Confidence Weighted Subspace Projection Techniques for Robust Face Recognition in the Presence of Partial Occlusions*

Vitomir Štruc, Simon Dobrišek and Nikola Pavešić
Faculty of Electrical Engineering, University of Ljubljana, Slovenia
{vitomir.struc, simon.dobrisek, nikola.pavesic}@fe.uni-lj.si

Abstract

Subspace projection techniques are known to be susceptible to the presence of partial occlusions in the image data. To overcome this susceptibility, we present in this paper a confidence weighting scheme that assigns weights to pixels according to a measure, which quantifies the confidence that the pixel in question represents an outlier. With this procedure the impact of the occluded pixels on the subspace representation is reduced and robustness to partial occlusions is obtained. Next, the confidence weighting concept is improved by a local procedure for the estimation of the subspace representation. Both the global weighting approach and the local estimation procedure are assessed in face recognition experiments on the AR database, where encouraging results are obtained with partially occluded facial images.

1. Introduction

Subspace projection techniques, such as PCA [11], or LDA [2], are among the most popular face recognition techniques found in the literature. Their popularity is fueled by their computational simplicity and evidenced by the vast number of studies addressing issues related to subspace-based face recognition. Despite the enormous research effort that is being directed towards subspace projection techniques by the academic community, some issues relating largely to robust recognition from partially occluded face images still pose great challenges to the existing methods. It should be noted that we do not imply that partial occlusion of the facial area is the only open issue in face recognition. However, next to pose, expression, lighting and aging effects, it represents one of the major factors affecting the performance of the existing face recognition technology.

The susceptibility of subspace projection techniques to partial occlusion is the consequence of the occluded pixels, which, from the statistical point of view, represent outliers in the image data. As such, they affect the projection step of the subspace methods and produce subspace representations which differ significantly from the subspace representation of unoccluded images of the same subject. Or in other word, the occluded pixels induce variability to the computed subspace representation, which is the consequence of the image characteristics rather than identity. An extreme case of the effect of outliers on the subspace representation was presented in [3], where it was shown that a single outlying pixel can affect the subspace representation of an image to an arbitrarily large extent.

To overcome this susceptibility researchers have proposed several modifications to the established subspace projection techniques. Leonardis and Bishof [7], for example, proposed an alternative procedure for the estimation of subspace coefficients in the PCA subspace using a sub-sampling procedure combined with a hypothesis and test paradigm. With this approach, first a number of coefficient-hypotheses is generated using only subsets of all image pixels, and then the most probable hypothesis is selected based on the minimum description length principle. Since only a subset of all image pixels is used for the generation of each hypothesis, the approach exhibits robustness to occlusions present in the image data. The described (PCA-based) subsampling approach was later extended for use with a number of discriminative methods by Fidler et. al [5], allowing for a robust estimation of the subspace coefficients of discriminative techniques as well.

Edwards and Murase [4] introduced a modification of the PCA technique, where robustness to partial occlusion was achieved by excluding all image pixels from the subspace projection that constituted occluded image areas. They developed an (coarse-to-fine) adap-

*Partially supported by the COST Action 2101 and the EU-FP7 project 217762 Homeland security, biometric Identification and personal Detection Ethics (HIDE).

tive masking concept that enabled a robust calculation of the subspace representation. More recently, a similar idea was also presented by Jia and Martinez in [6] and successfully applied to the problem of robust face recognition in the presence of partial occlusions.

Wright et al. [12] tackled the problem of subspace-based face recognition from partially occluded images with the theory of sparse representations. Here, the authors modeled the image occlusions as a spatially sparse error, which is efficiently estimated together with the sparse subspace representation via the minimization of a L_1-norm-penalized cost function. The sparse representation based approach resulted in state-of-the-art recognition results with partially occluded face images.

The confidence weighting concept presented in this paper is applicable to all linear subspace projection techniques and makes no assumptions regarding the characteristics of the projection basis as, for example, the sub-sampling techniques presented in [7] or [5], which exploit the reconstructive capabilities of the employed subspace method. To a certain extent our approach bears similarities with the approach of Edwards and Murase [4]; however, instead of simply excluding image pixels that constitute occluded image areas (i.e., masking the image with a binary mask), it weights each image pixel according to a confidence measure that quantifies the confidence that the pixel in question represents an outlier. By doing so, it reduces (or eliminates) the impact of the occluded pixels on the subspace representation of the given face image and adds robustness to partial occlusions of the underlying subspace projection technique. To further improve the basic confidence weighting paradigm, we also present a local technique for the estimation of the subspace representations, which can again be applied to any existing subspace projection technique.

