

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

**4,800**

Open access books available

**122,000**

International authors and editors

**135M**

Downloads

Our authors are among the

**154**

Countries delivered to

**TOP 1%**

most cited scientists

**12.2%**

Contributors from top 500 universities



**WEB OF SCIENCE™**

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.

For more information visit [www.intechopen.com](http://www.intechopen.com)



## Improving Face Recognition by Video Spatial Morphing

Armando Padilha, Jorge Silva, Raquel Sebastião  
*University of Porto, Faculty of Engineering  
Portugal*

### 1. Introduction

The focus of this chapter is in the problem of using technology to grant access to restricted areas by authorised persons, hereafter called 'clients', and to deny access to unauthorised or unidentified persons, the so called 'impostors'.

Conventional methods, such as magnetic or smart cards, user/password login and others, are being progressively recognised as insecure due to their many shortcomings, like the possibility of being lost, damaged or forged. Other methods, particularly those based on biometrics, are being increasingly used as they allow the verification of an individual's identity on the basis of precise and careful measures of biological and physiological characteristics, such as fingerprints, hand and palm print geometry, iris and retina patterns, voice and face recognition.

Automatic face recognition has very much progressed in the last few years, making its use practical in experimental or commercial systems. However, further research is still needed to make these systems more robust, reliable and less dependant on special constraints, particularly those imposed on the data acquisition process.

In order to be as flexible as possible, current face recognition systems must use a large database of facial views for each client, so that distinct poses and emotional states can be accommodated, as well as other short-term variations in appearance caused by cosmetics or beard size, and by the use of various accessories such as spectacles or earrings. These multiple views are intended to increase the individual's recognition rate for the capture of a single facial test image.

The large dimension of the faces database induces a number of problems, namely the requirement for more storage, the increased computing time for recognition and decision, and the need for more complex classifiers.

In an attempt to cope with the above problems we have devised an alternative approach, essentially consisting in keeping a much smaller facial image database, and in testing for valid matches a number of images extracted from a video fragment acquired during the person's path in direction to the protected entrance.

The main advantages to be expected from this approach can be summarised as: (a) the size of the reference face database is substantially reduced, as a single image or a small number of images for each individual are kept, (b) the clients are not subject to much discomfort when building the database, as a single neutral view (for each relevant appearance) is

required, (c) the training and updating of the database is performed much faster, as the total number of images is reduced, and (d) no special pose is normally required from the client, as the system relies on the fact that only one or a few valid matches of the images from the video sequence suffice for positive identification.

The overall system concept can be briefly described in the following terms:

- The reference face database is built by using a single image of each individual, in a frontal pose and with a neutral expression, or by adding a few other such images, one for each possible aspect (e.g., with or without glasses, beard, make-up);
- Each image from the video sequence (possibly under-sampled) is paired with its horizontal reflection, and a morphed image version is produced which emulates the frontal pose of the individual, using an underlying 3-D model;
- The matching between each morphed image from the sequence and every reference image in the database is performed by using an appropriate technique;
- When a pre-specified number of the images in an individual's sequence is matched to a certain reference image and no match is found to other reference images, the individual is assumed to be correctly identified;
- In an authentication context, when an impostor presents itself to the system claiming a false identity, the match rejection should occur for all morphed images of the sequence.

In our implementation both the database images and the video images were subject to some form of pre-processing, the morphing technique used is View Morphing (as proposed by Seitz, see Section 3 below) and the image matching process is based on a transformation that uses Independent Component Analysis.

The remainder of the chapter is organised as follows: (a) the next section details the image and video acquisition and pre-processing, (b) then the View Morphing methodology is explained in a separate section, (c) which is followed by another one detailing the application of the Independent Component Analysis methodology, (d) and by an implementation and results section, divided in two sub-sections dealing separately with the authentication of clients and the rejection of impostors, (e) which precedes the final section devoted to the conclusions and suggestions for further research.

## 2. Image and Video Acquisition and Processing

We constructed a database with a single view for each individual. The face images were captured in a frontal pose to the camera, with a neutral expression and with a background as homogenous as possible, for a total of 22 subjects. The database still images were taken with an off-the-shelf digital camera.

In order to get the test images, a short video sequence was captured by the same camera, in a movie mode, during the individual's approach path to the camera, and then 3 images were selected from each video. This small number of images was chosen for practical reasons.

The test image acquisition process ensures that virtually no restrictions are imposed on the subjects, and enables the use of more than one camera in practical situations.

One negative issue that must be stressed is that the images extracted from the video are often of poor quality as fixed focus and zoom factor were set in the camera, which means that these images may display a very small facial area and that they may also be slightly blurred. The original database image and one of the extracted images from the video are shown in Figure 1.



Figure 1. Original database image ( $2288 \times 1712$  resolution) and one image extracted from the video ( $640 \times 480$  pixels)

An area of interest, containing essentially the facial visual information, is extracted from each of these images and converted to a standard size. The video capture process is shown in Figure 2.

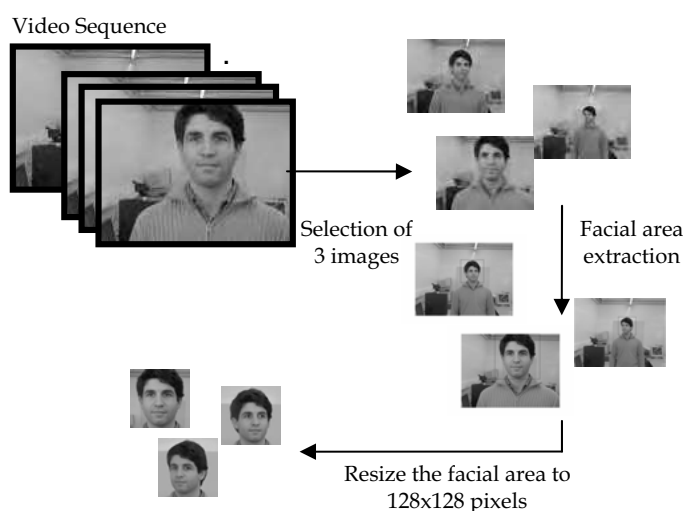


Figure 2. Image extraction from the video sequence and area of interest resizing

Having in mind the need for fast processing in an online system, on one hand, and the benefits of improving image contrast and normalisation, on the other, image pre-processing was performed over all captured images, using a combination of a photometric transformation to reduce the effects of brightness variation among different images, with a geometric normalisation to help the comparability of images by allowing some degree of distortion.

All images, both from the database and from the video fragment, are subject to a manual extraction of the area of interest and then converted to a standard resolution of  $128 \times 128$  pixels, using bi-cubic interpolation. The result of this process is shown in Figure 3. Moreover, the images are subject to histogram equalisation, so that comparable contrast is achieved throughout the whole image dataset. Pre- and post-equalised images are shown in Figure 4.



