Face verification using correlation filters

Article · January 2002

CITATIONS
110

READS
325

3 authors:

Marios Savvides
Carnegie Mellon University
223 PUBLICATIONS **3,481** CITATIONS
SEE PROFILE

B. V. K. Vijaya Kumar
Carnegie Mellon University
661 PUBLICATIONS **9,703** CITATIONS
SEE PROFILE

Pradeep K. Khosla
University of California, San Diego
480 PUBLICATIONS **15,788** CITATIONS
SEE PROFILE

Face Verification using Correlation Filters

Marios Savvides
Electrical and Computer Eng. Dept,
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.
msavvid@ri.cmu.edu

B.V.K. Vijaya Kumar
Electrical and Computer Eng. Dept,
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.
kumar@ece.cmu.edu

Pradeep Khosla
Electrical and Computer Eng. Dept,
Carnegie Mellon University
Pittsburgh, PA 15213, U.S.A.
pkk@ece.cmu.edu

Abstract

Face verification is an important tool for authentication of an individual and it can be of significant value in security and e-commerce applications. This paper deals with the application of correlation filters [1] for face verification. The performance of a specific type of correlation filter called the minimum average correlation energy (MACE) filter [2] is evaluated using a facial expression database collected at the Advanced Multimedia Processing Lab at Carnegie Mellon University (CMU). A comparison of verification performance between the correlation filter method and Individual Eigenface Subspace Method (IESM) is also presented. It is seen that these correlation filters offer significant potential for face verification.

Keywords

Correlation filters, minimum average correlation energy (MACE), face verification, eigenfaces, authentication.

INTRODUCTION

Correlation filters have been applied successfully to automatic target recognition (ATR) [3] problems. The most basic correlation filter is the matched spatial filter (MSF), whose impulse response (in 2-D, point spread function) is the flipped version of the reference image. While the MSF performs well at detecting a reference image corrupted by additive white noise, it performs poorly when the reference image appears with distortions (e.g., rotations, scale changes). Thus one MSF will be needed to detect each appearance of an object. Clearly this is computationally unattractive for practical pattern recognition. Hester and Casasent [4] addressed this challenge with the introduction of the synthetic discriminant function (SDF) filter. The SDF filter is a linear combination of MSFs where the combination weights are chosen so that the correlation outputs corresponding to the training images would yield pre-specified values at the origin. These pre-specified peak values are often referred to as *peak constraints*. The peak values corresponding to the authentics (also called the *true class*) are typically set to 1, and hence this SDF filter was known as the equal correlation peak (ECP) SDF filter. In principle, a single ECP SDF filter could replace many MSFs.

Object recognition is performed by cross-correlating an input image with a synthesized template or filter and processing the resulting correlation output. Figure 1 shows schematically how the cross-correlation is implemented efficiently using Fast Fourier Transforms (FFTs). The correlation output is searched for peaks, and the relative heights of these peaks are used to determine whether the object of interest is present or not. The locations of the peaks indicate the position of the objects.

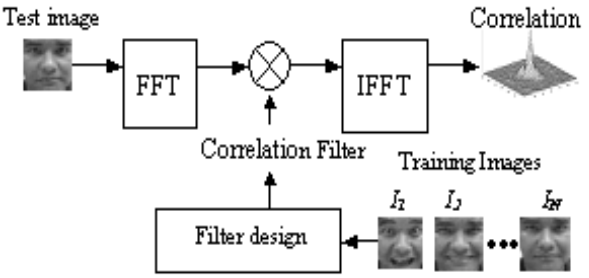

Figure 1. Block diagram showing the correlation process

Although the ECP SDF filter produces pre-specified correlation peak values, it also results in large sidelobes. Sometimes these sidelobes are larger than the pre-specified peak values leading to misclassifications. The reason for this is that ECP SDF design controls only the correlation values at the origin and nowhere else.

