Pose Normalization for Local Appearance-Based Face Recognition

Hua Gao, Hazım Kemal Ekenel, and Rainer Stiefelhagen

Computer Science Department, Universität Karlsruhe (TH)
Am Fasanengarten 5, Karlsruhe 76131, Germany
{hua.gao,ekenel,stiefel}@ira.uka.de
http://isl.ira.uka.de/cvhci

Abstract. We focused this work on handling variation in facial appearance caused by 3D head pose. A pose normalization approach based on fitting active appearance models (AAM) on a given face image was investigated. Profile faces with different rotation angles in depth were warped into shape-free frontal view faces. Face recognition experiments were carried out on the pose normalized facial images with a local appearance-based approach. The experimental results showed a significant improvement in accuracy. The local appearance-based face recognition approach is found to be robust against errors introduced by face model fitting.

1 Introduction

Most of the current face recognition technologies are based on the appearance of face. One of the typical problems, where the appearance-based recognition approaches fail, is 3D pose variation. As the human face is a 3D object, its visual appearance under different view angles varies significantly. In [2], it has been shown that facial appearance variations caused by different head poses is larger than the ones caused by different identities. The large difference in appearance makes face recognition difficult when the face pose in the probe image differs from the one in the gallery image.

One way to compensate the pose variation is to add enough training samples that cover the pose-space of the gallery set. This sample augmentation scheme is adopted in many video-based face recognition systems [24] since the multi-pose samples can be obtained in video recordings. But in practice, it is not possible to cover the whole pose-space for each subject in the gallery. Moreover, enlarging the size of gallery set always slows down the system. For some recent application scenarios, such as identifying faces in the images from the Internet [16], this problem is even more significant because only limited number of gallery images are available. Pose normalization is one of the ways to solve the pose problem.

Chai et al. [5] presented a statistical affine transformation for pose normalization. They first estimated the face pose with a pose subspace algorithm, and divided the face region into three rectangles. An affine transformation was performed independently in the three rectangles and the transformation parameters

were statistically learned from the correspondence information between the specific pose and the frontal face. Experiments on the FERET database [20] showed that the recognition rate increases significantly compared to recognition without pose correction.

In [7], the active appearance model (AAM) was proposed as a model-based approach for face alignment and pose normalization. The shape of the model is optimized by minimizing the texture representation error. Once the model is fitted on an input image, the optimized model parameters can be used for face recognition [9]. Later in [8], the authors found that the shape parameters of AAMs can be used to estimate the pose angle. A frontal view of an input face image can be synthesized by configuring the shape parameters that control the pose. Instead of synthesizing a frontal face using the optimized model texture, the authors in [15] warp the texture inside the fitted model shape to the pose corrected shape. This texture warping approach has the advantage of preserving the textural information such as moles and freckles contained in the original image, which will be lost in the synthesis-based approach where the model parameters only represent the principal components of the appearance.

Blanz et al. [3] extended the 2D AAM to a 3D morphable model (3DMM). The vertices in a 3DMM shape are much more dense than an AAM shape. The dense shape model enables a high resolution face representation and the 3D shape enables linear transformation in 3D space in a more natural way. The 3DMM achieved promising results for illumination and pose-invariant face recognition. However, fitting a dense model requires much higher computational effort, which is not applicable for real-time face recognition systems.

In this paper, we investigated a pose normalization approach based on fitting 2D AAMs. The simultaneous inverse compositional (SIC) fitting algorithm [14] was employed for fitting generic models. To initialize the model more precisely and robustly against cluttered backgrounds, a progressive model fitting scheme was used. Different from [15], we generated shape-free face images as a pose normalization step, in which all the salient facial feature points were warped to a common coordinate framework. Different texture warping methods were studied and the local appearance-based face recognition approach was performed on the pose-normalized face images. We conducted our experiments on the FERET b-series database with a setup similar to the one in [5]. However, the profile face images with all rotation angles were evaluated, while in other studies not all of them were used [5,22,21]. Furthermore, we evaluated all the images including those where the AAM fitting failed, to show the robustness of the face recognition algorithms. Experiments show that local appearance-based approach is more robust against the errors introduced by model fitting and texture warping.

