The HISCORE face recognition application: Affordable desktop face recognition based on a novel 3D camera

N. Mavridis, F. Tsalakanidou, D. Pantazis, S. Malasiotis, M. Strintzis

Informatics and Telematics Institute (ITI), 1st km Thermi-Panorama Road,
P.O.Box 361, GR-57001 Thermi, Thessaloniki, Greece

Email: nmav@iti.gr, filareti@iti.gr, malasiot@iti.gr

ABSTRACT

A toolchest of modules useful for 3D object recognition, and in particular for face and gesture recognition, based on color image and depth map data, is currently being developed. Using the early versions of these modules, a prototype face recognition application has already been built. Our results so far have indicated that our goal of an affordable desktop or even embedded face recognition system with robust and reliable performance using a novel cheap 3D camera subsystem based on the CCLA method (Color Coded Light Approach) is well within reach, and that the extra depth map information that is available in this case can significantly aid in simplifying and speeding up this application, which can help enhance many real-world surveillance, security and HCI (Human-Computer Interface) systems.

1. INTRODUCTION

Face recognition is an area which has received increased attention during the last decade. It now seems to be reaching a stage of maturity in terms of performance and cost, that will allow the widespread application of systems including a face recognition component. Some of the most important landmarks in the history of this area include the FERET international face recognition competitions that took place in the last decade, and several standard methods that have been developed including PCA (Principal Components Analysis) techniques (MIT), and elastic graph matching (Ruhr-Uni Bochum), as well as the emergence of commercially available software (such as FaceIt etc.) [1],[2].

In this paper we describe a face recognition system developed in the framework of the EU HiScore project. The main novel element of HiScore is the use of an affordable CCLA 3D+colour camera. Here we will describe our system, and show how the extra depth and colour information available is exploited.

This work was supported by the EU Project HiScore (IST-1999-10087)

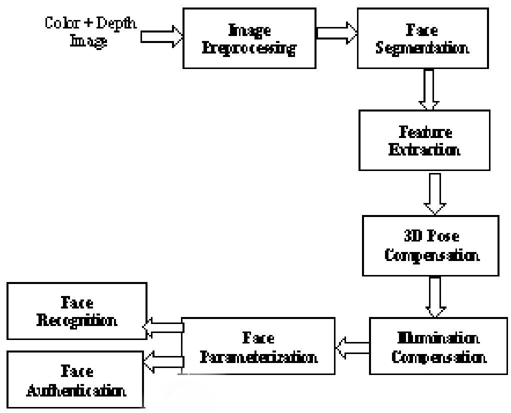

Figure 1: Face Recognition System Architecture

2. OVERVIEW OF THE SYSTEM

The input to the system is a color image together with an associated depth map taken by the camera subsystem. The system can be used in two modes: authentication and recognition. The output of the system is either a binary answer when the system is used for authentication of an a priori known person (i.e. this IS George or this IS NOT George), or in the case of recognition, the code of the nearest person of an a priori given face set together with a reliability estimate (i.e. out of the known faces this is most probably GEORGE with a 0.75 safeness factor).

Several types of recognition (or authentication) engines are considered, along with a number of variations to incorporate the extra 3D and colour data available. A PCA based approach is already working, as well as a simple 1D HMM based engine, and elastic graph matching techniques are considered, also in order to provide the authors with hands-on experience with these systems.

Something worth stressing here, that is well-known to everybody that has been involved in face recognition, is the importance of the stages performing image rectification and normalization, preceeding the actual recognition engine. Thus, image preprocessing, face detection and segmentation, facial feature extraction, pose compensation, illumination compensation and face parameterization stages have been created, and their prototypes are already operational (see 1). Here, again, the extra 3D+colour information available is very useful: as an example, facial feature extraction is greatly simplified by the use of 3D curvature, skin colour segmentation simplifies face detection etc.

3. IMAGE PREPROCESSING AND FACE SEGMENTATION

The images taken from the HISCORE 3D camera are expected to have holes due to occlusions, noise and other artifacts. The purpose of the image preprocessing module is to apply image conditioning by means of simple filtering operations, and more sophisticated hole covering techniques.

