Use of Depth and Colour Eigenfaces for Face Recognition

F. Tsalakanidou\textsuperscript{a}, D. Tzovaras\textsuperscript{b,*}, M. G. Strintzis\textsuperscript{a,b}

\textsuperscript{a}Aristotle University of Thessaloniki, Electrical and Computer Engineering Department, Information Processing Laboratory, Thessaloniki 54124, Greece
\textsuperscript{b}Informatics and Telematics Institute, Centre for Research and Technology Hellas, 1st km Thermi-Panorama Road, Therni 57001, P.O. Box 361, Thessaloniki, Greece

Abstract

In the present paper a face recognition technique is developed based on depth and colour information. The main objective of the paper is to evaluate three different approaches (colour, depth, combination of colour and depth) for face recognition and quantify the contribution of depth. The proposed face recognition technique is based on the implementation of the Principal Component Analysis algorithm and the extraction of depth and colour eigenfaces. Experimental results show significant gains attained with the addition of depth information.

Key words: colour, depth, 3D model, Principal Component Analysis, face recognition

1 Introduction

The problem of automatic face recognition is a complex task that involves detection and location of faces in a cluttered background followed by normalization and recognition. Considering it as a pattern, the face is a very challenging object to detect and recognize. The anatomy of the human face

\* Corresponding author.
Email addresses: filareti@panorama.ee.auth.gr (F. Tsalakanidou), tzovaras@iti.gr (D. Tzovaras).
is rigid enough so that all faces have the same structure, but they are at the same time very different from each other due to gender, race, and individual variations, as well as due to facial expressions. A robust detection-recognition system must also overcome variations due to lighting conditions, rotations of the head, complex background, etc.

Over the past 25 years many different face recognition techniques were proposed, motivated by the increased number of real world applications requiring recognition of humans. Turk and Pentland (1991) used eigen-faces for face detection and identification. Face recognition techniques using Linear/Fisher Discriminant Analysis (LDA), e.g. Zhao (1998), or Support Vector Machines, e.g. Heisele (2001), for the classification were also developed. Samaria (1994) employed Hidden Markov Models for face recognition and his method was further expanded with the use of embedded HMMs by Nefian (1999). Wiskott et al (1997) used elastic graph-matching (EGM) for face recognition. A variant of EGM based on multiscale erosion-dilation, was proposed by Tefas (2001). Local Feature Analysis (LFA), a technique related to eigenfaces, but being more capable of accommodating changes in appearance or facial aspect was presented by Penev (2000). 3D techniques based on the extraction of 3D facial features from depth images by means of differential geometry techniques, e.g. Tanaka (1998), or Point Signatures, e.g. Chua (2000), have also been proposed for face identification.

In the present paper a face recognition technique is developed based on depth and colour information. The main objective of the paper is to evaluate three different approaches (colour, depth, combination of colour and depth) for face recognition and quantify the contribution of depth in face recognition. The face recognition technique proposed is based on the Principal Component Analysis algorithm and the extraction of depth and colour eigenfaces. Experimental results show significant gains attained with the addition of depth information.

The paper is organized as follows. One methodology for the extraction of depth maps is described in Section 2. The general method for the generation of eigenfaces and the formulation of the eigenfaces method for use in face recognition based on colour and depth information is developed in Section
3. Experimental results evaluating the developed techniques are presented in Section 4. Finally conclusions are drawn in Section 5.

2 Depth Extraction

Most face recognition systems are based on the evaluation of 2D-intensity or colour images. Since the extraction of reliable features from 2D-images is difficult and is subject to a variety of possible interpretation errors, the recognition accuracy of such systems is limited on a small set of individuals. The use of additional 3D information is expected to improve the reliability of the recognition scheme, since biometric features are reliably extracted from 3D measurements, due to their relative independence from illumination influences, head rotation, face pigment and facial expressions.

One way of exploiting the 3-D information of the geometric characteristics of a human face is by building the “depth map” of the face (Tzovaras (1998)). Let us consider a virtual camera and an object in front of it. The depth map of the object is a function giving for each pixel of the image the depth of the corresponding 3D point, i.e. the distance of the point from the camera plane. The object, human face in this case, is considered to consist of many polygons (usually triangles), which properly arranged, give the impression of human face skin texture. Creating the depth map is actually a process of visible surface determination. In other words it is a procedure that requires examining all of these polygons to determine which is closest to the camera along the projector passing through the pixel, and that is done for all the pixels of the image. One well known visible surface determination technique is the z-buffer algorithm proposed by Catmull (1974).

