the approach models the skin color using a mixture of Gaussians and then employs morphological operators to tune the result. Two examples on non-occluded faces and two on occluded faces are given in Fig. 1. In the last step, the morphological operators of erosion and dilation are used to eliminate isolated segments and refill the eroded local areas, respectively. After these steps, we can delineate the face limits with the use of the color map previously obtained. The right and left most pixels with skin color for each image row and the top and bottom pixels in each column are used to achieve this. The result of the entire process is shown to the right of each of the occluded image examples in Fig. 1. The final occlusion detection results shown in the right-most image of each row in Fig. 1 are obtained after a second round of morphological erosion and dilation, where 0s (black) represent occluded facial pixels and 1s (white) non-occluded. The binary occlusion map is vectorized to get the occlusion masks $\mathbf{m}_{ij}$ and $\hat{\mathbf{m}}$.

Note that the process described above also determines the limits of the face. This will be later used to separate the face from its background – to prevent misclassifications due to background noise.

3. Experimental Results

We refer to our approach presented in the above as the Partial Within-Class Match (PWCM$_r$) method, where the subscript r specifies the metric (or quasi-metric) used in (5) and (6). Our approach is to use the 2-norm in (4) and the 1-norm or .5-quasi-norm in (5) and (6). Nonetheless, for comparison purposes, we also provide results where we have employed the 2-norm in (5). Our results are consistent with our theoretical argument and, hence, the 1-norm in (5) results in superior results to those of the 2-norm. We have also experimented with the use of the 1-norm in (4) and the 2-norm in (5). This resulted in slightly worse results than the ones reported below and with the added disadvantage of a high computational cost – typically, a 10-fold increase.

3.1. Database and experimental settings

In the present paper, we employ the AR face database [5], which is one of the most popular databases, and one of the very few to include natural occlusions. The AR face database consists of more than 100 people's frontal-view color images. Other image variations include different illuminations and distinct facial expressions. This database is considered very challenging, since $\sim 50\%$ of the images have large or very large occlusions. The first 13 images, for one of the subjects in the database, are shown in Fig. 2(a-m). These correspond to the images taken during a first session. Another set of 13 images taken under the same occlusions, illuminations, and expressions was taken two weeks after the first session. We will refer to the images in this second session as *a*' to *m*'. The images in the first session are labeled *a* to *m* following the notation shown in Fig. 2.

In our experiments, we first detect and warp the face (without the inclusion of any hair or background) to generate the registered face image with a fixed size, and calculate the occlusion mask for each of them (as described in Section 2.3). This localization and warping process is known to improve recognition rates [4]. Then, we convert the face images to gray-scale and resize them to an $a \times b$ size. These a and b are selected in each experiment to match those used by other authors. This facilitates a direct comparison to a large number of results reported in the literature. We randomly choose 100 persons (50 male and 50 female) from the database. Each face image is segmented using an oval-shaped masked as shown in Fig. 2(n).

3.2. Synthetic occlusions

Experiment 1: The first experiment is to test the proposed algorithm under synthetic occlusions and different facial expressions. We use the first image (neutral face) of the first session (*i.e.*, *a*) of each subject as the training set, and the happy (*b*), sad (*c*) and scream face (*d*) from the first session for testing. The occlusion is simulated by placing a random square mask of $s \times s$ to both the training and testing images. In Fig. 3(a), we can see that the proposed algorithm can successfully handle occlusions of up to 20×20 pixels, corresponding to a $\sim 20\%$ occlusion in *both* the training and testing sets. Note that this is different to the previous results reported in the literature, since we also enforce occlusions in the training set, not only the testing one.

Fig. 3(b) shows the reverse example. That is, using the face images *b*, *c* and *d* for training and *a* for testing. Here too, the maximum allowable occlusion is of $\sim 20\%$ of the image.