The feasibility of the proposed confidence weighting in local coefficient estimation paradigms for face recognition is assessed with three established subspace methods, namely, PCA, ICA, and LDA. The results of the experiments performed on the AR database [9] suggest that all three evaluated techniques in combination with the proposed modifications are capable of matching and in some cases even surpassing the performance of the state-of-the-art techniques for robust face recognition from the literature.

2. Confidence weighted subspace methods

2.1. Classical subspace techniques

Let $\mathbf{X} \in \mathbb{R}^{d \times n}$ denote a data matrix containing in its columns n centered (i.e., global-mean subtracted) training images (in vector form) each comprising d pixel values, i.e., $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$. Subspace projection techniques try to represent each of the training images $\mathbf{x}_i \in \mathbb{R}^d$ (for $i = 1, 2, ..., n$) as a linear combination of $d' \ll d$ vectors $\mathbf{w}_j$ (for $j = 1, 2, ..., d'$) spanning a d'-dimensional subspace. In matrix form this can be formally written as follows:

$$\mathbf{X} = \mathbf{W}\mathbf{Y}, \tag{1}$$

where the matrix $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_{d'}] \in \mathbb{R}^{d \times d'}$ contains in its columns d' vectors defining the given subspace and $\mathbf{Y} = [\mathbf{y}_1^T, \mathbf{y}_2^T, ..., \mathbf{y}_n^T]^T \in \mathbb{R}^{d' \times n}$ denotes the coefficient-vector-matrix, which in its rows contains the low-dimensional subspace representations of the n training images.

A visual example of the basic concept of subspace projection techniques is shown in Fig. 1. Note that the centered face image is represented as a linear combination of some canonical vectors, which, when shown in image form, often resemble faces.

Figure 1. Concept of subspace methods

Let us now assume that we have successfully computed a subspace $\mathbf{W}$ and we want to estimate the subspace representation of a centered test image $\mathbf{x}_t \in \mathbb{R}^d$ that was not included in the training set. The general expression for computing the low-dimensional representation of the test image $\mathbf{x}_t$ reads as follows:

$$\mathbf{y}_t = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{x}_t, \tag{2}$$

which, in case of mutually orthogonal column vectors in $\mathbf{W}$, is reduced to:

$$\mathbf{y}_t = \mathbf{W}^T\mathbf{x}_t. \tag{3}$$

Here, T denotes the transpose operator and $\mathbf{y}_t \in \mathbb{R}^{d'}$ stands for the coefficient vector corresponding to the test image $\mathbf{x}_t$.

2.2. The confidence weighted approach

To overcome the susceptibility of subspace projection techniques to partial occlusion, we here propose a confidence-based weighting scheme, which reduces the impact of the occluded pixels, i.e., the statistical outliers, on the estimated subspace representation, i.e., the coefficient vector[1]. While the concept of pixel

[1] Recall that it was shown in [3] that even a single outlying pixel value can affect the subspace representation of an image to an arbitrarily large extent

Figure 2. The basic concept of confidence weighted subspace projection techniques

weighting is not novel (see, for example, [3]), its use so far was (to the best of our knowledge) limited only to the subspace creation step. In this section, we extend the weighting concept to robust coefficient-vector-estimation and present the advantages of such a scheme.