The extraction of the depth map requires the existence of a 3D face model. One successful approach for 3D data acquisition is active triangulation based on structured light (e.g. Forster (2001)). This method is based on the projection of a known colour pattern onto the object to be measured. The luminance of this pattern is measured and classified in the image of the object for 3D-
reconstruction. From the result of the projection of the light to the human face, 3D information is extracted, which consists of a set of 3D coordinates of points. Then a triangulation technique is applied to the specific set of points and a 3D model of the head comprised of triangles is acquired. The depth map of the face is extracted by projecting the 3D face model to the camera plane as described above.

3 Face Recognition Based On Depth and Colour Eigenfaces

The main objective of the paper is to evaluate three different approaches (colour, depth, combination of colour and depth) for face recognition and quantify the contribution of depth in face recognition. The eigenface technique, which has already been used on intensity and colour images, is employed for the experimental evaluation of the aforementioned approaches. Depth information is treated exactly as intensity information, the only difference being that pixel intensity is replaced by depth values. Note here that most 3D face recognition techniques proposed in literature are based on the extraction of 3D facial features by means of differential geometry techniques (e.g. Tanaka (1998)) or Point Signatures (e.g. Chua (2000)). In this paper we propose a powerful global approach (PCA), in which the whole depth image serves as a feature vector, and the variability of the face is modelled by analyzing its statistical properties using a large set of depth maps.

PCA is a well-known, widely and successfully used method, based on the idea that face recognition can be based on a small set of features that best approximate the set of known face images and may not correspond to our intuitive notions of facial parts and features (Turk (1991)). The application of PCA for face recognition is very simple: PCA is performed in a well-defined set of images of human faces and a set of $M$ principal components is obtained. Every person is represented by number of different images, in which various expressions and slightly different poses are captured.

Given the eigen-faces, every face in the database can be represented as a vector
of weights; the weights are obtained by projecting the image into eigen-face components by a simple inner product operation. When a new test image whose identification is required is given, its vector of weights also represents the new image. The identification of the test image is done by locating the image in the database whose weights have the smallest Euclidean distance from the weights of the test image. The Nearest Neighbour Classification rule we employ (see Turk (1991)), is only one of the classifiers that can be used with PCA. Other classifiers (such as Neural Networks or Support Vector Machines) can also be used to classify the projected features. It has been observed experimentally that this approach degrades quickly as the scale changes. Intuitively, this is explained by the low correlation between face images at different scales. One way to overcome and eliminate the resulting problem is to apply the detection algorithm in linearly scaled versions of the test image. Of course, all faces in the database, including those in the training set must be uniformly scaled.

3.1 Use of colour images

The above analysis is valid for single channel signals. In case of face recognition from full colour images, the above analysis is still valid for each component of the colour signal. We assume that the colour images are available in (or have been transcribed to) YUV format to ensure independence. We use the MPEG-compatible YUV format 4:2:0 for colour images. This format has a horizontal subsampling period of 2 for both the U and V components indicating that U and V samples are taken for every second pixel across a line. Their vertical subsampling period is also 2 indicating that U and V samples are taken on every second line of the image. This means that if the image dimensions are \( N \times N \), then the U and V components of the image have dimensions \( (N/2) \times (N/2) \). Subsampling the colour components U and V not only does not affect the accuracy of the face recognition task, as shown by experimental tests performed, but it makes the system faster.

The PCA algorithm is performed in each of the three training sets (one for
each of the components YUV) and three sets of eigenvectors are obtained. For each class (face), three sets of projections to the respective PCA subspaces are calculated and stored. When a new image needs to be recognized, then its projections to the three subspaces are calculated and compared to those already stored. Three Euclidean distances are calculated (one for each component). The test image is assigned to the class \( k \), for which the smallest product of the Euclidean distances is calculated. Taking the Euclidean distance of the combined YUV space is an alternative approach. Experimental results have shown that using the product of the Euclidean distances offers an advantage of 2% over using the total Euclidean distance of the combined space.

\[ \text{3.2 Use of depth maps} \]

Up to this point the eigenface technique for face recognition was applied only on captured 2D gray-scale or colour images of faces. The use of 3D images of faces supplies additional features for face recognition, and thus, as the experimental results show, enhances the performance of the recognition algorithms. One way of exploiting the 3D information is by using depth maps of the faces as explained in detail in Section 2. Thus, the 3-D information of the human face is transformed into a gray scale 2-D image. The eigenface technique is extended to handle depth maps. The training and testing procedures are exactly the same, as in the case of single channel signals (see the analysis for gray-scale images in the second and third paragraph of Section 3), the only difference being that the training and testing sets consist of depth maps.

