

A FACE AND GESTURE RECOGNITION SYSTEM BASED ON AN ACTIVE STEREO SENSOR

S. Malassiotis, F. Tsalakanidou, N. Mavridis, V. Giagourta, N. Grammalidis and M. G. Strintzis

Informatics & Telematics Institute, Thessaloniki, Greece,
www.iti.gr
e-mail: malasiot@iti.gr

ABSTRACT

The paper presents several novel 3D image analysis algorithms, applied towards the segmentation and modeling of faces and hands. These are subsequently used to build a face-based authentication system and a system for human-computer interaction based on static and dynamic gestures. The system relies on an active stereo sensor that uses a structured light approach to obtain 3D information. In this paper we demonstrate how the use of 3D information may significantly improve the efficiency of traditional face and gesture recognition techniques that use 2D images only.

1. INTRODUCTION

A 3D & colour camera acquiring the 3D as well as the color images has been developed. This is based on an active 3D principle, making use of an improved and extended version of the well-known Coded Light Approach (CLA) for 3D- data acquisition. The CLA is extended to a Color Coded Light Approach (CCLA). The developed 3D camera achieves fast image acquisition (12 image pairs per second) and is based on low cost devices, an off-the-shelf CCTV-color camera and a standard slide projector [1].

The developed technology will be applied to two main application areas: face recognition and gesture recognition. In this paper several novel 3D-image analysis algorithms are presented aiming to face and hand segmentation and modeling. These constitute the building blocks of the face and gesture recognition applications. Experimental results demonstrate that improved efficiency and robustness may be achieved through the use of 3D information.

2. FACE RECOGNITION SYSTEM

Over the past 20 years extensive research has been conducted by psychophysicists, neuroscientists and engineers

This work has been supported by the EU project HISCORE (IST-1999-10087).

on various aspects of face recognition by humans and machines. Over the last five years, increased activity has been seen in tackling problems such as segmentation and location of a face in a given image, and extraction of features such as eyes, mouth, etc. Also, numerous advances have been made in the design of statistical and neural network classifiers for face recognition. A review of most representative face recognition techniques may be found in [2].

The majority of the work in face recognition has revolved around methods and systems, which use data obtained from a 2D intensity image. Another topic less studied by researchers is face recognition from range image data. A range image contains the depth structure of the object in question. Gordon in [3] describes a template based recognition system involving descriptors determined from curvature calculations of range image data. The data is obtained from a rotating laser scanner system with resolution of better than 0.4 mm. Segmentation is done by classifying surfaces into planar regions, spherical regions and surfaces of revolution. At each point on the surface of the curve the magnitude and direction of the minimum and maximum normal curvatures are calculated and subsequently used to locate facial features such as the nose, the eyes and mouth. This information is then used for depth template comparison. Using the location of the eyes, nose and mouth the faces are normalised into a standard position. The volume of space between two normalised surfaces is used as the criterion for a match.

Atick, Griffin and Redlich [4] applies the idea of PCA and a Gaussian model to three-dimensional heads. A set of "eigen-heads" is obtained by doing principal component analysis on a data set of 3D head scans, aligned with each other. A generative model is used to model image formation, which assumes a simple Lambertian lighting model. This is subsequently applied towards solving the shape-from-shading problem for face images.

The above techniques rely on a database of very accurate depth images. In our case, depth maps have limited accuracy and may have areas where depth could not be determined due to occlusions. Therefore, in our approach depth

information is used in combination with colour.

The developed face authentication system described in this paper consists of two distinct stages. An image rectification stage and a face recognition stage. At the first stage the raw input images are appropriately rectified or transformed. The role of this stage is to remove image variations that are irrelevant to the discrimination of the face. For example, this stage accounts for variations in location and orientation of the face, or variations caused by changes in illumination. The second stage performs the actual matching of the probe image with the images stored in a face database. This matching is commonly based on pertinent image features extracted from the rectified images.

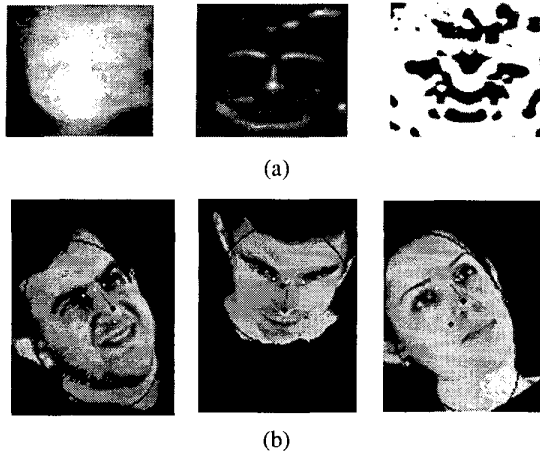


Fig. 1. 3D facial feature extraction. (a) Depth map, mean curvature, and Gaussian curvature. (b) Results showing nose, nose-ridge and eye detection on synthetically rotated heads.