Figure 3. Areas of interest of database (left) and video (right) images, resized to  $128 \times 128$

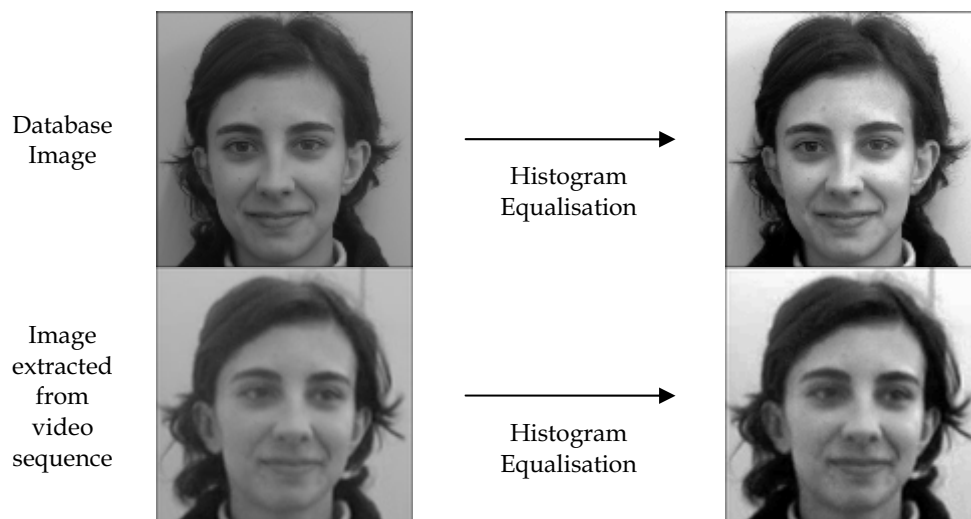


Figure 4. Effects of histogram equalisation

Then, a geometric correction is performed on all images, such that a normalised rendering is achieved. Basically, a planar image transformation is performed such that the eyes' centres and the midpoint of the horizontal line segment joining the corners of the lips are moved to fixed co-ordinates in the  $128 \times 128$  grid. This transformation uses bilinear interpolation. Then, the images are horizontally cropped from both sides, reducing the image resolution to  $128 \times 81$  pixels. Note that this geometric transformation, when applied to the video test images, is in fact performed after the view-morphing step described in the next section. Figure 5 shows the results of the geometric normalisation.



Figure 5. Results after geometric normalisation (database: left; view-morphed video: right)

In summary, the image processing steps are applied to all images in order to ease the authentication of clients and the rejection of impostors, both in terms of the radiometric and geometric properties of the images and in terms of the computation time required for online processing. These steps are summarised in Figure 6 for a database image; for a video image an intermediate view-morphing stage is to be included.

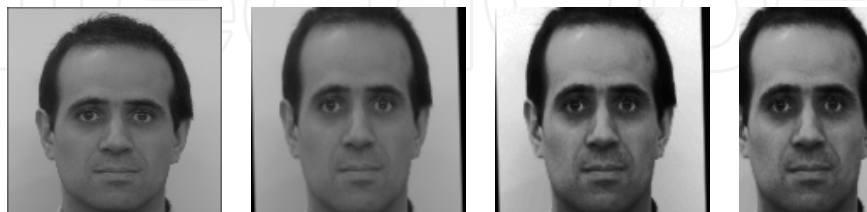


Figure 6. Image processing steps. From left to right:  $128 \times 128$  area of interest; geometric normalisation; histogram equalisation; cropping to  $128 \times 81$

### 3. View Morphing

In image acquisition the deviations from a frontal pose are often large, particularly for the video test images. One way of addressing this problem is to consider image synthesis as a preliminary step in order to simulate a facial image capture from a frontal viewpoint.

Ullman and Basri (Ullman & Basri, 1991) show that new views of an object can be expressed as a linear combination of other views of the same scene. However, their technique requires full correspondence between the original images and this is, quite often, very hard to achieve.

Chen and Williams (Chen & Williams, 1991) have proposed image synthesis based on linear interpolation of corresponding points between the original images. As this method causes a geometric bending effect, the interpolated frontal image can not be considered to represent a new view of the same object.

Lam and Yan (Lam & Yan, 1998) have used a snake model to extract the facial contours from the original images, and subsequently they detect 15 face feature points (such as the lips and eyes corners) and compute a 3D model based on these. The need to detect accurately the 15 feature points and the fact that the technique can only be applied to quasi-frontal images are two important limitations.

Beymer et al. (Beymer et al., 1993) have suggested the construction of a virtual view based on the theory that any 2D view can be expressed as a linear combination of other views. Their procedure, besides the need for more than 3 views to construct the new image, is also hampered by the requirement of a large number of image correspondences.

To avoid the difficulty of establishing a great number of corresponding points, Feng and Yuen (Feng & Yuen, 2000) presented an algorithm to detect facial landmarks, using these ones to estimate the face orientation in an image. After this estimation the authors propose the construction of a 3D model to transform an initial image into a frontal pose. Their technique only needs one image to construct a frontal view.

In spite of creating compelling 2D transitions between images, image morphing techniques often cause unnatural distortions. Seitz and Dyer (Seitz & Dyer, 1995; Seitz & Dyer, 1996; Seitz, 1997) propose a View-Morphing method to avoid that problem. From two images of the same object, in different poses, and with pixel correspondences in both images it is possible to compute any in-between view. The authors claim that view-morphing, using

basic principles of projective geometry and image geometric interpolation yields a more realistic transition that preserves the underlying 3D geometry. This technique works by pre-warping two images prior to computing a morph and then by post-warping the interpolated images. Because no knowledge of 3D shape is required the technique may be applied to photographs and drawings, as well as rendered scenes.

Xiao and Shah (Xiao & Shah, 2004) present an effective image-based approach without the explicit use of a 3D model. Based on the view-morphing methods, they propose a novel technique to synthesize a virtual view in a 2D space. Starting by establishing and refining corresponding feature points between each pair of images they determine the epipolar geometry for each pair and extract the trifocal plane by trifocal tensor computation. After marking a small number of feature lines, the correct dense disparity maps are obtained by using a trifocal-stereo algorithm developed by them. Finally, after self-calibration of the three cameras, they can generate an arbitrary novel view, synthesized by the tri-view morphing algorithm.

For our work we have developed a view-morphing algorithm closely following the Seitz and Dyer method. The view-morphing technique is shape-preserving, thus creating intermediate views that resemble the effect of rigid motion transformations in 3D. The intermediate images can be closer to any one of the two extreme images, this being controlled by a single transformation parameter.

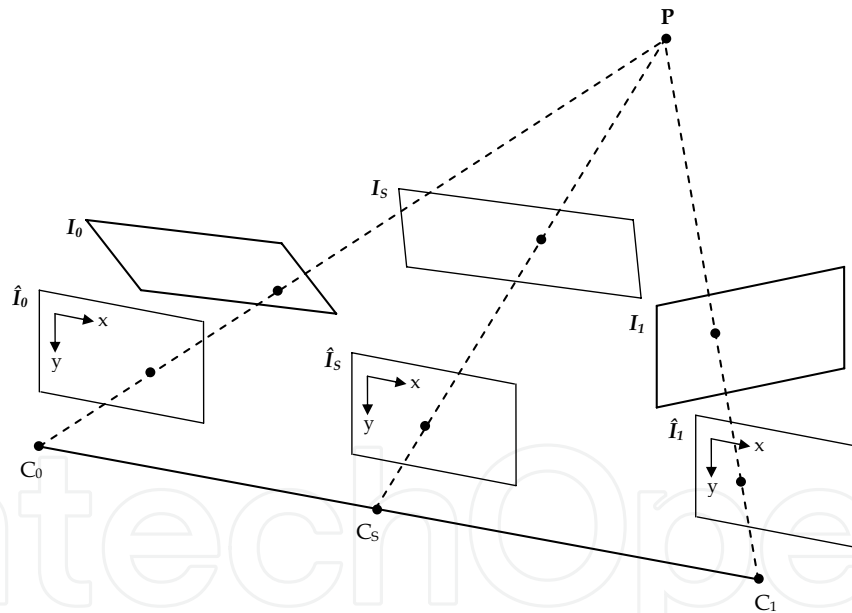


Figure 7. The pre-warping, morphing, and post-warping stages of the view-morphing technique (adapted from Seitz, 1997)