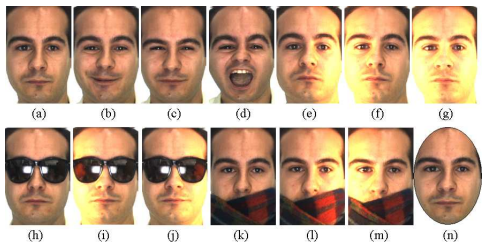

Figure 2. (a-m) Shown here are the 13 images of the first session for one of the subjects in the AR face database. (n) An oval-shaped cropped example.

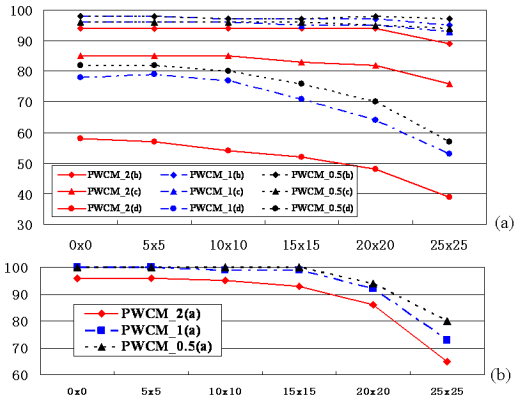

Figure 3. Classification accuracy with a synthetic occlusion mask of $s \times s$. (a) Training image *a*, testing images *b*, *c*, *d*. (b) Training images *b*, *c*, *d*, testing image *a*. The image size is 54×39.

3.3. Real occlusions

We divide the experiments in this section into two parts. The first set of experiments only considers occlusions in the testing set – allowing for a comparison with the state of the art. The second set of experiments considers occlusions in

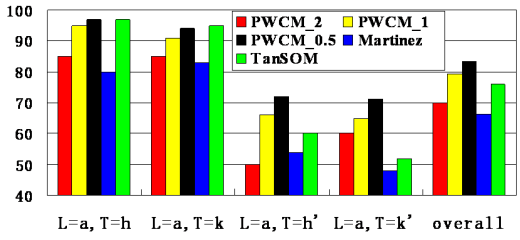

Figure 4. Successful classification rate using the proposed approach, PWCM$_r$ (with $r = 2$, 2-norm, $r = 1$, 1-norm, and $r = .5$, .5-quasi-norm). The results are compared to those in [4] and [9]. Here, $a = 170$ and $b = 120$.

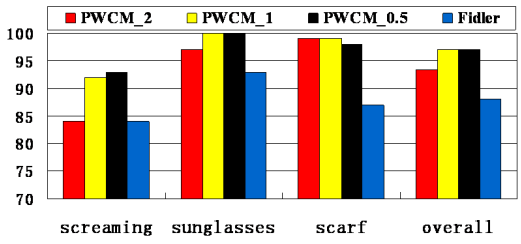

Figure 5. Training set {*a, b, c, a', b', c'*}. Testing set {*d, h, k, d', h', k'*}. Here, $a = 100$ and $b = 52$.

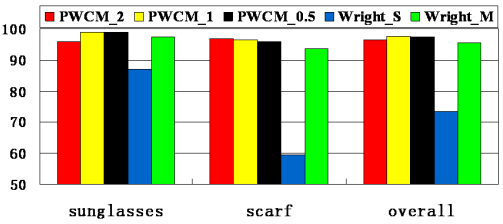

Figure 6. Training set {*a, b, c, d, a', b', c', d'*}. Testing set {*h, k, h', k'*}. Here, $a = 83$ and $b = 60$.

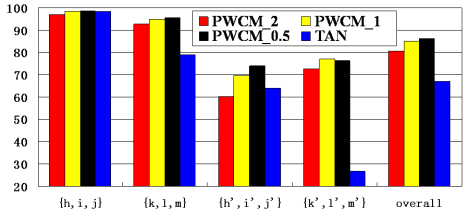

Figure 7. Training set {*a, b, c, d, e, f, g*}. Testing set {*h, i, j, k, l, m, h', i', j', k', l', m'*}. Here, $a = 66$ and $b = 48$.

both the training and testing sets.

Experiment 2: We use the neutral non-occluded face *a* for each of the 100 individuals for training, and the occluded faces *h* and *k* (of the first session) and *h'* and *k'* (of the second session) for testing. The results are compared with two other methods: the local probabilistic method of [4] and the Self-Organizing-Map (SOM) of [9]. The comparative results are in Fig. 4.