MINIMUM AVERAGE CORRELATION ENERGY (MACE) FILTER

In an effort to reduce the large sidelobes observed from the ECP SDF filter, Mahalanobis et al [2] developed the minimum average correlation energy (MACE) filter. MACE filter minimizes the average correlation energy (ACE) of the correlation outputs due to the training images while simultaneously satisfying the correlation peak constraints at the origin. The effect of minimizing the ACE is that the resulting correlation planes would yield values close to zero everywhere except at the location of a trained object, where it would produce a strong peak. Resulting MACE filter is as follows in a vector form:

$$\mathbf{h}_{\mathbf{MACE}} = \mathbf{D}^{-1}\,\mathrm{X}\,(\mathbf{X}^{+}\,\mathbf{D}^{-1}\,\mathrm{X})^{-1}\,\mathrm{c} \qquad (1)$$

Suppose that we have N training images from the true class with each image having d pixels in it. We perform 2-D FFTs on these images and convert the 2-D FFT arrays into 1-D column vectors by lexicographic ordering. These vectors are the column vectors of the $d \times N$ matrix $\mathbf{X}$ in Eq.

(1). Column vector **c** with N elements contains the pre-specified correlation peak values of the training images and the $d \times d$ diagonal matrix **D** contains along its diagonal the average power spectrum of the training images (i.e., average of the magnitude squares of the columns of **X**). Note that the synthesized $\mathbf{h}_{\mathbf{MACE}}$ is a column vector with d elements and the 2-D correlation filter is obtained by reordering the column vector back to a 2-D array. Throughout this paper we shall use the following notation: uppercase, bold letters indicate matrices and lowercase bold letters denote column vectors. The + symbol represents the complex conjugate transpose and * denotes the complex conjugate.

DATABASE USED FOR FACE VERIFICATION

The computer simulations described in this paper utilize a facial expression database collected at the Advanced Multimedia Processing (AMP) Lab at the Electrical and Computer Engineering Department of CMU [5]. The database consists of 13 subjects, whose facial images were captured with varying expressions. Each subject in the database has 75 images of varying facial expressions. The faces were captured in a video sequence where a face tracker [6] tracked the movement of the user's head and based upon an eye localization routine and extracted registered face images of size 64x64. Example images are shown in Fig. 2.

Figure 2. Sample images from the Advanced Multimedia Processing Lab's facial expression database.

FACE VERIFICATION USING MACE FILTERS

We have evaluated, using the above facial expression database, the performance of MACE filter for face verification. The computer simulation proceeded as follows. A single MACE filter was synthesized for each of the 13 persons using a variable number of training images from that person. In the test stage, for each filter, we performed cross correlations with all the face images from all the people (i.e., 13 x 75 = 975 images). For authentics, the correlation output should be sharply peaked and it should not exhibit such strong peaks for impostors. Peak to sidelobe ratio (PSR) defined below is used to measure the peak sharpness.

$$\mathrm{PSR} = \frac{peak - mean}{\sigma} \qquad (2)$$

Figure 3 illustrates how the PSR is estimated. First, the peak is located (shown as the bright pixel in the center of the figure). The mean and the standard deviation of the 20x20 sidelobe region (excluding a 5x5 central mask) centered at the peak are computed. The PSR is the ratio of (peak-mean) to standard deviation as shown in Eq. (2).

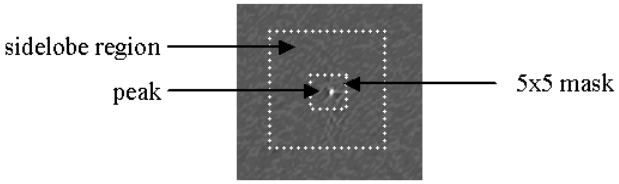

Figure 3. This figure shows how the peak to side lobe ratio (PSR) is estimated.

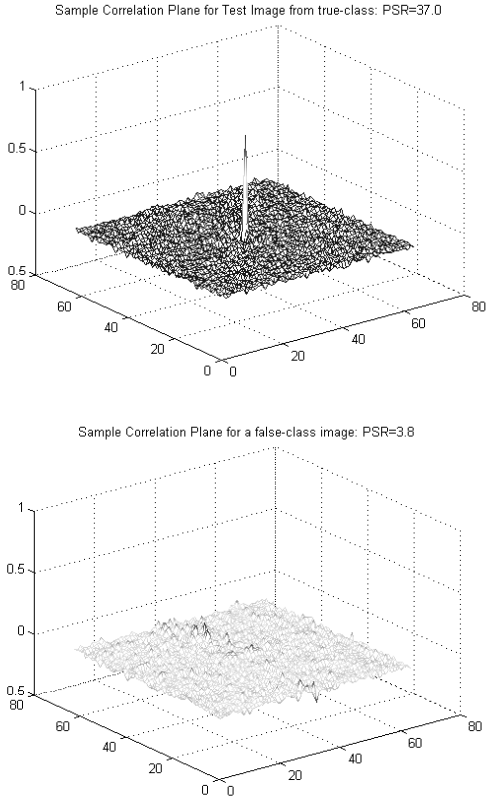

Figure 4. Correlation outputs when using a MACE filter designed for Person A. *(Top):* Input is a face image belonging to Person A. *(Bottom):* Input is a face image not belonging to Person A .