For non-parallel original views, i.e. images obtained by cameras with non-parallel optical axes, the first step is to backproject those views into a single plane where they become parallel. This is done by using the adequate projection matrices, whose construction details are beyond the scope of this document. Considering Figure 7, the original images  $I_0$  and  $I_1$  are backprojected



to  $\hat{I}_0$  and  $\hat{I}_1$ ,  $C_0$  and  $C_1$  representing the optical centres of the two cameras and  $P$  being a generic point in space (this is the pre-warping stage). New perspective views for virtual optical centres located along the segment joining  $C_0$  and  $C_1$  can be synthesised through linear positional and intensity interpolation, using a parameter  $s$  that may vary from 0 to 1; for a particular value of  $s$  the image  $\hat{I}_s$  is created, which has a virtual camera optical centre located at  $C_s$  (this is the morphing stage). Finally, a forward projection is performed, resulting in the construction of the desired virtual view  $I_s$  (this is the post-warping stage).

The extreme images for frontal pose estimation are an original image and its reflection on a vertical mirror, as shown in Figure 8.



Figure 8. Original image and its horizontal reflection

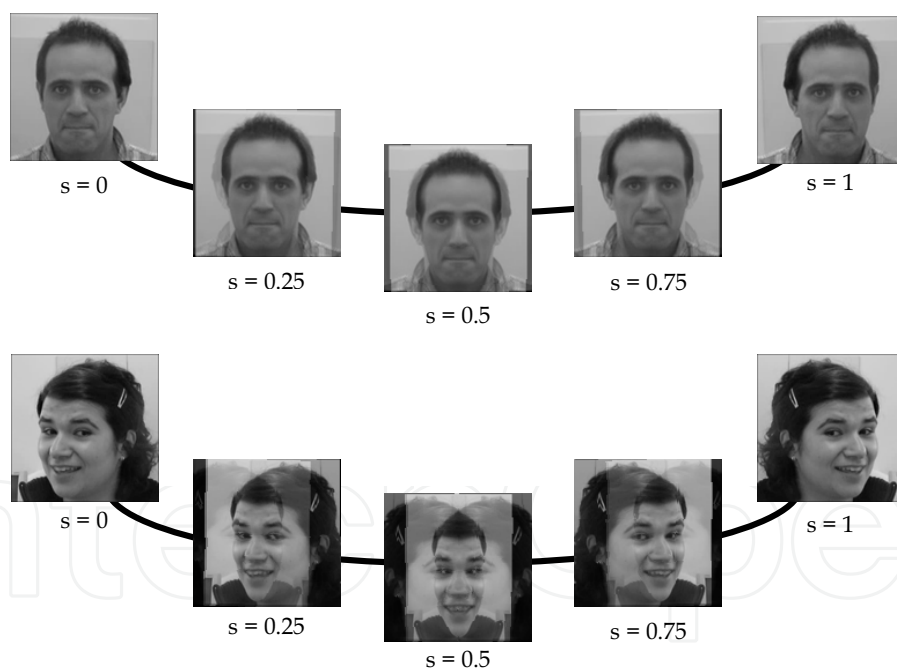


Figure 9. Examples of application of the view-morphing technique

Figure 9 shows the results of applying view-morphing to two pairs of extreme images. The intermediate images represent three different degrees of 3D rotation, obtained for values of  $s$  of 0.25, 0.5 and 0.75. The case  $s = 0.5$  represents the frontal view pose.



Observe in Figure 9 that while the original male image is not far from a frontal view, the original female image is quite far from a frontal pose. As a consequence, a significant part of the synthesized views of the male individual is subjectively correct, whereas for the female individual only a small vertical strip of the image is satisfactory.

The aforementioned problem is due to the practical implementation of the method, where only a few corresponding points in the original and reflected image were established, namely located on the eyes, nose and mouth, so that a very inaccurate extrapolation results in the outer vertical regions of the images.

#### 4. Independent Component Analysis

Image recognition has been the objective of numerous studies and investigations in several scientific disciplines, giving rise to many different approaches of which we will only briefly review a few.

Given the large data volumes present in image analysis, it is necessary to simplify the data by reducing its dimensionality. This goal can be reached through varied robust and efficient techniques, carrying out linear transformations that transform the data to a new co-ordinate system. The following representation, a linear statistical model, describes the observed data ( $\mathbf{x}$ ) through a mixing process (represented by matrix  $\mathbf{A}$ ) that depends on the latent variables or sources  $\mathbf{s}$ :

$$\mathbf{x} = \mathbf{A}\mathbf{s} \quad (1)$$

However, in most problems the mixing matrix  $\mathbf{A}$  is unknown. So, the independent variables must be recovered by a separation process ( $\mathbf{W}$  is the separation matrix):

$$\mathbf{s} = \mathbf{W}\mathbf{x} \quad (2)$$

There are a lot of available methods to use for this purpose, namely second and higher order methods. Principal Component Analysis (PCA) and Common Factor Analysis (CFA) are methods that approximate the intrinsic structure of image data up to its second order statistics, assuming data gaussianity, and thus find a data representation using only the information in the covariance matrix. These methods are easier to implement than the higher order techniques (Hyvärinen, 1999).

Nevertheless, the goal of many signal processing problems (such as speech enhancement, telecommunications and medical signal processing) is to reduce the information redundancy by exploiting the statistical structure of the data that is beyond second order. Independent Component Analysis (ICA) and Projection Pursuit (PP) are techniques that examine thoroughly the higher-order statistical structure in the data. Addressing higher-order statistical dependencies, these methods allow the separation of multivariate signals into additive sub-components, assuming the mutual statistical independence of the non-gaussian source signals (blind source separation is a special case of this), which is a much stronger condition than uncorrelatedness.

Many authors have been using PCA for face recognition (Bartlett et al., 1998; Bartlett et al., 2002; Draper et al., 2003; Fortuna et al., 2002; Torres et al., 2000; Turk & Pentland, 1991). PCA consists in finding the principal components of the distribution of faces. These components are the eigenvectors of the covariance matrix of the database of facial images, each one accounting for a different amount of the variation among the images, such that the greatest

variance of the data lies on the first co-ordinate (first principal component), the second greatest variance on the second co-ordinate, and so on. Each face can then be represented as a linear combination of the eigenfaces (eigenvectors) or approximated using only the “largest” eigenfaces, ordered according to the associated eigenvalues. (PCA is useful as a pre-processing step for other methods, as it reduces data dimensionality by keeping higher-order principal components and ignoring lower-order ones).

Common Factor Analysis is used to explain variability among observed random variables, modelling those as linear combinations of a smaller number of unobserved random variables called factors. This method allows the identification of the relationships between data items and shows the similarities and differences between them.