2   Active Appearance Models and Model Fitting

An active appearance model (AAM) is a generative parametric model which utilizes both shape and appearance information to represent a certain object such as the human face. A shape in AAM is defined as a set of normalized

facial landmarks. To obtain enough variations in both shape and appearance, a set of representative facial images should be collected for training an AAM. Each facial image is manually labeled with a set of 2D landmarks, $[x_i, y_i], i = 1, 2, ...v$. A Procrustes analysis [13] is performed to make the shapes invariant to Euclidean transformation. Eigen-analysis is applied to reduce the redundancy in the normalized shape set and the resulting linear shape model represents a shape instance as $s = s_0 + \sum_{i=1}^{n} p_i s_i$, where s_0 is the mean shape, s_i is the ith shape basis, and $\mathbf{p} = [p_1, p_2, \ldots, p_n]$ are the shape parameters.

After building the shape model, each facial image is warped into the mean shape using a piecewise affine transformation. The transformation results in shape-free facial appearances which allows texture comparison in the coordinates of the mean shape. Eigen-analysis is performed on the shape-free appearances from all training images and the resulting model represents an instance appearance as $A = A_0 + \sum_{i=1}^{m} \lambda_i A_i$, where A_0 is the mean appearance, A_i is the ith appearance basis, and $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_m]$ are the appearance parameters.

The original AAM [7] learned the correlations between the shape model and appearance model by performing another eigen-analysis on the concatenated shape and texture parameters. This combined model is more compact. However, the choice of model fitting algorithms is restricted. Only numerical approaches are suitable for fitting the combined model. The authors in [18] proposed a more precise analytical approach for fitting AAMs using independent models in the framework of the inverse compositional (IC) image alignment algorithm. The IC algorithm is efficient and effective for fitting a person-specific AAM. However, the fitting accuracy degrades significantly on unseen faces. They extended their work in [14] for fitting a generic model to unseen persons with the simultaneous inverse compositional (SIC) algorithm, in which the shape parameter $\mathbf{p}$ and appearance parameter λ are optimized simultaneously during the search.

Given an input facial image $I(\mathbf{x})$, the goal of fitting an AAM is to find the optimal model parameters such that the synthesized model instance is as similar to the image observation as possible. This leads to a minimization of a cost function defined as,

$$E = \sum_{\mathbf{x} \in s_0} \left[I(\mathbf{W}(\mathbf{x}; \mathbf{p})) - A(\mathbf{x}, \lambda) \right]^2, \tag{1}$$

where $I(\mathbf{W}(\mathbf{x}; \mathbf{p}))$ is the warped input facial image, and $A(\mathbf{x}, \lambda)$ is the synthesized appearance instance. The minimization problem is usually solved by iteratively estimating the incremental update of the shape parameter $\Delta\mathbf{p}$ and the appearance parameter $\Delta\lambda$ and updating the current model parameters respectively. The IC and SIC algorithm, however, switch the role of the appearance template and the input image when computing $\Delta\mathbf{p}$. The shape is updated by composing an inverse incremental warping $\Delta W(\mathbf{x}; \mathbf{p})$ which is estimated where $\mathbf{p} = 0$. This IC framework enables the time-consuming steps of parameter estimation to be pre-computed and performed outside of the iterations.

It is known that the gradient-descent-based optimization problem usually requires a reasonable initialization. A poor initialization may cause the search to
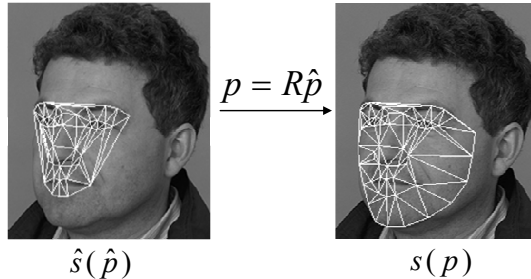
$$p = R\hat{p}$$