Consistent with our theory, the use of the 1-norm in (5) results in more accurate classifications. We also see that the results of the proposed approach are (on average) superior to those of [4] and [9]. The method of [9] provides comparative results with ours for those images within the same session, but not with those in the other session.

Experiment 3: Our previous experiments only considered a single training image per class. We now consider the case where the number of training images is larger, as it was done in [2, 8, 10]. In the first of these experiments, the training set is {*a,b,c,a',b',c'*} (*i.e.*, the non-occluded faces), and the testing set is {*d,h,k,d',h',k'*} (*i.e.*, the scream face and the occluded set). The experimental results are plotted in Fig. 5. In this comparison, we note that the results reported in [2] only used 50 people from the AR database. The second experiment uses {*a,b,c,d,a',b',c',d'*} as the training set and {*h,k,h',k'*} as the testing. Comparative results against the method of Wright *et al.* [10] are in Fig. 6, where "Wright_S" is a method using a single block and "Wright_M" a set of multiple blocks as defined in [10]. The method of [10] does provide comparative results with ours. However, as reported in [10], their algorithm requires about 75 seconds of processing time for a single 83×60 test image (on a PowerMac G5). The proposed PWCM algorithm achieves a slightly superior result with each image while requiring of less than 1 second of processing time. Finally, comparative results with Tan [8] are in Fig. 7.

Experiment 4: There are very little results reported in the literature where a method can deal with partially occluded faces in both the training and testing sets. In [12], Zhu and Martinez provide experimental results using several subspace algorithms when using the images of the first session in the AR database for training and the images of the second session for testing. The authors report superior results for the Subclass Discriminant Analysis (SDA) algorithm [12]. In Table 1, we show the results obtained with the proposed approach and that of SDA. We also included the results obtained with a Nearest Neighbor approach with the r-norm (denoted as NN$_r$), and those obtained when using Eq. (2) in place of (3) (denoted Within-Class Match, WCM$_r$). Again, our approach consistently outperforms the others, with the .5-quasi-norm providing the top results.

Experiment 5: Our next experiment considers the extreme cases where all the training images have partial occlusions. There are many such cases. In Table 2, we report on six possible scenarios. Since the occlusion in the AR face database each occludes almost 50% of the image, this is a very challenging case. We see that, as expected, the

SDA	$PWCM_2$	$PWCM_1$	$PWCM_{.5}$	NN_2	NN_1	$NN_{.5}$	WCM_2	WCM_1	$WCM_{.5}$
78.2	85.0	89.6	**90.6**	64.0	66.9	66.2	76.0	83.2	84.6

Table 1. Training set {*a*,*b*,*c*,*d*,*e*,*f*,*g*,*h*,*i*,*j*,*k*,*l*,*m*}. Testing set {*a'*,*b'*,*c'*,*d'*,*e'*,*f'*,*g'*,*h'*,*i'*,*j'*,*k'*,*l'*,*m'*}. Here, $a = 54$ and $b = 39$.

Training set	Testing Set	$PWCM_2$	$PWCM_1$	NN_2	NN_1	WCM_2	WCM_1
[*h*,*k*]	[*a*]	90.0	**96.0**	50.0	90.0	55.0	32.0
[*h*,*k*]	[*a'*]	60.0	**77.0**	33.0	54.0	32.0	27.0
[*h*,*k*,*h'*,*k'*]	[*a*,*a'*]	94.0	**99.0**	53.5	89.5	37.5	28.5
[*h*,*k*]	[*b*,*c*,*d*]	70.3	**87.6**	39.0	69.7	43.0	31.3
[*h*,*k*]	[*b'*,*c'*,*d'*]	44.7	**60.3**	21.7	37.7	29.0	19.7
[*h*,*k*,*h'*,*k'*]	[*b*,*c*,*d*,*b'*,*c'*,*d'*]	72.5	**90.0**	38.5	70.5	27.0	15.5

Table 2. Successful recognition rate (in percentages), with $a = 54$ and $b = 39$.