Simply stated, Projection Pursuit is a kind of statistical technique which involves finding the most “interesting” possible projections in multidimensional data.

In the context of face recognition, ICA and NMF (Non-Negative Matrix Factorization) were analysed in (Rajapakse & Wyse, 2003). Both approaches yield a sparse representation of localized features (from human faces) and the authors discuss the strengths and weaknesses of each one.

For use in face recognition, J. Yang et al. (Yang et al., 2004) developed a new Kernel Fisher Discriminant analysis (KFD) algorithm, called Complete KFD (CKFD), which is claimed to be simpler and more powerful than the existing KFD algorithms.

There are also studies comparing the use of Spatial-Temporal Networks (STN) and Conditional Probability Networks (CPN) in human face recognition (A. Fernández-Caballero et al., 2001). These authors evaluate both techniques by testing, without any kind of pre-processing, 16 image faces with 6 different poses and expressions for each one. The results obtained with CPN slightly outperform those with STN.

A simple application of Independent Component Analysis (ICA) is the “cocktail party problem”, where the underlying speech signals (sources) are separated from the mixed data consisting of people talking simultaneously in a room (a blind source separation problem).

Many authors (Bartlett et al., 1998; Bartlett et al., 2002; Draper et al., 2003; Fortuna et al., 2002) have compared face recognition results under PCA and ICA, generally claiming that ICA provides better performance.

Te-Won Lee and colleagues (Lee et al., 2000; Lee & Lewicki, 2002) apply ICA to find statistically significant structures in images construed by classes of image types, such as text overlapping with natural scenes, or the natural scene itself composed by diverse structures or textures. Developing what they called the ICA mixture model, the authors mould up the underlying image with a mixture model that can capture the different types of image textures in classes, categorizing the data into several mutually exclusive classes and assuming that the data in each class are generated by a linear combination of independent, non-Gaussian sources, as is the case with ICA.

Jung et al. (Jung et al., 2001), in the context of basic brain research and medical diagnosis and treatment, apply ICA to analyse electroencephalographic (EEG), magnetoencephalographic (MEG) and functional magnetic resonance imaging (fMRI) recordings. Removing artefacts and separating/recovering sources of the brain signals from these recordings, ICA allows the identification of different types of generators of the EEG and its magnetic counterpart (the MEG) and can be used to analyse hemodynamic signals from the brain recorded by the fMRI.

In general terms, ICA is a statistical method that finds the independent components (or sources) by maximizing the statistical independence of the estimated components, according

to equation (1). Mutual Information and non-gaussianity (measured by kurtosis or by approximations to negentropy) are popular criteria for measuring statistical independence of signals. Typical algorithms for ICA use centring and whitening as pre-processing steps in order to reduce the complexity of the problem for the actual iterative algorithm. The Newton method, the gradient ascent method, Infomax and FastICA are possible algorithms for ICA (Hyvärinen, 1999; Hyvärinen & Oja, 2000; Hyvärinen et al., 2001).

The ICA representation can be constructed under architectures I and II. The former considers images as random variables and pixels as observations, the second treats pixels as random variables and images as observations (Bartlett et al., 1998; Bartlett et al., 2001). Architecture II produces more global features while architecture I produces spatially localized features that are mostly influenced by small parts of the image, leading to better object recognition results.

Based on reasons given in (Sebastião, 2006b) and detailed in (Sebastião, 2006a), we constructed the ICA representation considering architecture I and implemented it by the FastICA algorithm, using  $g(y) = \tanh(y)$  as the objective function. The data,  $\mathbf{X}$ , is first centred, i.e. made zero-mean, according to the following equation:

$$\mathbf{X}_c = \mathbf{X} - \boldsymbol{\mu} \quad (3)$$

where  $\boldsymbol{\mu}$  represents the data mean.

Then the data is whitened (or sphered), meaning it is normalised with respect to the variance. The sphering matrix,  $\mathbf{V}$ , is defined by the eigenvalues and eigenvectors matrices of the covariance matrix:

$$\mathbf{V} = \mathbf{E} \mathbf{D}^{-\frac{1}{2}} \mathbf{E}^T \quad (4)$$

where  $\mathbf{E}$  is the eigenvectors and  $\mathbf{D}$  is the eigenvalues matrix of the covariance matrix  $\boldsymbol{\Sigma}$ . The data matrix  $\tilde{\mathbf{X}}$  with zero-mean and covariance matrix equal to the identity matrix is obtained by the following transformation:

$$\tilde{\mathbf{X}} = \mathbf{V} \mathbf{X}_c = \mathbf{V} \mathbf{A} \mathbf{s} = \tilde{\mathbf{A}} \mathbf{s} \quad (5)$$

where  $\tilde{\mathbf{A}}$  is an orthogonal matrix. Thus, the problem of finding an arbitrary matrix  $\mathbf{A}$  in the model given by equation (1) is reduced to the simpler problem of finding an orthogonal matrix  $\tilde{\mathbf{A}}$ . For convenience the data matrix  $\tilde{\mathbf{X}}$  will be renamed to  $\mathbf{X}$ , in the sequel.

The evaluation/comparison of the two image representations, one pertaining to the training set (the still database images) and the other to the test set (the view-morphed images extracted from the video sequence), was measured by the cosine distance, as suggested by Bartlett et al. (Bartlett et al., 1998):

$$d(i, j) = 1 - \frac{\mathbf{X}_{\text{test}}(j) \cdot \mathbf{X}_{\text{train}}^T(i)}{\|\mathbf{X}_{\text{test}}(j)\| \|\mathbf{X}_{\text{train}}^T(i)\|} \quad (6)$$

where  $\mathbf{X}_{\text{test}}(j)$  and  $\mathbf{X}_{\text{train}}^T(i)$  represent the row vectors  $j$  and  $i$  of matrices  $\mathbf{X}_{\text{test}}$  and  $\mathbf{X}_{\text{train}}^T$ , respectively.

## 5. Implementation and Results

As previously stated, we constructed a database containing the still images of 22 subjects, who volunteered to participate in this work. In Figure 10 some of the still images are shown, while in Figure 11 the same images are already pre-processed.

Figure 13 shows the frontal view synthesis of the test images (extracted from video) of the same subjects represented in Figure 10, while Figure 12 shows these images prior to the view-morphing transformation.

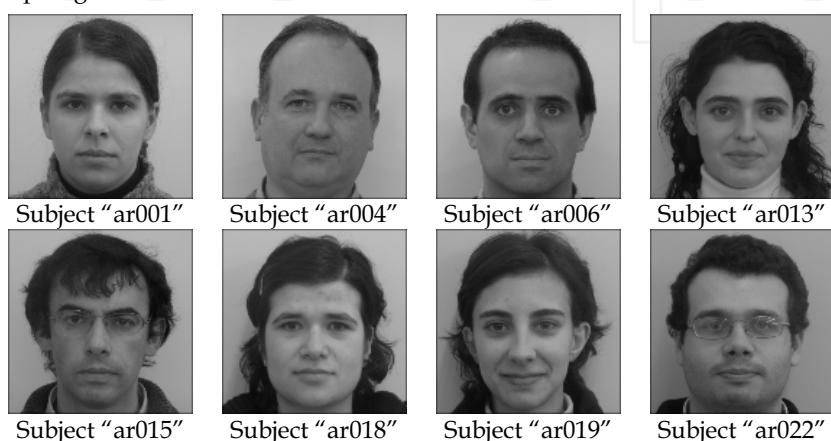


Figure 10. Some database images, stored as the areas of interest of the original still images

It is clear, particularly from the analysis of Figure 13, that the view-morphing only works well in the central vertical area of the face, as previously observed. This is due to the fact that the morphing is based on the actual image and its horizontal reflection, using only a few corresponding points, manually selected, located on the eyes, nose and mouth, thus forcing an inaccurate extrapolation in the outer vertical regions of the faces. This effect is much more noticeable when the deviation angle from frontal pose is larger.