$$\hat{s}(\hat{p}) \qquad\qquad s(p)$$

Fig. 1. Progressive AAM fitting

stuck in a local minimum. By utilizing some prior-knowledge of the face in question, the face shape can be initialized by transforming the mean shape close enough to the target. This initialization usually suffices for fitting frontal faces. However, when fitting semi-profile faces, part of the initialized shape cannot even cover the face. Thus the optimization can be affected by the included background pixels. Inspired by the work in [17], an inner-face model that only models the inner part of the face, i.e. eyes, nose, mouth, is built. The fitting is then performed progressively in two stages. In the first stage, only the inner-face model is initialized and fitted. In the second stage, the initial shape parameter $\mathbf{p}$ of the whole-face model is estimated from the obtained inner-face shape parameter $\hat{\mathbf{p}}$ with $\mathbf{p} = R\hat{\mathbf{p}}$, here we assume R is constant and it is obtained by performing linear regression on the training data of both models. A second fitting is carried out to fit the whole-face model using the initial parameters $\mathbf{p}$. Figure 1 depicts the bi-stage AAM fitting scheme, and it provides robust fitting accuracy in [17].

3   Pose Normalization

The most straightforward method to normalize the pose of a face image is the piecewise affine warping which is also used in the fitting iteration for sampling the texture inside the face mesh. The warp is realized by mapping the pixels in the fitted triangular mesh s to the base mesh s_0. For each pixel $\mathbf{x} = (x, y)^T$ in a triangle in the base mesh s_0, it can find a unique pixel $\mathbf{W}(\mathbf{x}; \mathbf{p}) = \mathbf{x}' = (x', y')^T$ in the corresponding triangle in the mesh s by an affine mapping in the barycentric coordinates. The implementation for the piecewise affine warp is detailed in [18].

The piecewise affine warping is simple but the deformation field is not smooth. Straight lines may usually be bended across triangle boundaries. To overcome this discontinuity, we investigated another warping technique based on thin-plate splines (TPS) [4] to improve the smoothness of the deformation. However, the behavior of the method is not always clear in-between feature points, especially in the case of large pose variations.

4   Face Recognition

The pose normalized shape-free facial images are masked with the AAM mean shape mask. All salient facial features are warped to the canonical locations. However, feature points around the chin area might be misaligned, which may create strong edges. As demonstrated in [11], the chin area does not contribute too much discriminative information compared to other facial features. For this reason, we decide to crop the chin area in the pose normalized facial image. Following the approach in [10], we scaled the cropped images to 64×64 pixels size and then divided them into 64 non-overlapped blocks of 8×8 pixels size. On each local block, the discrete cosine transform (DCT) is performed. The obtained DCT coefficients are ordered using a zig-zag scanning. The first component is skipped because it represents the average pixel intensity of the entire block. The following ten low frequency coefficients are retained which yields a ten dimensional local feature vector. Finally, the 64 local feature vectors are concatenated to construct the feature vector of a whole face image.

The DCT preserves the total image energy of the processed input block; therefore blocks with different brightness levels lead to DCT coefficients with different magnitudes. In order to balance each local block's contribution to the classification, the local feature vector is normalized to unit norm. To balance the contribution of each DCT frequency, each DCT coefficient is divided by its corresponding standard deviation [12] before the unit normalization. Nearest neighbor with L1 distance measure is used for classification.

5   Experiments

We built a generic AAM to fit it on all facial images of different subjects. The study in [14] showed that building generic AAMs is far more difficult than building person-specific AAMs. In order to reduce the effects of reconstruction errors, more training images are needed to cover all possible variations in pose, illumination, expression and identity.

In this work, we collected a set of 649 images from four publicly available databases: the IMM database [1], the ND1 database [6], the CMU PIE database [23], and the FERET database. The example images selected from these databases are displayed in Figure 2. Each training face image was manually labeled with 58 landmarks. The manual labeling is a time-consuming operation and is error-prone due to the accuracy limitations of a manual operation. Moreover, the interpretation of correct landmark locations differs between annotators. The inconsistent labeling introduces noise in the shape data which affects the face modeling and the performance of fitting, consequentially. To solve this problem, we employed an iterative data refitting method as suggested in [14] to eliminate the noise in the manual labels.

The face recognition experiments were carried out on the FERET b-series which contains 200 subjects, where each subject has nine images corresponding to different pose angles. Sample images of a single subject are shown in the first

row of Figure 3. The label *ba* indicates the frontal series, which was used as the training set (gallery). The remaining eight series are non-frontal face images with different pose angles ($\pm 60^\circ$, $\pm 40^\circ$, $\pm 25^\circ$, and $\pm 15^\circ$). All except the *ba* series were used for testing (probe). None of the subjects used for training the AAM were used during the recognition experiments.