The aim of the face segmentation module is to locate faces in the image. What this face region that is segmented should contain depends on the further stages. The module should also cope with small partial occlusion of the face, people wearing glasses or having beards etc. The output of the module is a list of candidate image segments with high probability of being faces. Here, we have developed a module based on skin colour segmentation as well as one based on PCA and maximum likelihood detection methods (see[7]) ,and are currently developing HMM-based modules for face detection and segmentation.

4. FACIAL FEATURE EXTRACTION AND POSE COMPENSATION

A face candidate image is the input to the facial feature extraction stage. The output is a structure containing the estimates of the positions of certain useful face points, including nose, eyes, the point immediately below the nose and above the mouth etc.

Here, we are currently using a technique based on local extrema of gaussian and mean curvature of the depth map, and a priori knowledge of the allowable geometry of the extracted points [3]. However, we are also developing a module based on symmetry and attentional fixation points [4], probably complementary to the curvature module. Some results from the curvature based module are shown in 2.

Figure 2: Facial Feature Extraction Results

Figure 3: Pose Compensation Results

The aim of the pose compensation module is to calculate a rigid transformation that aligns the 3D-face surface with a predefined face pose (e.g. a frontal facing head). This transformation is calculated as a function of the positions of the facial features found by the facial feature extraction module. Based on this transformation, and camera calibration parameters, warping of depth and colour images and associated segmentation mask is performed. An example of the attained pose compensation is shown in 3.

A thorough review of some aspects of pose compensation can be found in [5]. Some internal implementation problems that have been addressed here also deal with adaptive fine-tuning of parameters, obstructed area reconstruction, accuracy of compensation estimation etc.

5. ILLUMINATION COMPENSATION

The aim of this module is to compensate for the effect of the illumination on the colour images. A simplified lighting model is assumed. This consists of a distant directional diffuse light source (this is similar to the daylight coming through a window), while the object surface is assumed to have Lambertian reflection properties. This assumption, although idealized, turns out to be a fairly realistic approximation of many surfaces including human skin. Making the above assumptions, the rendering equation is:

I(x,y)=Io(x,y)(a+d n(x,y) L)

where I(x,y) is the intensity image, Io(x,y) is called the surface albedo, and it represents the deviation in reflectance properties due to pigmentation or markings on the surface, n(x,y) is the surface normal computed from the depth images, L is a 3D vector representing the direction of the light, and a,d are scalars that are proportional to the contribution of the ambient and diffuse illumination on the surface.

We simplify the problem by assuming a surface with constant albedo, taken to be equal to the mean skin-colour. After the illumination model parameters have been estimated the illumination compensated image is easily computed. Assuming that the environmental lighting conditions do not change, estimation of the illumination model may be performed only once and not for every captured image frame.

6. FACE RECOGNITION AND AUTHENTICATION

The eigenface technique was proposed by Turk and Pentland [6] for the purpose of face authentication and recognition. The application of PCA for face authentication and recognition is very simple: the Principal Component Analysis is performed in a well-defined set of images of human faces and a set of M principal components (eigen-vectors) is obtained. The training set consists of the images of K different persons. Every person is represented by a predefined number of different images, in which various expressions and slightly different poses are captured. It is often desirable to additionally use images in which people have different make up and hairstyles, wear or not wear glasses etc., in order to overcome the problems that the aforementioned variations in human face appearance could cause accurate authentication. The number M of the principal components can be chosen dynamically depending on the database of face images. It is usually smaller than the number NT of the images of the training set.

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. It has been observed experimentally, that this approach is fairly robust to changes in lighting conditions, but degrades quickly as the scale changes. Intuitively, this is explained by the significant correlation present between images under varying in illumination conditions and 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 as described in detail in the previous section. Of course, all faces in the database, including those in the training set must be uniformly scaled.

Figure 4: The first 3 color and depth eigenfaces

Since there will always be one person in the training sample, minimizing the Euclidean distance regardless of the fact that the person on the new image may not belong to the training sample, we must impose the constraint that the minimum distance is smaller than a predefined threshold. This way we avoid the possibility of a face outside the database is mistaken for one within it. The threshold separates the images in two categories: those of people belonging to the database set and those of "strangers".