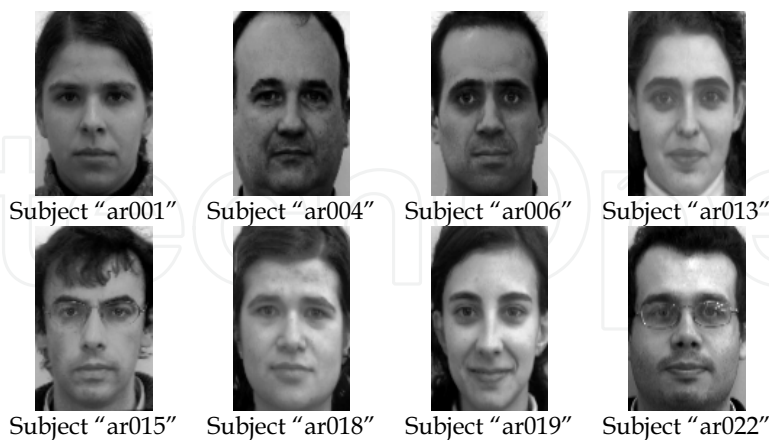


Figure 11. Same database images as in Figure 10, but after pre-processing

To evaluate the advantage of using view-morphing synthesis we compare the identification results obtained with the original images selected from the video sequence and those obtained with the images synthesized from the former ones.

The identification results were evaluated using the following rate:

$$R_{identification} = \frac{N_{correct}}{N_{total}} \times 100\% \quad (7)$$

This rate is the percent quotient between the number of images correctly identified and the total number of test images. The image face  $j$  from the test set is correctly identified as the image face  $i$  from the training set if the distance given by (6) is the minimum for all  $i$ 's and sufficiently close to zero.

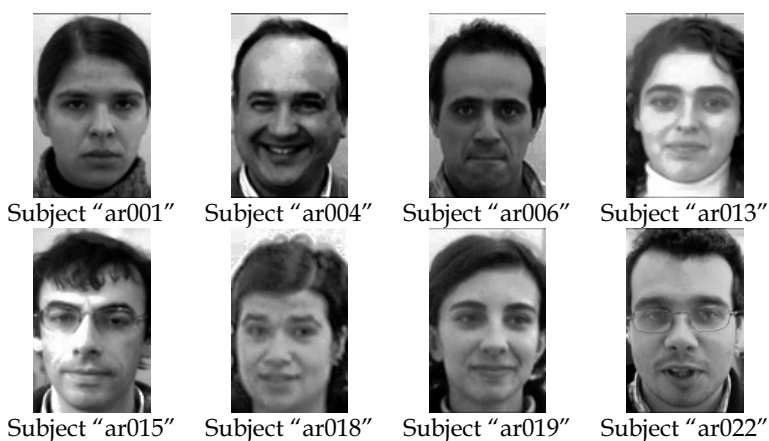


Figure 12. Images extracted from video (same subjects as in Figure 10)



Figure 13. Frontal view synthesis of the test images in Figure 12

To authenticate subjects we define a threshold for each individual using the distances of each face of subject  $i$ , selected from the video sequence, to all the representations in the training set. This threshold, for each individual, is given by:

$$T_i = \mu_i - \sigma_i \quad (8)$$

where

$$\mu_i = \frac{\sum_{k=1}^{N_{\text{images}}} \sum_{j=1}^{N_{\text{ind}}} d_k(i, j)}{N_{\text{images}} \cdot N_{\text{ind}}} \quad \text{and} \quad \sigma_i = \sqrt{\frac{\sum_{k=1}^{N_{\text{images}}} \sum_{j=1}^{N_{\text{ind}}} [d_k(i, j) - \mu_i]^2}{(N_{\text{images}} \cdot N_{\text{ind}} - 1)}} \quad (9)$$

with  $N_{\text{images}} = 3$  (3 frames from a video fragment) and  $N_{\text{ind}} = 22$  (total number of images in the database).

Moreover, a subject is authenticated if the distances between the representation of at least two of the three selected test images and the subject's training set representation are lower than the threshold value.

Comparing the identification results obtained with the view-morphed images and those obtained with the original images from the video sequence, we get the following rates:

	Video Images	View-morphed Images
Identification rate - $R_{\text{identification}}$	$\frac{32}{66} \times 100\%$	$\frac{35}{66} \times 100\%$

Table 1. Identification rates obtained for the video images and the view-morphed images

These results support the advantage of synthesizing frontal images with the view-morphing method, to achieve individual identification, even in the experimental case where the view-morphing synthesis is hampered by the uneven location of the corresponding points. The authentication of a subject takes approximately 8 seconds with code far from optimised, including full training of the database and test (using a 2.8 GHz Pentium 4 processor with 1.00 GB of RAM).

With the aim of improving image quality as well as the recognition rate, image pre-processing was applied to all images, using a photometric transformation and a geometric normalisation, as explained in section 2. Some of the pre-processed view-morphed images are shown below, in Figure 14.

Table 2 compares the identification results obtained with the view-morphed images without and with pre-processing.

	Without pre-processing	With pre-processing
Identification rate - $R_{\text{identification}}$	$\frac{35}{66} \times 100\%$	$\frac{40}{66} \times 100\%$

Table 2. Identification rates obtained for the view-morphed images, with and without pre-processing





Figure 14. Sample pre-processed and view-morphed images (video frames)

The above results suggest that one should use the pre-processed view-morphed images to get the best identification results. This way, the database is formed by the pre-processed images in frontal and neutral pose, and every image extracted from the video fragments is also subject to the full pre-processing stage.

### 5.1 Authentication of Clients

To authenticate clients we use a simple criterion, namely: at least two of the three test images from a video fragment must be correctly identified (meaning that the identity announced by the client gives the correct match in the database – a distance smaller than the respective threshold). The thresholds that allow the decision are associated to each individual, as previously defined, so that if we want to add a new subject to the database or to remove another one it is necessary to perform a full training test in order to establish the new threshold value for each subject.

Considering the above criterion, a full authentication test was conducted and an authentication rate (defined similarly to the identification rate) was computed in two situations: (a) test images without the geometric transformation step, and (b) test images with the geometric transformation step. Table 3 presents the results achieved.

	Without geometric transformation	With geometric transformation
Authentication rate - $R_{authentication}$	$\frac{15}{22} \times 100\%$	$\frac{22}{22} \times 100\%$

Table 3. Authentication rates obtained for the full set of view-morphed images, with and without geometric transformation

The results in Table 3 stress the importance of the geometric transformation.

On the other hand, the fact that an entirely correct authentication rate could be achieved should not be overstated. In fact, the decision thresholds definition was designed for the best performance with the existing test images ( $22 \times 3 = 66$  images), but it is not guaranteed that similar results would be achieved for other test images.