The basic steps for image rectification is pose and illumination compensation. Pose compensation relies on a novel automatic feature extraction algorithm. This algorithm uses 3D principal curvature images, extracted from 3D data, for robust detection of pertinent facial features such as the nose, and the eyes (figure 1). Exploiting the results of a skin color segmentation technique provides additional evidence. The 3D position of detected features is subsequently used to calculate a 3D transformation that aligns the head with a predefined up-right pose (figure 2).

The objective of the illumination compensation module is to reduce the effect of environmental illumination on the recorded image surface. The algorithm is based on an inverse shape-from-shading approach that uses 3D and color images to calculate the direction and intensity of the light source. A single diffuse light source and Lambertian surface reflectance are assumed. The effect of the illumination



Fig. 2. 3D pose compensation results. The original images are in the first column, artificially rotated images in second column, compensated images in third column.

is subsequently subtracted from the original color image. It is well known that face recognition algorithms do not cope well with changes in illumination, therefore the developed technique leads to a significant improvement to the recognition rate under different conditions.

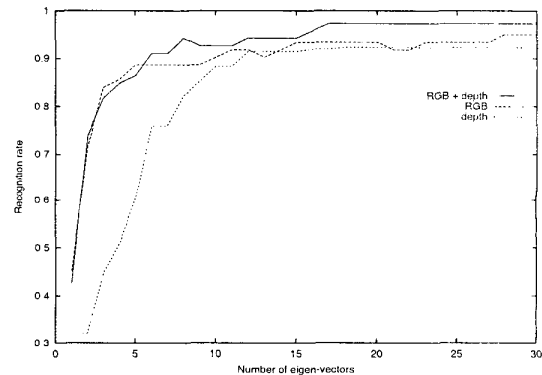


Fig. 3. Plot of correct recognition rate versus the number of eigen-vectors used. Results using depth, colour and combined depth and colour information are shown.

Two face recognition techniques were developed that exploit both depth and colour information. The eigen-face technique is extended and adapted to handle color images and depth maps. Also a face recognition algorithm based on the elastic graph matching approach has been developed. Both algorithms have demonstrated similar performance when applied to rectified versions of input images. The incorporation of depth information resulted in the improvement of

the recognition rate up to 5% (figure 3).

3. GESTURE RECOGNITION SYSTEM

The industrial interest in gesture recognition for Human-Computer interfaces comes from the vast number of potential applications. In desktop applications it can stand beside or replace the "classical" input devices. A large potential interest comes from the possibility to develop advanced interfaces made of virtual objects. These objects can be images on a computer screen. The user can "manipulate" the objects by moving his hand and performing actions like "grasping" and "releasing". The computer uses gesture recognition to reproduce the user actions on the virtual object and the result of the operation is shown in the graphical interface so that the user can have a feedback. Applications are in simulation, robot teaching, graphical interface control, device control and in Virtual Reality [5].

In this paper we present several novel 3D-image analysis algorithms for the segmentation and modeling of the human hand using depth and color information. The results obtained are applied to two application scenarios. The first application involves the recognition of static gestures, such as gestures corresponding to the numbers 0-9. The second application is about the control of a 3D object rendered on a computer screen by means of dynamic manipulative gestures such as translation, scaling, activation etc.

The developed gesture recognition system consists of various steps that proceed from a coarse hand modeling to a detailed 3D representation.

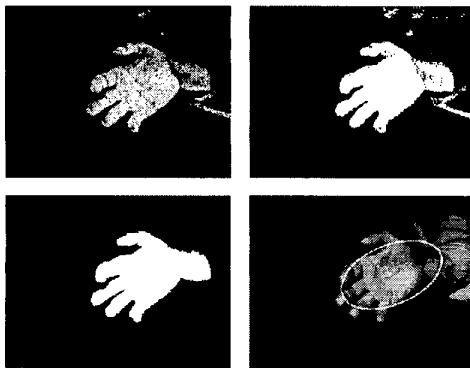


Fig. 4. Skin color segmentation steps. Estimation of $p(\text{skin}|x, y)$, thresholding, post-processing, extraction of geometric information.

The results of a skin color segmentation algorithm (figure 4) are combined with depth-based foreground-background segmentation to obtain an initial segmentation of the user hand(s).

A-priori knowledge about the geometry of the hand is subsequently exploited to achieve a refined segmentation of the hand into its composing surfaces such as the palm, the arm, and the forearm. Intensity or colour images may be used to achieve a segmentation of the hand e.g. by analysing the hand silhouette after background segmentation. However this approach can not cope well with rotations and/or occlusions. Depth information on the other hand may be used to achieve a 3D segmentation of the hand under arbitrary rotations and even occlusions.