\[ \text{3.3 Use of both depth and colour images} \]

Up to this point, most face recognition systems rely only on a single type of facial information: 2D gray or colour images or 3D depth maps. We argue that integration of colour and depth may lead in increased recognition rate, since such an approach will exploit the benefits of both 2D and 3D information and will make it possible to overcome their individual shortcomings (Mavridis (2001)). The eigenface method is once again used. The depth map and one
front view of each person are fed on two independently working recognition systems (one using colour information and the other using depth information). For each person of the training ensemble two Euclidean distances are calculated (one by each system) and multiplied. The person to whom the smallest product corresponds to, is recognized as the person in the test images.

4 Experimental Results

For experimental evaluation of the proposed approaches we used the XM2VTS database of the University of Surrey (Messer (1999)), which is made up of the faces of 295 volunteers and provides 12 images for each person. The images were taken from the video recordings made during the acquisition of the database. There were four sessions recorded at approximately 1-month intervals. From each session 3 images were taken for each person: 1 frontal view, 1 left profile and 1 right profile. The colour images were stored in portable pixmap format (ppm) with resolution $720 \times 576$ pixels. In addition, for each person a 3D VRML model was given, followed by a JPEG file providing the necessary texture information. The 3-D VRML model of the head was used for the creation of the depth map of each person. The “structured light” approach (Forster (2001)) was used for capturing the 3D facial surface and thus creating the VRML files (Messer (1999)).

4.1 Using colour information alone

The training procedure was based on a set of 80 images from the first and the fourth session of XM2VTS. Each person is represented by 2 frontal views. The original images were scaled down to a size of $100 \times 80$ pixels and a window with dimensions $42 \times 66$ covering the head area, was used for the training procedure (Figure 1). The colour eigenfaces resulted after training are shown in Figure 2. The testing set consisted of the remaining 80 images (sessions 2 and 3) of the same 40 people, again scaled down to a $100 \times 80$ pixel size.

The algorithm presented by Tzovaras (1999) was used to detect the position of
the head in various predetermined scales. For face detection the first 3 eigenvectors of each colour component (YUV) were used. As shown by Tzovaras (1999) this technique has an detection rate of 99%. Indeed in all cases the position of the head was determined with almost perfect accuracy. The result is a window covering the head area in which the face recognition algorithm described in section 3 is applied. The recognition rate in terms of the number of eigenvectors used is shown in Figure 6a. The recognition technique we implemented reached a recognition rate as high as 87.5% with the use of 30 colour eigenfaces.

The rate of successful recognition may be improved if two more images for each person are used: one left and one right profile. This of course presupposes the simultaneous use of three cameras for the recognition. In this case we used three different recognition systems and their corresponding training sets: one for the front views, one for the right profiles and one for the left profiles. Each training set consisted of 80 images from the first and the fourth session of XM2VTS, as described above (See Figure 1). For each set a corresponding PCA subspace was extracted. The testing set consisted of 80 triplets of images from sessions 2 and 3 (again the 3 images of each person in both sessions).
The testing procedure was performed as follows: in the entrance of the full recognition system 3 different images of the same person are present: 2 profiles and 1 frontal view. Each one of them is fed in the corresponding subsystem, which operates independently from the other two, precisely as described for the case of a single image. In this way 40 Euclidean distances are computed, one for every person of the sample as in Section 3. The 3 distances of each person are then multiplied and the person in the entrance of the system is recognized as that person of the training set, for whom the smallest product was calculated. Results show significant improvement of the performance of the recognition task. A correct-recognition rate of 87.5% is achieved using 10 eigenvectors, whereas in the former approach this percentage was reached using 30 eigenvectors (Figure 6a). Using 20 eigenvectors the recognition rate verged to 93.75%. Use of higher numbers of eigenvectors did not further improve the recognition rate.

4.2 Using depth information alone

For the experimental evaluation of the recognition technique of Section 3.2, which uses depth information, we used the VRML models of the XM2VTS database. Such a model consisting of triangles is shown in Figure 3. Using the z-buffering algorithm (Catmull (1974)) we obtained the respective depth maps from the 3-D surfaces (see the second paragraph of Section 2). These are
2-D gray-scale images, with the pixel intensity 0(black) corresponding to the background and the pixel intensity 255(white) corresponding to that point of the face, which is nearest to a virtual camera in front of the face (this point, if someone is staring at the camera face to face, is the tip of the nose). The basic idea of the z-buffer algorithm is to test the \(z_{\text{depth}}\) of each surface to determine the closest (visible) surface. Declare an array \(z_{\text{buffer}}(x,y)\) with one entry for each pixel position. Initialize the array to the maximum depth.