		Training Set																	
																			
		ar001			ar004			ar015			ar018			ar019			ar022		
		I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3
Test Set	ar001	0,26	0,14	0,18	0,85	0,75	0,69	0,37	0,30	0,26	0,43	0,36	0,41	0,47	0,39	0,44	0,52	0,57	0,47
	ar002	0,55	0,56	0,61	0,71	0,66	0,67	0,64	0,64	0,61	0,83	0,82	0,62	0,92	0,80	0,85	0,97	0,99	0,64
	ar003	0,49	0,49	0,52	0,84	0,79	0,94	0,65	0,31	0,30	0,18	0,23	0,48	0,24	0,20	0,42	0,42	0,41	0,55
	ar004	0,55	0,60	0,58	0,31	0,32	0,19	0,81	1,00	0,95	0,84	0,70	0,61	0,59	0,73	0,45	0,53	0,45	0,46
	ar005	0,27	0,29	0,32	0,61	0,59	0,47	0,38	0,54	0,51	0,48	0,44	0,32	0,49	0,42	0,27	0,32	0,45	0,19
	ar006	0,33	0,33	0,43	0,54	0,47	0,37	0,61	0,62	0,60	0,71	0,67	0,57	0,55	0,54	0,31	0,49	0,49	0,32
	ar007	0,27	0,18	0,25	0,68	0,59	0,54	0,45	0,41	0,36	0,36	0,34	0,39	0,37	0,25	0,25	0,32	0,37	0,33
	ar008	0,60	0,60	0,63	0,31	0,33	0,37	0,71	0,71	0,63	0,54	0,46	0,64	0,54	0,42	0,30	0,42	0,48	0,32
	ar009	0,30	0,32	0,38	1,01	0,90	0,83	0,26	0,22	0,27	0,48	0,54	0,50	0,65	0,46	0,58	0,66	0,74	0,57
	ar010	0,35	0,26	0,35	0,88	0,76	0,69	0,49	0,32	0,28	0,49	0,42	0,49	0,52	0,41	0,44	0,59	0,65	0,49
	ar011	0,37	0,42	0,40	0,66	0,69	0,66	0,31	0,42	0,40	0,37	0,42	0,41	0,46	0,31	0,31	0,42	0,54	0,24
	ar012	0,32	0,38	0,29	0,77	0,67	0,74	0,54	0,56	0,51	0,48	0,48	0,33	0,42	0,44	0,47	0,46	0,48	0,38
	ar013	0,57	0,60	0,63	0,72	0,72	0,85	0,66	0,49	0,45	0,19	0,22	0,59	0,19	0,16	0,34	0,35	0,31	0,49
	ar014	0,42	0,45	0,46	0,96	0,85	0,99	0,47	0,32	0,33	0,27	0,34	0,43	0,42	0,32	0,58	0,53	0,57	0,65
	ar015	0,47	0,43	0,52	1,16	1,08	1,03	0,28	0,08	0,15	0,39	0,52	0,46	0,65	0,49	0,62	0,61	0,76	0,67
	ar016	0,37	0,30	0,33	0,98	0,96	0,84	0,20	0,22	0,18	0,40	0,41	0,46	0,54	0,38	0,48	0,60	0,68	0,50
	ar017	0,26	0,23	0,22	0,81	0,65	0,67	0,45	0,41	0,39	0,36	0,32	0,32	0,41	0,35	0,42	0,37	0,46	0,44
	ar018	0,53	0,51	0,52	0,89	0,86	0,98	0,50	0,29	0,24	0,09	0,11	0,45	0,23	0,16	0,40	0,33	0,37	0,52
	ar019	0,54	0,50	0,53	0,66	0,64	0,82	0,60	0,42	0,39	0,18	0,23	0,51	0,24	0,13	0,36	0,32	0,35	0,49
	ar020	0,45	0,45	0,43	0,86	0,80	0,91	0,57	0,36	0,33	0,29	0,29	0,51	0,31	0,27	0,47	0,46	0,44	0,55
	ar021	0,32	0,27	0,30	0,52	0,44	0,41	0,42	0,65	0,62	0,53	0,57	0,29	0,57	0,47	0,30	0,30	0,42	0,22
	ar022	0,45	0,42	0,44	0,63	0,62	0,67	0,53	0,43	0,44	0,31	0,31	0,44	0,37	0,25	0,33	0,24	0,36	0,20
T= $\mu$ - $\sigma$		<b>0,28</b>			<b>0,50</b>			<b>0,27</b>			<b>0,27</b>			<b>0,26</b>			<b>0,32</b>		

Table 4. Distances among the ICA representations of the view-morphed test images of some subjects and the representations of the training set formed by the database pre-processed images

However, note also that, in practice, it is possible to extract much more than only three images from each video sequence, which will possibly increase the number of correct matches. In fact, it is in the essence of the adopted approach that the increased number of test images, while necessarily generating many false negatives which are un-consequential, will augment the probability of detecting a few true positives in the case where the identification should occur.

Table 4 shows the distances between the ICA representations of the view-morphed images of some subjects and the representations of the training set formed by the database pre-processed images.

The analysis of the distances between all of the ICA representations of the test set and all the representations of the training set, led us to verify that, with a control system constructed under the conditions previously established, all the clients are correctly authenticated, giving an authentication rate of 100%.

### 5.2 Rejection of impostors

We have also performed a sort of “leave-one-out” tests to evaluate the capacity and the accuracy that this system could have to automatically reject the access of impostors (subjects announcing themselves with a false identity). This type of tests consists on the construction of a training set with  $n - 1$  subjects (in our case, 21), and then evaluating the differences and similarities between the ICA representations of this set and the representations of the test set formed by the images of the subject that was left out. This procedure is repeated until all the individuals of the database have been left out. Considering that subject  $j$  is left out, a training set with the remaining subjects is constructed. The distances between the ICA representations of the  $j$  subject images and the representations of the training set allows the definition of a threshold given by:

$$T_i = \mu_i - 2\sigma_i \quad (10)$$

where  $\mu_i$  and  $\sigma_i$  can be obtained by using equation (9) with  $N_{\text{images}} = 3$  and  $N_{\text{ind}} = 21$ .

In this way, the impostors will be (wrongly) authenticated if the distances between the representation of at least two of the three selected images and a representation of any subject on the training set are lower than the threshold value given by equation (10). Table 5 shows the results obtained.

With these results we can conclude that, using the threshold given by equation (10), no impostor can gain access. The previously mentioned computer takes about 12 seconds to construct the training set and to evaluate an impostor authentication. These results are valid considering each one of the 22 subjects as an impostor and the training set formed by the other 21. Aside from these results, it can also be observed that considering these 21 training sets and the threshold defined by equation (8), all the clients (the 21 individuals that formed the training sets) are correctly authenticated.