The example pose normalized images are depicted in the second row of Figure 3. The automatic AAM initialization is not always stable because of the large rotation angle in some series such as *bb* and *bi*. Automatic facial feature detectors fail to detect eyes and mouth, which are important for a precise AAM initialization. In this sense, we used a semi-automatic AAM initialization where the annotation of eye-centers and mouth-centers were used for model initialization.

In general, fitting near-frontal faces performs better than fitting semi-profile faces. When fitting semi-profile faces, even a small misalignment in the chin area may cause a large error in the warped face image. The reason is that the partially
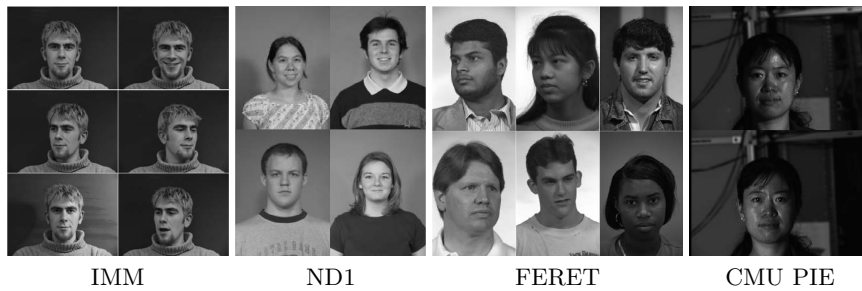

IMM   ND1   FERET   CMU PIE

Fig. 2. Sample images from IMM, ND1, FERET and CMU PIE face databases for training a generic model

ba   *bb*   *bc*   *bd*   *be*   *bf*   *bg*   *bh*   *bi*

Fig. 3. The first row shows the example images from the FERET *b*-series. The face images in the series *ba* are all frontal. The series $b[b - i]$ contains faces in different yaw pose-angles: *bb* (60°), *bd* (40°), *bc* (25°), *be* (15°), *bf* (-15°), *bg* (-25°), *bh* (-40°), and *bi* (-60°). The second row shows the corresponding fitted faces with pose normalized.

self-occluded face part is over-sampled during the warping, which enlarges the misalignment error. Another fact which we have noticed is that even if a semi-profile face is fitted perfectly, the warped frontal-view face still looks different from the real frontal face. Take the image labeled *bi* in Figure 3 for example: the left half-face is over-sampled and the right half-face is down-sampled after texture warping. This effect makes the left-eye wider and the right-eye narrower, which results in a different local appearance compared to the training image *ba* in Figure 3.

We carried out experiments with different face warping methods. First, the piecewise affine warp was considered. The original piecewise affine (PA) warp stretches zero-mean unit-length texture vector to intensity values in [0-255]. In addition to PA, piecewise affine warping with original pixel intensities was also considered and we named it PAN (Piecewise Affine warp No stretch). The third method (HIST) was also based on the piecewise affine warp, however, the histogram of the warped face was matched to the histogram of the mean face. In addition, the TPS-based warping technique was also evaluated. The face recognition results based on these four face warping approaches are plotted in Figure 4(a). Note that the recognition performances of these four methods are similar on series *be* and *bf*, where the face rotation angles are $\pm 15°$. With the increasing degree of face rotation in depth, the recognition results differ from each other. Face warping with PAN outperforms the others in most cases. The performance of the other piecewise affine-based methods decreases dramatically when the fitting performance decreases due to large degree of face rotation. The reason for this behavior is that both methods change the histogram of the original input face image. In case of misalignment of the outline shape points, the background pixels may affect the histogram in the face area. The TPS warping, however, does not help in terms of face recognition. It achieves the worst results when testing on *bi*, *bh* and *bf*.
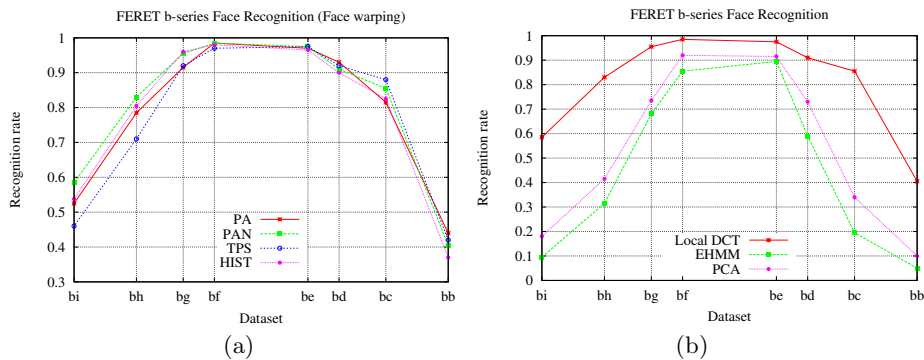