Figure 4 (top) shows a typical correlation output for an authentic face image. Note the sharp correlation peak resulting in a large PSR value of 37. The bottom correlation output in Fig. 4 shows a typical response to an impostor face image exhibiting low PSRs (<10).

We first used only 3 training images for the synthesis of each person's MACE filter. These three images were at a uniform interval in order to capture some of the expression

variations in the dataset (e.g., images # 1, 21 and 41). To evaluate the performance of each person's MACE filter, cross-correlations of all the images in the dataset were computed using a person's MACE filter resulting in 13*75=975 correlation outputs (corresponding to 75 true class images and the 900 false class images) and the corresponding PSRs were measured and recorded.

Figure 5 shows the best MACE filter PSR performance (upper plot, Person 1), and the worst PSR performance (lower plot, Person 2) as a function of image index. PSRs of authentics are shown in solid line and those of impostors using dotted lines.

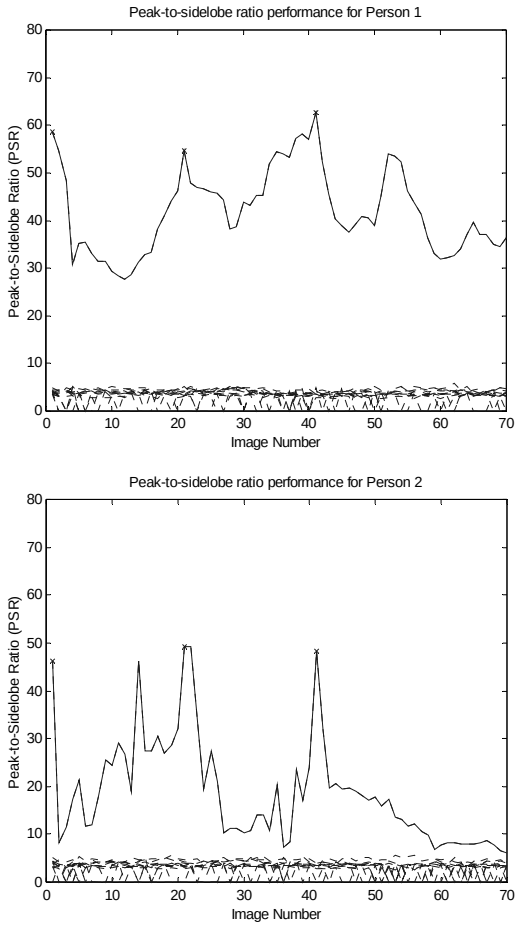

Figure 5. (*Top*): PSRs for Person 1, and (*Bottom*): PSRs for Person 2.

One very important observation from all 13 PSR plots (we showed only two PSR plots in Fig. 5) is that all the false class images (12*75=900) yielded PSR values consistently smaller than 10 (those are the dotted lines at the bottom of the plot) for all 13 subjects. The three 'x' symbols indicate the PSRs for the three training images used to synthesize the MACE filter for that person, and as expected they yield high PSR values. Person 2 whose filter yields the worst performance (exhibiting the smallest margin of separation between the true and false class PSR values), suggests that the expected distortions in the test set were not adequately captured by the training set, and indeed a close look at the dataset shows that person 2 exhibits significantly more variation in facial expressions than others. Thus more training images may be needed to improve authentication of face images belonging to person 2. Nevertheless even person 2's filter designed using only 3 training images performed reasonably well yielding a 99.1% verification performance.

Table 1 Error percentages for all 13 MACE filters synthesized using only 3 training images.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13
FAR, FRR=0	0	1.3	0	0	1	0	0	0	0	0	0	0	0
EER	0	0.9	0	0	1	0	0	0	0	0	0	0	0
FRR, FAR=0	0	0.2	0	0	2.6	0	0	0	0	0	0	0	0

Table 2 Error percentages for all 13 MACE filters synthesized using the first 5 training images.