		Training Set																	
																			
		ar001			ar004			ar015			ar018			ar019			ar022		
		I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3	I1	I2	I3
Test Set	ar001				0,78	0,70	0,61	0,35	0,25	0,23	0,42	0,34	0,41	0,47	0,38	0,44	0,52	0,57	0,47
	ar002	0,55	0,56	0,60	0,65	0,62	0,60	0,61	0,59	0,58	0,82	0,81	0,62	0,92	0,80	0,85	0,98	0,99	0,65
	ar003	0,49	0,49	0,51	0,75	0,72	0,83	0,60	0,25	0,25	0,17	0,20	0,48	0,24	0,19	0,42	0,42	0,41	0,55
	ar004	0,55	0,60	0,57				0,77	0,94	0,90	0,83	0,69	0,61	0,59	0,74	0,45	0,53	0,45	0,46
	ar005	0,27	0,29	0,29	0,54	0,53	0,39	0,35	0,49	0,47	0,47	0,42	0,32	0,49	0,41	0,27	0,32	0,45	0,18
	ar006	0,33	0,32	0,42	0,47	0,43	0,29	0,58	0,56	0,56	0,71	0,65	0,57	0,55	0,53	0,31	0,49	0,49	0,31
	ar007	0,27	0,17	0,23	0,61	0,54	0,46	0,42	0,35	0,32	0,36	0,31	0,39	0,37	0,24	0,25	0,32	0,37	0,33
	ar008	0,60	0,60	0,63	0,24	0,29	0,29	0,67	0,65	0,59	0,53	0,44	0,64	0,54	0,41	0,30	0,42	0,48	0,31
	ar009	0,30	0,31	0,36	0,94	0,84	0,75	0,24	0,17	0,24	0,48	0,52	0,50	0,65	0,45	0,58	0,66	0,74	0,57
	ar010	0,35	0,25	0,33	0,81	0,71	0,62	0,46	0,27	0,25	0,48	0,40	0,49	0,52	0,40	0,44	0,59	0,65	0,49
	ar011	0,37	0,42	0,39	0,58	0,63	0,57	0,28	0,36	0,36	0,37	0,40	0,41	0,46	0,30	0,31	0,42	0,54	0,24
	ar012	0,32	0,38	0,26	0,69	0,61	0,65	0,50	0,51	0,46	0,48	0,46	0,33	0,42	0,43	0,47	0,46	0,48	0,38
	ar013	0,57	0,60	0,62	0,64	0,66	0,76	0,62	0,43	0,42	0,18	0,20	0,59	0,19	0,14	0,34	0,34	0,31	0,49
	ar014	0,42	0,44	0,45	0,88	0,79	0,90	0,43	0,26	0,29	0,27	0,31	0,43	0,42	0,31	0,58	0,53	0,57	0,66
	ar015	0,47	0,43	0,50	1,08	1,02	0,94				0,38	0,50	0,46	0,65	0,49	0,62	0,62	0,76	0,68
	ar016	0,37	0,29	0,31	0,90	0,90	0,75	0,18	0,17	0,14	0,39	0,39	0,46	0,54	0,37	0,48	0,60	0,69	0,50
	ar017	0,26	0,22	0,20	0,75	0,61	0,59	0,42	0,36	0,36	0,35	0,30	0,32	0,40	0,34	0,42	0,37	0,46	0,44
	ar018	0,53	0,50	0,51	0,80	0,80	0,88	0,46	0,23	0,20				0,22	0,15	0,40	0,33	0,37	0,52
	ar019	0,54	0,49	0,51	0,60	0,59	0,74	0,57	0,37	0,35	0,17	0,21	0,51				0,32	0,35	0,49
	ar020	0,45	0,44	0,41	0,78	0,74	0,82	0,54	0,31	0,29	0,29	0,27	0,51	0,31	0,26	0,47	0,46	0,44	0,55
	ar021	0,32	0,26	0,28	0,44	0,39	0,32	0,39	0,59	0,58	0,53	0,55	0,29	0,57	0,47	0,30	0,30	0,42	0,22
	ar022	0,45	0,42	0,43	0,57	0,57	0,59	0,50	0,38	0,40	0,31	0,29	0,44	0,37	0,24	0,33			
<b>T= <math>\mu-2\sigma</math></b>		<b>0,17</b>		<b>0,29</b>		<b>0,08</b>		<b>0,13</b>		<b>0,11</b>		<b>0,17</b>							

Table 5. Distances among ICA representations of view-morphed images of some subjects and of the training set formed by the pre-processed images of the remaining subjects

## 6. Conclusion

This work addressed the applicability of Independent Component Analysis to face recognition through image analysis, namely to the facial authentication problem.

We defined and tested the main functionality of a potentially automatic vision system. In fact, some aspects of the process have been performed manually, but it is well known and the literature clearly documents successful automated methods for those operations.

The main purpose of the system is to validate the clients' access to restricted areas and to avoid the entrance of impostors, using as few restrictions on the environment as possible, and causing virtually no discomfort on the clients, both by not forcing them to multiple image acquisitions in the database construction phase, and by allowing them to walk naturally (with no specific pose for image capture) in the direction of the entrance.

The results obtained have demonstrated the usefulness of the image pre-processing steps described and of the view-morphing techniques to generate virtual frontal views from slightly lateral images of the individuals.

The identification of clients was 100% accurate, although it can be observed from the tables in the text that the robustness of the method is not high, as there is a significant number of "near" false-positives.

On the other hand, the tests for the rejection of impostors have also proved that this most important goal could be successfully achieved for the data available.

Nevertheless, because the two experiments were conducted separately, there are limits to the usability of the results, as is implied by the fact that the computation of the thresholds for identification and for rejection resulted different (that is, one and two standard deviations from mean). This indicates that further work must be done to counter this obvious limitation.

One must take into consideration, however, that the building of the database was not thoroughly addressed, as timing constraints in the development phase of the work forced this process to be done very fast. The experience in analysing the results shows that test images of much higher quality can be acquired with only a minor investment in using one or two standard video cameras, and in setting up an adequate illumination and environment conditioning system, in particular to avoid specular reflections in the field of view.

Also, the view-morphing point-correspondence problem was solved by manually setting the correspondences and, as previously pointed out, using only a small number of feature points located on a relatively narrow vertical strip of the faces. More accurate methods for matching pairs of points in an image and its reflection can be used in an automatic mode, not only for feature points but also for other points apart from the vertical symmetry line of the face, using epipolar geometry.

As a final conclusion, one can say that the results achieved so far are quite encouraging and that the ideas for improving the methodology are well founded and promise a good outcome.