(a)   (b)

Fig. 4. Face recognition on the FERET *b*-series. (a) Face recognition rate with different face warping approaches. (b) Comparison of different face recognition techniques on pose-normalized face images with PAN face warping.

In addition to the different face warping methods, we also compared various face recognition techniques. Since the training set only contains one example per subject, the selection for other face recognition techniques is limited. Techniques which utilize intra-class information are not suitable for this evaluation. We selected two well-known face recognition approaches: the holistic Eigenfaces approach [25] and embedded hidden Markov models (EHMM) [19], which utilize DCT coefficients as local features as well. The selected algorithms were tested on face images warped with the PAN method. Figure 4(b) plots the corresponding recognition rates. Local DCT denotes the local appearance-based face recognition approach. For the Eigenfaces approach, 40 eigen-components were used, which yielded optimal performance. We noticed that Local DCT outperformed the other two approaches significantly over all the eight series. Moreover, the recognition rates of EHMM and PCA decreased drastically when the rotation angles increased. This demonstrates that the performance of the other approaches can be easily effected by local deformation of the face appearance.

We end this section by comparing the proposed face fitting and alignment method to a simple affine transform-based face alignment approach without pose correction. The results demonstrate that the face recognition performance increases significantly with the pose correction. As listed in Table 1, recognition with simple affine alignment only achieved 62.5% and 78.5% with $\pm15\,^\circ$ degree of face rotation, and it did not work at all when the face rotation angles were larger than $40\,^\circ$.

Table 1. Face recognition results on the FERET series *bb-bi* with PAN and simple affine (SA) face warping. Recognition with local appearance-based approach.

Probe set	bb	bc	bd	be	bf	bg	bh	bi
Rec. rate (PAN)	44.0%	81.5%	93.0%	97.0%	98.5%	91.5%	78.5%	52.5%
Rec. rate (SA)	0.0%	5.5%	26.0%	62.5%	78.5%	26.5%	4.0%	1.0%

6   Conclusions

A face registration approach based on AAM fitting was studied in this work. The pose of an input face was normalized with a piecewise affine warp resulting in a frontal view of the input face. To initialize the model more precisely and to increase robustness against cluttered backgrounds, an inner-face AAM was built and the fitting was done progressively. Face recognition was performed on the pose-normalized face images using the local appearance-based approach. We extracted the local features from each block on a pose normalized image using the DCT, and then concatenated the normalized local features for classification.

We conducted experiments on the FERET *b*-series to evaluate the proposed approach. With the AAM-based approach we achieved 97.8% recognition rate on the data set with $\pm15\,^\circ$ yaw pose-angle. The simple affine warp approach however, only recognized 70.5% of the testing images with this pose angle. When the rotation angle increased to $\pm60\,^\circ$, recognition based on simple affine warp could

hardly recognize any images in the probe set, while the AAM-based approach still achieved a recognition rate of 48.3%. The proposed pose normalization helped a lot to recognize faces with small angles of rotation as we observed in the experiments. We also compared the local appearance-based face recognition to other approaches such as the Eigenfaces and the EHMM. The local appearance-based approach significantly outperformed the other two approaches and it was more robust against the error introduced by AAM fitting and face warping.