Let $\mathbf{x}_t$ denote a centered test image (containing occlusions), for which we want to compute the subspace representation. If we characterize each pixel in the test image using a confidence measure that takes values near one for valid pixels and values near zero for pixels that represent statistical outliers, then we can define a novel test image $\mathbf{x}'_t$ and a new subspace $\mathbf{W}' \in \mathbb{R}^{d \times d'}$ as follows:

$$\mathbf{x}'_t = \mathbf{c} \odot \mathbf{x}_t, \tag{4}$$

and

$$\mathbf{W}' = [\mathbf{c} \odot \mathbf{w}_1, \mathbf{c} \odot \mathbf{w}_2, ..., \mathbf{c} \odot \mathbf{w}_{d'}], \tag{5}$$

where $\odot$ denotes the Schur or Hadamard (i.e., element-wise) product, $\mathbf{c} \in \mathbb{R}^d$ stands for the vector of confidence measures corresponding to the test image $\mathbf{x}_t$, and the Schur products $\mathbf{c} \odot \mathbf{w}_i$ (for $i = 1, 2, ..., d'$) are normalized to unit length. If we arrange the confidence measures in $\mathbf{c}$ into a diagonal confidence matrix $\mathbf{C}$, i.e., $\mathbf{C} = \mathrm{diag}([c_1, c_2, ..., c_d]) \in \mathbb{R}^{d \times d}$, then Eq. (5) can be rewritten as follows:

$$\mathbf{W}' = \lfloor \mathbf{C}\mathbf{W} \rfloor_{L_2}, \tag{6}$$

where $\lfloor \cdot \rfloor_{L_2}$ denotes an operator that normalizes each column of its argument to unit norm.

The confidence weighted subspace projection is then performed using the following expression:

$$\mathbf{y}_t = (\mathbf{W}'^T \mathbf{W}')^{-1} \mathbf{W}'^T \mathbf{x}'_t. \tag{7}$$

Note that even for initially orthogonal subspaces, such as the PCA subspace, the confidence weighting results in non-orthogonal subspace vectors comprising the matrix $\mathbf{W}'$. Thus, the general procedure for computing the subspace representation (Eq. (2)) has to be used instead of the simplified one given in Eq. (3).

A visualization of the presented confidence weighting concept is shown in Fig. 2. Here, the squares in the lower right corner relate to the local approach presented in Section 3 and should be ignored at this stage. We can see that the pixel values corresponding to outliers (i.e., the occluded pixels) are weighted with a confidence measure near zero in both the original image as well as the subspace vectors. Hence, the effect of the occluded pixels on the estimate of the subspace representation is reduced significantly. The theoretical justification for such weighting stems for the fact that the linear combination must be true for every pixel. However, since the subspace representation is computed in the least square sense (see Eq. (2)), it is susceptible to outlying values. Thus, by weighting the outlying pixels with a small weight, we obtain a better estimate of the true subspace representation of the given image.

2.3. Outlier detection

To be able to apply the confidence weighting scheme to the problem of face recognition, we require an outlier detection procedure, on the basis of which we can derive a confidence measure quantifying the "outlyingness" of each image pixel. To this end, we turn to the work of De la Torre and Black [3]. The authors presented a PCA-based outlier detection procedure for their robust subspace learning framework. From their approach we can directly derive a measure of confidence for the validity of each image pixel. We define the confidence vector $\mathbf{c}$ as follows:

$$\mathbf{c} = \mathbf{1}_d - f(\boldsymbol{\sigma}), \tag{8}$$

where

$$\sigma_i = \max(1.4826 \cdot \mathrm{med}_\Omega(|\mathbf{e}_i - \mathrm{med}_\Omega(|\mathbf{e}_i|)|), \sigma_e), \tag{9}$$

and $f(.)$ denotes a linear scaling procedure that maps the values of its argument vector to the interval $[0, 1]$, $\mathbf{1}_d$ denotes a d-dimensional vector of all ones, med_Ω stands for the median operator applied to the region Ω around the i-th pixel, $\mathbf{e}_i$ stands for the vector of PCA reconstruction errors in the i-th pixels neighborhood Ω (obtained with a PCA basis that preserved approximately 55% of the variance in the training data), σ_e represents the median absolute deviation taken over the entire image, and last but not least, $\boldsymbol{\sigma} \in \mathbb{R}^d$ is defined as $\boldsymbol{\sigma} = [\sigma_1, \sigma_2, ..., \sigma_d]$. Two examples of confidence vectors computed from two images of the AR database are shown in the second column of Fig. 3.

3. The local extension

The global confidence weighting approach presented in the previous section should already result in enhanced recognition performance when compared to the classical subspace methods. However, since face recognition is notorious for its lack of training data, the high-confidence pixels (which represent only a subset of all image-pixels) may exhibit an insufficient correlation with the training data and may still result in a poor estimation of the true subspace representation.