Training set	Testing Set	Image size	$PWCM_2$	$PWCM_1$
[*a*∼*m*]	[*a'*∼*m'*]	108×78	85.2	89.8
[*a*∼*m*]	[*a'*∼*m'*]	54×39	85.0	89.6
[*a*∼*m*]	[*a'*∼*m'*]	36×26	84.5	89.2
[*a*∼*m*]	[*a'*∼*m'*]	27×20	83.5	88.4
[*h*,*k*,*h'*,*k'*]	[*a*,*a'*]	108×78	94.0	99.0
[*h*,*k*,*h'*,*k'*]	[*a*,*a'*]	54×39	94.0	99.0
[*h*,*k*,*h'*,*k'*]	[*a*,*a'*]	36×26	93.5	99.0
[*h*,*k*,*h'*,*k'*]	[*a*,*a'*]	27×20	92.0	99.0

Table 3. Successful recognition rate (in percentage) obtained using the specified image size and training and testing sets.

1-norm provides superior results to the 2-norm and that our results are consistently high. The only slightly low classification rate is for the case where we train with the set {*h*,*k*} and test with {*b'*,*c'*,*d'*}. This is challenging because the images in the training and testing sets are quite distinct and correspond to different sessions. Since there is no other method available that allows a direct comparison, we show the results obtained with the nearest neighbor and the WCM approaches defined above. The results of the proposed approach are much superior.

3.4. Effect of different image sizes

Although all the algorithms aforementioned [2, 4, 8, 10, 12] use the AR database, the image size tends to be different in each experiment. This is important, because the performance of most algorithms goes up as the image size increases. In our final experiment, we demonstrate that the proposed algorithm achieves similar results to those shown above for a variety of image sizes. The results are in Table 3.

4. Conclusion

This paper introduced a new mechanism to do reconstruction of partially occluded faces. We argued for the use of the 1-norm and the .5-quasi-norm for comparison with this reconstruction. In a large number of experimental results, we demonstrated the superiority of the proposed approach to those reported in the literature. Our Matlab implementation of the algorithm classifies a new test image in less than a second.

Acknowledgments

This research was supported in part by the National Science Foundation, grant 0713055, and the National Institutes of Health, grant R01 DC 005241.

References

[1] T. C. Faltemier, K. W. Bowyer, and P. J. Flynn. A region ensemble for 3-D face recognition. *IEEE Transactions on Information Forensics and Security*, 3(1):62–73, 2008. 1

[2] S. Fidler, D. Skočaj, and A. Leonardis. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):337–350, 2006. 1, 5, 6

[3] R. L. Hsu, M. Abdel-Mottaleb, and A. K. Jain. Face detection in color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):696–707, 2002. 3

[4] A. M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002. 1, 3, 4, 5, 6

[5] A. M. Martinez and R. Benavente. The AR face database. *CVC Tech. Rep. No. 24*, 1998. 4

[6] J. S. Park, Y. H. Oh, S. C. Ahn, and S. W. Lee. Glasses removal from facial image using recursive error compensation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(5):805–811, 2005. 1

[7] S. L. Phung, A. Bouzerdoum, and D. Chai. Skin segmentation using color pixel classification: Analysis and comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):148–154, 2005. 3

[8] X. Tan, S. Chen, Z.-H. Zhou, and J. Liu. Learning non-metric partial similarity based on maximal margin criterion. *Proc. Conf. Computer Vision and Pattern Recognition 2006*, 1:168–175, 2006. 1, 5, 6

[9] 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. *IEEE Transactions on Neural Networks*, 16(4):875–886, 2005. 1, 5

[10] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *University of Illinois Tech Report UILU-ENG-08-2203 DC-233*. 1, 5, 6

[11] W. Zhao, R. Chellappa, P. J. Phillips, and A. Reosenfeld. Face recognition: A literature survey. *ACM Computing Survey*, 34(4):399–485, 2003. 1

[12] M. Zhu and A. M. Martinez. Subclass discriminant analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8):1274–1286, 2006. 5, 6