Acknowledgments

References

1. The IMM face database, `http://www2.imm.dtu.dk/~aam/`
2. Beymer, D., Poggio, T.: Face recognition from one example view. In: ICCV, Boston, MA, pp. 500–507 (1995)
3. Blanz, V., Vetter, T.: Face recognition based on fitting a 3D morphable model. IEEE Trans. on PAMI 25(9), 1063–1074 (2003)
4. Bookstein, F.L.: Principal warps: thin-plate splines and the decomposition of deformations. IEEE Trans. on PAMI 11(6), 567–585 (1989)
5. Chai, X., Shan, S., Gao, W.: Pose normalization for robust face recognition based on statistical affine transformation. In: Chai, X., Shan, S., Gao, W. (eds.) ICICS-PCM 2003, Singapore, vol. 3, pp. 1413–1417 (2003)
6. Chang, K., Bowyer, K., Flynn, P.: Face recognition using 2D and 3D facial data. In: Proc. ACM Workshop on Multimodal User Authentication, pp. 25–32 (2003)
7. Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 484–498. Springer, Heidelberg (1998)
8. Cootes, T.F., Wheeler, G.V., Walker, K.N., Taylor, C.J.: View-based active appearance models. Image and Vision Computing 20(9-10), 657–664 (2002)
9. Edwards, G.J., Cootes, T.F., Taylor, C.J.: Face recognition using active appearance models. In: Burkhardt, H., Neumann, B. (eds.) ECCV 1998. LNCS, vol. 1407, pp. 581–595. Springer, Heidelberg (1998)
10. Ekenel, H.K., Stiefelhagen, R.: Local appearance based face recognition using discrete cosine transform. In: 13th European Signal Processing Conference (EUSIPCO 2005), Antalya, Turkey (2005)
11. Ekenel, H.K., Stiefelhagen, R.: Block selection in the local appearance-based face recognition scheme. In: CVPR Biometrics Workshop, New York (2006)
12. Ekenel, H.K., Stiefelhagen, R.: Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In: CVPR Biometrics Workshop (2006)
13. Goodall, C.: Procrustes methods in the statistical analysis of shape. Journal of the Royal Statistical Society, Series B 53(2), 285–339 (1991)
14. Gross, R., Matthews, I., Baker, S.: Generic vs. person specific active appearance models. Image and Vision Computing 23(11), 1080–1093 (2005)

15. Guillemaut, J., Kittler, J., Sadeghi, M.T., Christmas, W.J.: General pose face recognition using frontal face model. In: Martínez-Trinidad, J.F., Carrasco Ochoa, J.A., Kittler, J. (eds.) CIARP 2006. LNCS, vol. 4225, pp. 79–88. Springer, Heidelberg (2006)

16. Huang, G.B., Ramesh, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: a database for studying face recognition in unconstrained environments. University of Massachusetts, Amherst, Technical Report 07-49 (October 2007)

17. Kim, D., Kim, J., Cho, S., Jang, Y., Chung, S.-T., Kim, B.-G.: Progressive AAM based robust face alignment. In: Proc. of World Academy of Science, Engineering and Technology, vol. 21, pp. 488–492 (2007)

18. Matthews, I., Baker, S.: Active appearance models revisited. IJCV 60(2), 135–164 (2004)

19. Nefian, A.: A hidden Markov model-based approach for face detection and recognition, Ph.D. dissertation, Georgia Institute of Technology (1999)

20. Phillips, P.J., Wechsler, H., Huang, J., Rauss, P.: The FERET database and evaluation procedure for face recognition algorithms. Image and Vision Computing 16(5), 295–306 (1998)

21. Sanderson, C., Shan, T., Lovell, B.C.: Towards pose-invariant 2D face classification for surveillance. In: Zhou, S.K., Zhao, W., Tang, X., Gong, S. (eds.) AMFG 2007. LNCS, vol. 4778, pp. 276–289. Springer, Heidelberg (2007)

22. Shan, T., Lovell, B.C., Chen, S.: Face recognition robust to head pose from one sample image. In: 18th Int'l. Conf. on Pattern Recognition, vol. 1, pp. 515–518 (2006)

23. Sim, T., Baker, S., Bsat, M.: The CMU pose, illumination and expression (PIE) database. In: Proc. IEEE Int'l. Conf. on Automatic Face and Gesture Recognition (May 2002)

24. Stallkamp, J., Ekenel, H.K., Stiefelhagen, R.: Video-based face recognition on real-world data. In: ICCV 2007, Rio de Jenario, Brasil (October 2007)

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