Up to this point, the eigen-faces technique for face authentication, was applied only on greyscale or colour images of faces. The use of 3D information supplies additional features for face authentication, and thus as the experimental results show, enhances performance of authentication algorithms. The eigen-face technique described above was extended to also handle depth maps. The depth map and one front view of each person are read lexicographically, so that the vector that represents the image has the following arrangement: R G B Depth R G B Depth. Of course for the success of the particular technique an accurate alignment between the depth map and the corresponding colour image is required. The three first color and depth eigenfaces when a combination of depth and colour information is used, are shown in 4.

At this stage, we already have prototypes of PCA and HMM based engines. However, we are considering several variations in the above themes, such as LFA (Local Feature Analysis) [8], several 2D embedded HMM variants [9] etc. Also, the partial use of neural networks is under consideration, for example for parts of the classification problem (see the future work paragraph).

7. FUTURE WORK

Future work, apart from the refinements required to everything mentioned above, will evolve in the following directions:

Modules for classification of faces according to gender, race or other identifiable features such as hair color, eye color etc. (see for example [10])

Expression compensation, which probably will also include some form of expression classification as a byproduct.

Creation of a standard face database taken from the HiScore camera or at least with quality similar to that expected from our camera, has started. This database consists of pictures with predetermined intraclass variation (i.e. different poses, expressions, lighting conditions etc.)

Techniques for storage, maintenance and updating of the known database of persons is a problem under consideration. For example, several papers exist on themes such as sequential KL-basis extraction, merging and splitting eigenmodels etc. ([11]). Extensions and practical considerations for these methods are being developed, for the case of PCA engines. Also, several other techniques can be used in the case of HMM engines.

Quantitative evaluation of the performance of our system, using FERET-type and post-FERET methodologies ([12]).

8. CONCLUSION

A compact and affordable real-world face recognition system is being developed, which is robust to intra-class variations, and which effectively exploits the 3D data that is available from the novel structured light cameras. The first results seem very promising, and justify our conjecture that the depth map information can significantly aid to simplifying the implementation and enhancing the performance of a face recognition system as a whole.

9. REFERENCES

[1] M. Turk, A. Pentland, “Eigenfaces for Recognition” *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991.

[2] J. Zhang, Y. Yan and M. Lades “Face Recognition: Eigenface, Elastic Matching, and Neural Nets” *Proceedings IEEE*, Vol. 85, No. 9, Sep. 1997.

[3] Hallinan et al,“Two and Three Dimensional Patterns of the Face” *A. K. Peters Natick MA*.

[4] G. Sela and M. D. Levine, “Real-Time Attention for Robotic Vision”, *Real-Time Imaging* 3, 173-194 (1997).

[5] T. S. Jebara, “3D Pose Estimation and Normalisation for Face Recognition”, *Thesis EE Dept, McGill* 1995.

[6] M. A. Turk and A. P. Pentland, “Face recognition using eigenfaces”,*Proc. International Conference on Pattern Recognition*, 1991, pp. 586-591

[7] B. Moghaddam and A. Pentland, “Probabilistic Visual Learning for Object Detection”, 5th International Conference on Computer Vision, Cambridge, MA, June 1995

[8] P.S. Penev and J.J. Atick. “Local feature analysis: a general statistical theory for object representation”, *Network: Computation in Neural Systems*, 7(3):477–500, 1996.

[9] Ara V. Nefian and Monson H. Hayes Iii (Georgia Institute of Technology), “An Embedded HMM Based Approach For Face Detection And Recognition”, *IEEE International Conference on Acoustics, Speech and Signal Processing 1999, Phoenix AZ.*

[10] S. Gutta, J. Huang, P. Jonathon, H. Wechsler ,“Mixture of Experts for Classification of Gender, Ethnic Origin, and Pose of Human Faces”, *IEEE Transactions on Neural Networks*, Vol.11, No.4, July 2000.

[11] Levy and Lindenbaum, “Sequential Karhunen-Loeve Basis Extraction and its Application to Images”, *IEEE Transactions on Image Processing*, Vol.9, No.8, Aug.2000

[12] D. M. Blackburn, M. Bone, P. Phillips, “Facial Recognition Vendor Test 2000, Evaluation Report”