Person	1	2	3	4	5	6	7	8	9	10	11	12	13
FAR, FRR=0	0	2.4	0	0	0	0	0	0	0	0	0	0	0
EER	0	1.3	0	0	0	0	0	0	0	0	0	0	0
FRR, FAR=0	0	2.6	0	0	0	0	0	0	0	0	0	0	0

Table 1 shows the error rates achieved using MACE filters designed from only 3 training images. FAR, FRR and EER refer to false acceptance rate, false rejection rate and equal error rate, respectively. EER is the case when FRR equals FAR. Table 1 shows that the overall EER (13 filters each tested on 975 images) is only 0.15% from MACE filters designed from only 3 training images per person.

We also performed a similar experiment using the first 5 training images from each person in the dataset to design that person's filter. These 5 images exhibit a different range of variability and have been placed there out of sequence. Table 2 summarizes the results of using 5 training images per person. There is some improvement in that Person 5 is now 100% correctly classified. However, class 2 gives 1.3% EER for an overall EER of 0.1%.

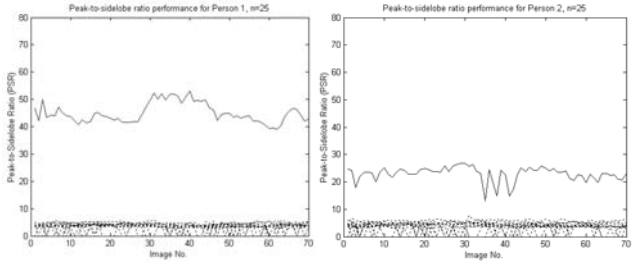

Figure 6. PSR plots for Person 1 *(left)*, Person 2 *(right)* for MACE filters designed from 25 training images.

Another simulation was performed where we increased the size of the training dataset size to 25 face images per

person sampled at regular intervals of the 75-image video sequence. Figure 6 shows the PSR plots for Person 1 (left) and Person 2 (right) for MACE filter synthesized from 25 training images. This figure shows a larger margin of separation than in the previous cases. In this case, the 13 MACE filters yielded 100% verification for all people.

Although we have shown that the verification accuracy of the MACE filters increases as more training images are used for filter synthesis, it is attractive that this method can work well with as few as 3 training images per class for this database.

INDIVIDUAL EIGENFACE SUBSPACE METHOD

Turk and Pentland introduced the eigenface method [7] for performing face recognition. This approach used the training faces from all the people to compute a *universal* eigenface subspace. While such a universal subspace is optimal for representing all the people's training faces in the minimum mean squared error sense, it may not adequately capture or describe the detailed information that discriminates one person's face from another. A better approach for face verification may be to use each person's training faces to build an individual eigenface subspace [8].

Let $\mathbf{f}_{i,n}$ denote the d-element column vector representing the n-th face image from the i-th person. Then the mean vector and the covariance matrix representing the i-th person are as follows.

$$\bar{\mathbf{f}}_i = \sum_{n=1}^{N} \mathbf{f}_{i,n} \tag{3}$$

$$\mathbf{C}_i = \sum_{n=1}^{N} (\mathbf{f}_{i,n} - \bar{\mathbf{f}}_i)(\mathbf{f}_{i,n} - \bar{\mathbf{f}}_i)^T \tag{4}$$

Then eigenvectors $\mathbf{e}_{i,j}$ and eigenvalues $\lambda_{i,j}$ are computed from the covariance matrix $\mathbf{C}_i$. More details of this method can be found in [9] and efficient ways to calculate these eigenvectors are described in [10]. Since each individual's orthonormal eigenface basis best spans that person's faces, we test input face images as follows. We first project the test image on to each person's individual eigenface subspace, then reconstruct the face $\mathbf{r}_{i,j}$ as shown in Eq. (5) using the projection coefficients in that subspace and finally measure the residual error between the original test input face image and the reconstructed face image.

$$\mathbf{r}_{i,j} = \bar{\mathbf{f}}_i + \sum_{n=1}^{M} \{(\mathbf{f}_{i,j} - \bar{\mathbf{f}}_i)^T \mathbf{e}_{i,n}\}\mathbf{e}_{i,n} \tag{5}$$

In Eq (5), M denotes the number of eigenvectors used to span the i-th class eigenface subspace. M is chosen to keep the average reconstruction squared error of the training images spanned by the M-eigenvectors basis.