We overcome this problem by applying a local procedure for the estimation of the subspace representation of the given face image similarly as it was done by Leonardis and Bishof for the case of PCA in [7]. Nevertheless, our approach differs from the one in [7], as we rely on the concept of confidence weighting and are, hence, not bound to any specific subspace method.

Let us assume that we have a confidence weighted test face image $\mathbf{x}_t'$ and in general a non-orthogonal confidence weighted subspace $\mathbf{W}'$. We estimate the subspace representation of $\mathbf{x}_t'$ based on coefficient vectors computed from local blocks of the image $\mathbf{x}_t'$, where the number of pixels m in each block has to be larger than the dimensionality of the subspace representation d'. An example of the image blocks used for computing one estimate of the subspace representation is shown in Fig. 2 - the lower right corners of the images.

With this local estimation procedure we generate k initial hypotheses $\mathbf{y}_1^{h_i}, \mathbf{y}_2^{h_i}, ..., \mathbf{y}_k^{h_i}$ for the subspace representation (i.e., coefficient vector) using image blocks of different sizes and extracted from different locations. We assign every hypothesis a confidence measure ρ_i ($i = 1, 2, ..., k$), which represents the mean value of the elements of $\mathbf{c}$ corresponding to the currently observed image block. Next, we order the hypotheses according to their confidence and discard all hypotheses with a confidence ρ_i smaller than the median of the hypothesis-confidence-vector $\boldsymbol{\rho}_t$, i.e., we discard ρ_i if:

$$\rho_i < \rho_{thresh} = \text{med}(\boldsymbol{\rho}_t), \qquad (10)$$

where $\boldsymbol{\rho}_t = [\rho_1, \rho_2, ..., \rho_k]^T$.

In this way, we obtain a new set of p subspace representation hypotheses $\mathcal{Y}_h = \{\mathbf{y}_i^h, i = 1, 2, ..., p\}$, where $p < k$. The final subspace representation $\mathbf{y_{tf}}$ is ultimately computed as the median of the remaining hypotheses in $\mathcal{Y}_h$, i.e.,

$$\mathbf{y}_{tf} = \text{med}([\mathbf{y}_1^{(h)T}, \mathbf{y}_2^{(h)T}, ..., \mathbf{y}_p^{(h)T}]^T). \qquad (11)$$

We can visually assess the appropriateness of our proposed confidence weighting schemes using the PCA subspace. Since PCA is best suited for data representation, we can use the subspace representation estimated with our procedure and reconstruct the input images with the original (un-weighted) subspace vectors. The results tell us how good our estimate of the coefficient vector is. An example of the reconstruction of two images from the AR database is shown in Fig. 3. Here, the first column represents the input images, the second column represents the confidence vectors and the third and fourth columns depict the reconstructed images using the presented global and local approaches, respectively.

Figure 3. Effect of confidence weighting

4. Experiments and results

The proposed confidence weighting (CW) scheme is assessed on the popular AR face database [9], which contains more than 4000 facial images that correspond to 126 subjects. Following the experimental setup of [8] and [5], we select a subset of images corresponding to 50 subjects (25 males and 25 females), geometrically normalize the images according to manually marked eye coordinates and crop the facial region to the predefined size of 100×100 pixels. We select 300 training images (6 per subject) depicting the neutral, smile and anger expressions, to build our subspaces and use 200 images (4 per subject) for testing. The test images are taken from the "scarves" (100 test images) and "glasses" (100 images) subsets of the database.

We use the Pearson correlation for matching the computed subspace representations to the mean subspace representations of the training images of each subject. We implement three popular subspace projection techniques, i.e., PCA [11], ICA [1] and LDA [2], and subject them to our global confidence weighting procedure and local estimation approach (denoted as GCW and LCW, respectively). For comparison purposes we also provide results from other researchers. In particular, we provide results for the ARG approach of [10], the approach of Martinez in [8] - MAR, the technique of Fidler et al. [5] - FID, and the global and local sparse representations of [12] - G-SRC and L-SRC.