Then the algorithm proceeds as follows:

for each polygon \(P\) (triangle in our case)

for each pixel \((x,y)\) in \(P\)

compute \(z_{\text{depth}}\) at \((x,y)\)

if \(z_{\text{depth}} < z_{\text{buffer}}(x,y)\) then

\(z_{\text{buffer}}(x,y) = z_{\text{depth}}\)

Computation of the \(z_{\text{depth}}(x,y)\) uses coherence calculations similar to the x-intersection calculations. It is actually a bi-linear interpolation, i.e., interpolation both down \((y)\) and across scan lines \((x)\).

The training procedure we followed is exactly the same as the one using grayscale information. The training set was comprised of 40 depth maps, one for each person, with dimensions \(74 \times 66\) (see Figure 4). The PCA algorithm
was performed on the ensemble and the most important eigenvectors were stored (see Figure 5). The detection and recognition tasks are performed exactly as in the case of single channel signals (Moghaddam (1995)). Since there is only one VRML model available for each person, we also used for the testing set depth maps resulting after small rotations of the 3-D models around the vertical axis y and the horizontal axis x. For a rotation of ±2° around the vertical axis the recognition rate is as high as 93%, while for a rotation of ±5° it falls off to 89%. For larger rotations (±10°) the recognition rate reduces to 85%.

The above technique was also applied to depth maps corrupted by Gaussian noise of zero mean value and statistical variation \( \sigma = 20dB \). The recognition rate in terms of number of eigenvectors used is shown in Figure 6b.

4.3 Combining colour and depth information

For the experimental evaluation of the recognition technique based on both colour and depth described in Section 3.3, the test set of the second system consisted of depth maps resulting after a rotation of ±10° of the 3-D models around the vertical axis y. The combination of colour images and depth maps leads to a remarkably high recognition rate, even when only one colour image is used. Using 20 eigenvectors the recognition rate becomes 97.5%, while when we use only frontal colour images the maximum recognition rate is 87.5% with use of 30 eigenvectors. When we additionally use a profile view, then a recognition rate of 98.75% is achieved (Figure 6c).

4.4 Comparison of face recognition techniques

As already stated, the main contribution of this paper is the evaluation of different approaches that use colour, depth or combination of colour and depth information. The use of different views of the human face was also exploited. In Figure 6d the results obtained using the methods described above are summarized. These results are also depicted in Table 1. The maximum recognition
rate achieved by each method is highlighted. It can be clearly seen that the use of depth information leads in a increased recognition rate. Use of different views of the face may also significantly help the recognition task, mainly when colour images are used, but offers only a relatively small benefit (1%), when depth images are also used, since depth alone is a very powerful cue for recognizing a face.

5 Conclusions

In the present paper a face recognition technique is developed based on depth and colour eigenfaces. The recognition technique presented is based on the implementation of the Principal Component Analysis algorithm and the extraction of the colour and depth eigenfaces. The main objective of the paper was to evaluate three different approaches (colour, depth, combination of colour and depth) for face recognition and quantify the contribution of depth in face recognition. Experimental results show significant gains with the use of depth information. The improvement varies according to the method chosen and ranges between 2% and 5% as shown in Figure 6d. The technique based on depth alone appears to be robust in variations due to noise or slightly different poses and rotations of the human head. The algorithm based on the use of colour and luminance as well as depth was seen to lead to very high recognition rates.

References


Fig. 6. Cumulative recognition rates: (a) Colour frontal views and combination of colour frontal and profile views, (b) Depth maps (The test set is comprised of rotated or embedded to noise depth maps), (c) Combination of depth and colour information, (d) Comparison of all methods tested.

<table>
<thead>
<tr>
<th>Recognition Method</th>
<th>Number of eigenvectors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Use of colour frontal views</td>
<td>75%</td>
</tr>
<tr>
<td>Use of colour frontal view &amp; left profile &amp; right profile</td>
<td>80%</td>
</tr>
<tr>
<td>Use of depth maps (rotation 10 degrees)</td>
<td>55%</td>
</tr>
<tr>
<td>Use of depth map &amp; colour frontal view</td>
<td>91.25%</td>
</tr>
<tr>
<td>Use of depth map &amp; colour frontal view</td>
<td>93.75%</td>
</tr>
</tbody>
</table>

Table 1
Cumulative recognition rates obtained using different recognition methods.