$$\sum_{k=1}^{M} \frac{\lambda_{i,k}}{T_i} > R \tag{6}$$

Where $T_i = \sum_{k=1}^{N} \lambda_{i,k}$ and where $\lambda_{i,k}$ denotes the i-th class eigenvalues, arranged in decreasing magnitude. Typically we choose M so that the squared reconstruction error is 1%, i.e., $R = 0.99$. The reconstruction error is the squared norm between the test image $\mathbf{f}_{i,j}$ and its reconstruction $\mathbf{r}_{i,j}$.

$$\text{Err} = \left\| \mathbf{f}_{i,j} - \mathbf{r}_{i,j} \right\|^2 \tag{7}$$

An authentic face image and its expression variations are expected to be modeled well in its individual eigenface subspace; thus leading to small residual errors. Similarly, impostor face images are not well represented by some one else's individual eigenface subspace and should result in larger residual errors.

FACE VERIFICATION USING INDIVIDUAL EIGENFACE SUBSPACE METHOD (IESM)

To provide a benchmark of the MACE filter verification performance, we repeated the face verification experiment using the same training images as before, but this time using the individual eigenface subspace method. For each person's eigenface subspace, we projected all the face images from each person and reconstructed the face images to record the reconstruction error. Experiments were performed using 3, 5, and 25 training images per person.

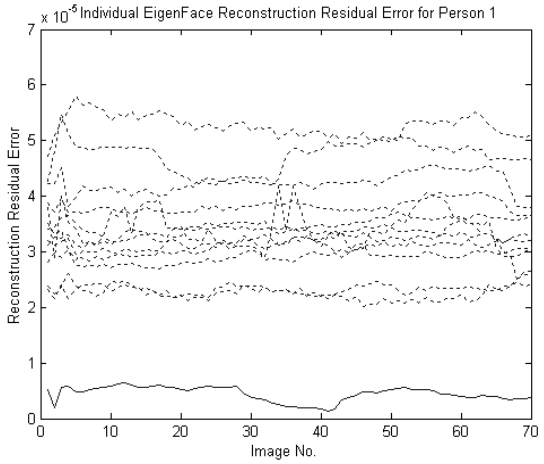

Figure 7. Reconstruction error plot for Person 1's individual eigenface subspace

Figures 7 and 8 show reconstruction errors for Person 1 and Person 2, respectively as a function of image index while using individual eigenspaces constructed using with n=3 training images per person. The solid line here represents the reconstruction error of faces belonging to the authentic person, while the dotted lines represent the recon-

struction errors from the impostors. Each dotted line represents a particular false class person.

Figure 8. Reconstruction errors for Person 2's individual eigenface subspace

Table 3 Error percentage for all 13 individual eigenface subspaces using 3 training images per person

Person	1	2	3	4	0	6	7	8	9	10	11	12	13
FAR, FRR=0	0	5.3	2.6	0	0	0	0	7.6	0	0	0	0	0
EER	0	3.5	2.1	0	0	0	0	5.4	0	0	0	0	0
FRR, FAR=0	0	8	10.6	0	0	0	0	14.7	0	0	0	0	0

Table 4 Error percentage for all 13 individual eigenface subspaces using 5 training images per person

Person	1	2	3	4	5	6	7	8	9	10	11	12	13
FAR, FRR=0	0	0	0	0	0	0	0	1.5	0	0	0	0	0
EER	0	0	0	0	0	0	0	1.3	0	0	0	0	0
FRR, FAR=0	0	0	0	0	0	0	0	9	0	0	0	0	0

Tables 3 and 4 show the error percentages for all 13 classes using the individual eigenface subspace method. In Table 3, the average EER is 0.85% which is higher than the 0.15% obtained from the MACE filters designed from only 3 training images per person. Table 4 shows that the IESM method achieves the same EER of 0.1% as the MACE filter synthesized using 5 training images per person. It appears that the MACE filter yields smaller FRR,FAR=0, FRR=0,FAR values. We also repeated the IESM simulations using 25 training images, and it was observed that 100% verification performance on this data set was also achieved using the IESM method. Figure 9 shows the reconstruction errors for IESM based on 25 training images. There is a larger margin of separation for Person 2 yielding 100% verification performance.