## 7. References

- Bartlett, M.; Lades, H. & Sejnowski, T. (1998). Independent Component Representations for Face Recognition, *Proceedings of the SPIE Symposium on Electronic Imaging: Science and Technology; Human Vision and Electronic Imaging III*, Vol. 3299, pp. 528-539, ISBN 9780819427397, California, July 1998, SPIE Press
- Bartlett, M.; Movellan, J. & Sejnowski, T. (2002). Face Recognition by Independent Component Analysis, *IEEE Transactions on Neural Networks*, Vol. 13, No. 6, November 2002, pp. 1450-1464, ISSN 1045-9227
- Beymer, D.; Shashua, A. & Poggio, T. (1993). Example based image analysis and synthesis, In *A. I. Memo 1431*, C. B. C. L: Paper No. 80, Artificial Intelligence Laboratory, MIT (November 1993)
- Chen, S. & Williams, L. (1993). View Interpolation for Image Synthesis, *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 279-288, ISBN 0-89791-601-8, California, August 1993, ACM Press, New York
- Draper, B.; Baek, K.; Bartlett, M. & Beveridge, J. (2003). Recognizing faces with PCA and ICA, *Computer Vision and Image Understanding*, Vol. 91, No. 1-2, July 2003, pp. 115-137, ISSN 1077-3142
- Feng, G. & Yuen, P. (2000). Recognition of Head-&-Shoulder Face Image Using Virtual Frontal-View Image, *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Cybernetics*, Vol. 30, No. 6, November 2000, pp. 871-883, ISSN 1083-4427
- Fernández-Caballero, A.; Gómez, F.; Moreno, J. & Fernández, M. (2001). Redes STN y CPN para el Reconocimiento de Rostros, *Technical Report (in Spanish)*, Department of Computer Science, University of Castilla-La Mancha, 2001
- Fortuna, J; Schuurman, D. & Capson, D (2002). A Comparison of PCA and ICA for Object Recognition Under Varying Illumination, *Proceedings of the 16th International Conference on Pattern Recognition (ICPR'02)*, Vol. 3, pp. 11-15, ISBN 0-7695-1695-X, Canada, August 2002, IEEE Computer Society, Washington DC
- Hyvärinen, A. (1999). Survey on Independent Component Analysis, *Neural Computing Surveys*, Vol. 2, 1999, pp. 94-128, ISSN 1093-7609
- Hyvärinen, A. & Oja, E. (2000). Independent Component Analysis: Algorithms and Applications, *Neural Networks*, Vol. 13, No. 4-5, May-June 2000, pp. 411-430, ISSN 0893-6080
- Hyvärinen, A.; Karhunen, J. & Oja, E. (2001). *Independent Component Analysis*, John Wiley & Sons, ISBN 978-0-471-40540-5, New York
- Jung, T.-P.; Makeig, S; Mckeown, M.; Bell, A.; Lee, T.-W. & Sejnowski, T. (2001). Imaging Brain Dynamics Using Independent Component Analysis (Invited Paper). *Proceedings of the IEEE*, Vol. 89, No. 7, July 2001, pp. 1107-1122, ISSN 0018-9219
- Lam, K. & Yan, H. (1998). An Analytic-to-Holistic Approach for Face Recognition Based on a Single Frontal View, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 7, July 1998, pp. 673-686, ISSN 0162-8828
- Lee, T.-W.; Lewicki, M. & Sejnowski, T. (2000). ICA mixture models for unsupervised classification with non-Gaussian classes and automatic context switching in blind signal separation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 10, October 2000, pp. 1078-1089, ISSN 0162-8828

- Lee T.-W. & Lewicki, M. (2002). Unsupervised Image Classification, Segmentation, and Enhancement Using ICA Mixture Models. *IEEE Transactions on Image Processing*, Vol. 11, No. 3, March 2002, pp. 270-279, ISSN 1057-7149
- Rajapakse, M. & Wyse, L. (2003). NMF vs ICA for Face Recognition, *Proceedings of the 3rd International Symposium on Image and Signal Processing and Analysis*, pp. 605-610, ISBN 953-184-061-X, Italy, September 2003, IEEE Computer Society Press
- Sebastião, R. (2006a). Autenticação de Faces a partir da Aquisição de Sequências de Imagens, *MSc thesis (in Portuguese)*, Faculdade de Ciências e Faculdade de Engenharia da Universidade do Porto, Porto, Janeiro 2006
- Sebastião, R.; Silva, J. & Padilha, A. (2006b). Face Recognition from Spatially-Morphed Video Sequences, *Proceedings of International Conference on Image Analysis and Recognition*, pp. 365-374, ISBN 3-540-44894-2, Póvoa de Varzim, September 2006, Springer LNCS 4142, Berlin
- Seitz, S. & Dyer, C. (1995). Physically-valid View Synthesis by Image Interpolation, *Proceedings of IEEE Workshop on Representations of Visual Scenes*, pp. 18-25, ISBN 0-8186-7122-X, Massachusetts, June 1995, IEEE Computer Society Press
- Seitz, S. & Dyer, C. (1996). View Morphing, *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH'96)*, pp. 21-30, ISBN 0-89791-746-4, New Orleans, August 1996, ACM Press, New York
- Seitz, S. (1997). Image-Based Transformation of Viewpoint and Scene Appearance, *PhD Thesis*, University of Wisconsin-Madison, Madison, 1997
- Torres, L.; Lorente, L. & Vila, J. (2000). Automatic Face Recognition of Video Sequences Using Self-eigenfaces, *International Symposium on Image/video Communication over Fixed and Mobile Networks*, Rabat, Morocco, April 2000
- Turk, M. & Pentland, A. (1991). Eigenfaces for Recognition, *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, Winter 1991, pp. 71-86
- Ullman, S. & Basri, R. (1991). Recognition by Linear Combination of Models, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 13, No. 10, October 1991, pp. 992-1006, ISSN 0162-8828
- Xiao, J. & Shah, M. (2004). Tri-view Morphing, *Computer Vision and Image Understanding*, Vol. 96, No. 3, December 2004, pp. 345-366, ISSN 1077-3142
- Yang, J.; Frangia A. & Yang, J.-y. (2004). A new kernel Fisher discriminant algorithm with application to face recognition, *Neurocomputing*, Vol. 56, January 2004, pp. 415-421, ISSN 0925-2312





## **Face Recognition**

Edited by Kresimir Delac and Mislav Grgic

ISBN 978-3-902613-03-5

Hard cover, 558 pages

**Publisher** I-Tech Education and Publishing

**Published online** 01, July, 2007

**Published in print edition** July, 2007

This book will serve as a handbook for students, researchers and practitioners in the area of automatic (computer) face recognition and inspire some future research ideas by identifying potential research directions. The book consists of 28 chapters, each focusing on a certain aspect of the problem. Within every chapter the reader will be given an overview of background information on the subject at hand and in many cases a description of the authors' original proposed solution. The chapters in this book are sorted alphabetically, according to the first author's surname. They should give the reader a general idea where the current research efforts are heading, both within the face recognition area itself and in interdisciplinary approaches.

### **How to reference**

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Armando Padilha, Jorge Silva and Raquel Sebastiao (2007). Improving Face Recognition by Video Spatial Morphing, Face Recognition, Kresimir Delac and Mislav Grgic (Ed.), ISBN: 978-3-902613-03-5, InTech, Available from:

[http://www.intechopen.com/books/face\\_recognition/improving\\_face\\_recognition\\_by\\_video\\_spatial\\_morphing](http://www.intechopen.com/books/face_recognition/improving_face_recognition_by_video_spatial_morphing)

**INTECH**  
open science | open minds

### **InTech Europe**

University Campus STeP Ri  
Slavka Krautzeka 83/A  
51000 Rijeka, Croatia  
Phone: +385 (51) 770 447  
Fax: +385 (51) 686 166  
[www.intechopen.com](http://www.intechopen.com)

### **InTech China**

Unit 405, Office Block, Hotel Equatorial Shanghai  
No.65, Yan An Road (West), Shanghai, 200040, China  
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元  
Phone: +86-21-62489820  
Fax: +86-21-62489821



© 2007 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the [Creative Commons Attribution-NonCommercial-ShareAlike-3.0 License](https://creativecommons.org/licenses/by-nc-sa/3.0/), which permits use, distribution and reproduction for non-commercial purposes, provided the original is properly cited and derivative works building on this content are distributed under the same license.

IntechOpen

IntechOpen