The results of the experiments are presented in Table 1. Note that both the global and local CW schemes significantly improve upon the baseline performance of the un-weighted subspace projection techniques. While the relative ranking of the subspace techniques with the two proposed modifications remained the same as in the case of un-weighted subspace methods for the majority of experiments, there are, nevertheless, some points that

Table 1. Recognition rates for the challenging ”glasses” and ”scarves” subsets - AR database

Method	ARG [10]	MAR [8]	FID [5]	G-SRC [12]	L-SRC [12]	PCA	ICA
glasses	80.7%	80.0%	84.0%	87.0%	97.5%	54.0%	52.0%
scarves	85.2%	82.0%	93.0%	59.5%	93.5%	41.0%	47.0%
Method	LDA	GCW-PCA	GCW-ICA	GCW-LDA	LCW-PCA	LCW-ICA	LCW-LDA
glasses	61.0%	86.0%	85.0%	92.0%	85.0%	79.0%	89.0%
scarves	70.0%	58.0%	62.0%	88.0%	99.0%	95.0%	97.0%

require a more detailed discussion.

Note that all implemented subspace techniques combined with the global CW scheme perform better on the “glasses” subset than on the “scarves” subset, while the local CW approach ensures better recognition rates on the “scarves” subset than on the “glasses” subset with all three subspace techniques. This result can mainly be ascribed to the fact that the global CW scheme uses the entire (weighted) face image to compute its subspace representation, and, hence, performs better if less pixels are occluded[2]. The local scheme, on the other hand, relies on local image blocks to compute the subspace representation, and, when doing so, adopts only the blocks with the highest confidence for computation (see Eq. (10)). Most of the outliers in the images of the “glasses” subset are located in the eye-area, therefore, the local approach focuses on the lower part of the face during estimation of the subspace representation. However, as the face images are registered based on manually marked eye-coordinates, the lower face areas are not aligned properly, introducing additional errors into the process of estimating the true subspace representation. The result of this setting is a small drop in the recognition performance of the local CW techniques on the “glasses” subset and a significant increase on the “scarves” subset, when compared to the global CW approaches. In any case, these results suggest that with face images registered on the basis of eye and mouth locations, the recognition rates for both proposed approaches could be even further improved upon.

Last but not least, let us take a look at the performance of our better performing method, i.e., the local CW approach, in comparison with techniques from the literature. We can notice that on the “scarves” subset, the proposed technique results in a competitive performance, easily surpassing all state-of-the-art techniques considered. On the “glasses” subset, however, our technique outperforms most methods from the literature, except for the L-SRC approach, whose performance could only be matched with facial images registered on the basis of eye as well as mouth locations.

[2]Note that in the “glasses” subset approximately 20% of each image is occluded, while in the “scarves” subset the percentage of occluded pixels in an image can easily surpass 40%.

5. Conclusion

We have presented a confidence weighting scheme for subspace methods. The proposed scheme significantly improved the performance of the traditional subspace methods in the presence of partial occlusions. When combined with a local procedure for the estimation of the subspace representation, the scheme resulted in a competitive performance when compared to the state-of-the-art methods from the literature.

References

[1] M. Bartlett, J. Movellan, and T. Sejnowski. Face recognition by independent component analysis. *IEEE Trans. on Neural Networks*, 13(6):1450–1464, 2002.

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

[3] F. de la Torre and M. Black. A framework for robust subspace learning. *Int. J. Comp. Vis.*, 54:117–142, 2003.

[4] J. Edwards and H. Murase. Coarse-to-fine adaptive masks for appearance matching of occluded scenes. *Machine Vis. and Appl.*, 10(5-6):232–242, 1998.

[5] S. Fidler, D. Skočaj, and A. Leonardis. Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling. *TPAMI*, 28(3):337–350, March 2006.

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

[7] A. Leonardis and H. Bishof. Robust recognition using eigenimages. *Com. Vis. Im. Und.*, 78(1):99–118, 2000.

[8] A. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from single sample per class. *TPAMI*, 24(6):748–763, June 2002.

[9] A. Martinez and R. Benavente. The ar face database. *CVC Technical Report*, (24), June 1998.

[10] B. Park, K. Lee, and S. Lee. Face recognition using face-arg matching. *TPAMI*, 27(12):1982–1988, 2005.

[11] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.

[12] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma. Robust face recognition via sparse representation. *TPAMI*, 31(2):210–227, February 2009.