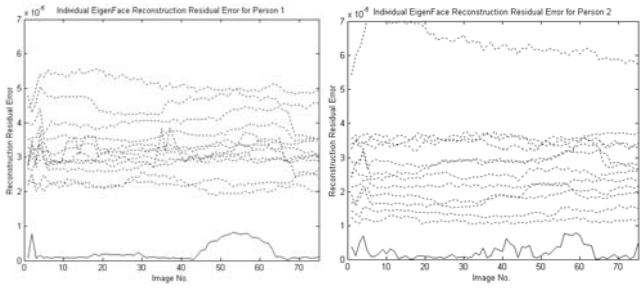

Figure 9. Reconstruction Errors for IESM using 25 training images of Person 1 (left) and Person 2 (right)

DISCUSSION

We have shown results using MACE correlation filter approach to performing face verification and compared it to the individual eigenface subspace method. MACE has shown to perform better with fewer images. However, error rates do not convey the full story and it is instructive to view the plots of the measured feature (e.g., the PSRs and the reconstruction residual errors) and examine the margins of separations. Figures 5 and 6 for the MACE filter show that the PSR values for the impostor faces seem bounded below a fixed threshold regardless of the person (i.e., all impostor PSRs are smaller than 10). In contrast, for the IESM method, the impostor reconstruction residue errors exhibit larger variations sometimes leading to smaller margins of separation between the true and false classes. The larger the margin of separation, the more confidence we have in the system to handle larger variations of the true class faces beyond that seen in the training set (e.g., possibly due to lighting variations which is a common hurdle in many face verification systems).

CONCLUSION

This paper has provided a short overview of the application of correlation filters to face verification, focusing mainly on how MACE correlation filters can be used. The simulation results shown are promising, demonstrating that correlation filters can be an alternative solution for performing face verification. Advantages of the correlation method include shift-invariance and ability to suppress impostor faces using a universal threshold. We are currently improving the filter design methods and testing the correlation filters on the much larger PIE [11] (pose, illumination and expression) database collected at CMU. Results of this investigation will be presented in the near future.

ACKNOWLEDGMENTS

We would like to thank Xiaoming Liu of AMP Lab at CMU for his help in obtaining the facial expression database used in our investigations.

REFERENCES

[1] B.V.K. Vijaya Kumar, “Tutorial survey of composite filter designs for optical correlators,” *Appl. Opt*. 31, pp. 4773-4801 (1992).

[2] A. Mahalanobis, B.V.K. Vijaya Kumar, and D. Casasent, “Minimum average correlation energy filters,” *Appl. Opt*. 26, pp. 3633-3630 (1987).

[3] A. Mahalanobis, D. W. Carlson, and B.V.K. Vijaya Kumar, “Evaluation of MACH and DCCF correlation filters for SAR ATR using MSTAR public database,” *Algorithms for Synthetic Aperture Radar Imagery V* (E. G. Zelnio, ed.), Vol. 3370, pp. 460-468, Photo-Opt. Instrum. Eng, (1998).

[4] C. F. Hester and D. Casasent, “Multivariant technique for multiclass pattern recognition,” *Appl. Opt*. 19, pp.1758-1761 (1980).

[5] <http://amp.ece.cmu.edu> - Advanced Multimedia processing Lab web page at Electrical and Computer Engineering Department at CMU.

[6] F. J. Huang and T. Chen, "Tracking of Multiple Faces for Human-Computer Interfaces and Virtual Environments", *IEEE Intl. Conf. on Multimedia and Expo.*, New York (2000).

[7] M. Turk, A. Pentland, “Eigenfaces for Recognition”, *Journal of Cognitive Neuroscience*, 3(1), pp.71-86, 1991.

[8] X. Liu, T. Chen, B.V.K. Vijaya Kumar, "Face Authentication for Multiple Subjects Using Eigenflow". Accepted for publication, *Pattern Recognition*, special issue on Biometrics (2001).

[9] R. Duda and P Hart, *Pattern classification and scene analysis*, John Wiley, New York (1973).

[10] H. Murakami, B.V.K. Vijaya Kumar, “Efficient calculation of primary images from a set of images,” *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 4 (5), pp. 511-545 (1982).

[11] T. Sim, S. Baker, and M. Bsat “The CMU Pose, Illumination, and Expression (PIE) Database of Human Faces,” tech. report CMU-RI-TR-01-02, Robotics Institute, Carnegie Mellon University, January (2001).