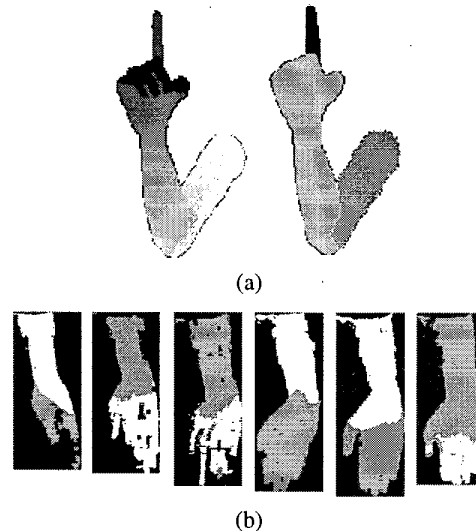


Fig. 5. 3D hand segmentation results on (a) synthetic data, and (b) real data.

This is achieved by assuming a 3D-blob model of the hand. Each hand sub-surface is represented by a 3D ellipsoid. Hand segmentation then reduces to the estimation of the centre and 3D orientation of individual 3D blobs. A probabilistic approach has been used, where each 3D point is assumed to be a sample of a distribution modelled as a mixture of Gaussians. The parameters corresponding to every Gaussian component are estimated by applying a restricted version of the Expectation Maximisation algorithm. A-priori knowledge regarding the geometry of the human hand is incorporated in the model to achieve fast and accurate convergence of the algorithm (figure 5).

The 3D principal curvatures are subsequently calculated from the input depth images. The curvature images are analysed in order to locate feature points that may be associated with fingertips. Spatial constraints posed by the structure of the hand are exploited to reject erroneous measurements. Also the center and orientation of the palm is available from the above hand segmentation module. Based on these measurements (fingertip position and orientation and palm cen-

ter and orientation) a coarse modelling of the hand may be achieved. This is sufficient for static and dynamic gesture recognition and has a real-time performance.

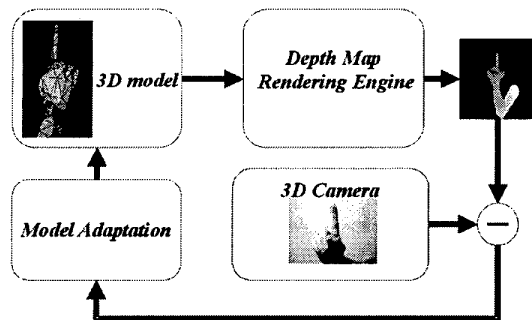


Fig. 6. Architecture of 3D hand modelling algorithm.

Finer 3D modelling of the hand is achieved by adapting an articulated parametric 3D-hand model on the depth images using an analysis-by-synthesis approach. In order to estimate the animation parameters of the hand, a generic VRML hand model, which has been used for BAP synthesis experiments in the context of MPEG-4 Synthetic-Natural Hybrid Coding (SNHC) Group is used. This model consists of 16 separate meshes, three for each phalanx of each finger, and one describing the palm. The motion of the hand is controlled by means of three flexion parameters for each finger, each corresponding to a finger joint, and one parameter, describing the pivot of the thumb, in the palm plane.

An initial model configuration is obtained by using the coarse hand model estimated above. Then an iterative optimisation algorithm is applied that minimises the distance of the synthesised depth map from the original depth map provided as input. Several distance measures have been investigated including chamfer and Euclidean. A block diagram of illustrating the above technique is shown in figure 6.

After an initial adaptation of the 3D model on the first image frame is obtained the algorithm may be used for tracking the hand from frame to frame. Experimental results demonstrate robustness of the developed algorithms to occlusion, illumination changes and erroneous depth measurements.

4. REFERENCES

- [1] F. Forster, M. Lang, and B. Radic, "Real-time 3d and color camera," in *Proc. ICAV3D 2001*, Mykonos, Greece, May 2001.
- [2] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–740, May 1995.

- [3] G. Gordon, "Face recognition based on depth maps and surface curvature," in *Proc. SPIE: Geometric Methods in Computer Vision*, 1991, vol. 1570, pp. 234–247.
- [4] J. Atick, P. Griffin, and N. Redlich, "Statistical approach to shape from shading," *Neural Computation*, vol. 8, pp. 1321–1340, 1996.
- [5] R. Sharma, V. I. Pavlovic, and T. S. Huang, "Toward multimodal humancomputer interface," *Proc. IEEE*, vol. 86, no. 5, pp. 853–869, May 1998.