The HISCORE gesture recognition application:
A gesture recognition system based on an active stereo sensor

V. Giagourta, N. Grammalidis, S. Malasiotis, F. Nebia-colombo, M. Strintzis
Informatics and Telematics Institute (ITI), 1st km Thermi-Panorama Road,
P.O.Box 361, GR-57001 Thermi, Thessaloniki, Greece
Email: giagourta@panorama.ee.auth.gr, ngramm@iti.gr, malasiot@iti.gr

ABSTRACT
In recent years there has been increasing interest in gesture-based human-computer interaction in order to develop more natural and efficient human-computer interfaces. The paper presents several novel 3D image analysis algorithms, applied towards the segmentation and modelling of hands. These are subsequently used to build 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 gesture recognition techniques that use 2D images only.

1. INTRODUCTION
Using machine vision to recognise hand gestures is a key feature in the pursuit of designing machines capable of interacting intelligently and effortlessly in a human-inhabited environment. 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. Several applications exist in the areas of simulation, robot teaching, graphical interface control, device control and Virtual Reality.

An overview of this challenging area can be found in [1]. Depending on the "environment" where the hand movement takes place (local or global), there have been several examples of systems, with varying degrees of complexity. In the case of local environments, a mouse substitution system has been described by Rehg [2]. Also a system where different hand configurations lead to mouse click equivalents is described by Toshiba [3]. In the global case, the user may move freely in a room, which may be useful when gestures are used for remote control of devices [4,5] or interactions with a video projector [6,7,8]. Also, several physiologic factors must be taken into account when designing such a system: for example, limits on the allowable latency time in order to make the interaction effective [9].

In this paper we present a system based on a novel 3D color camera using the Color Coded Light Approach (CCLA). Several novel 3D-image analysis algorithms for the segmentation and modeling of the human hand using depth and color information constitute the building blocks of the 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 concerned with the control of a 3D object rendered on a computer screen by means of dynamic manipulative gestures for actions such as translation, scaling, activation etc. The experimental results demonstrate that improved efficiency and robustness may be achieved through the use of 3D information.

2. DESCRIPTION OF THE SYSTEM
A classical image-processing pipeline is used for hand gesture recognition. The first step is image acquisition. The input to the system is a color image together with an associated depth map taken by the camera subsystem. The special camera performing the 3D as well color images acquisition 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.

The developed gesture recognition system consists of various steps that proceed from a coarse hand modelling to a detailed 3D representation. The problem is commonly investigated using the following steps:

**Hand segmentation:** the aim of this module is to divide the hand image from the background image. This task can come after a preprocessing step including illumination compensation and noise filtering. Hand segmentation is different from common image segmentation aimed at dividing an image into homogeneous regions on the basis of some suitable mathematical or logical criteria. In hand segmentation the objective is to identify the hand in the background i.e. there is a semantic criterion not a mathematical one. This can imply the need of several processing steps involving some recognition tasks.

The results of a skin color segmentation algorithm are combined with depth-based foreground-background segmentation to obtain an initial segmentation of the user hand(s). Some results are shown in Figure 1.

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.

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. As an example, we present here some palm-forearm segmentation results in Figure 2.

**Feature extraction:** this module is responsible for extracting features that are measured and used to model the hand. The initial phase is mainly focused to the development of a fingertip detection module and a gesture classification module based on measured features. These modules should start from the image segmentation results to detect fingertip position and to classify static gesture. After analyzing several alternatives a curvature analysis algorithm has been selected for detecting fingertips. The 3D principal curvatures are subsequently calculated from the input depth images. The curvature

Figure 1: Skin color segmentation results

Figure 2: Palm-forearm segmentation results
images are analysed in order to locate feature points that may be associated with fingertips. The algorithm uses the hand contour curvature in color images and can be extended to 3D images by using the “occluding boundaries”. After hand location and segmentation, the curvature of the hand contour is computed and fingertips are detected on the basis of their high (and convex) curvature value, see Figure 3. The same approach allows to extract the points at the attachment of fingers to hand palm. They have high concave curvature value. In the 3D image case, the 2D projection of occluding boundaries can be used to improve the algorithm. In this case depth information could be used too, but the computation of 3D surface curvature seems to be too difficult to be computed with reliable results with low processing effort. The classification module should start from fingertip position, as well as from all available data, to recognize the static hand gesture. A neural network approach has been considered for this task.

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 center 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.

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 palmar 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 described 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. For an overview of the proposed BAP estimation system, see Figure 4.

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. Here we present a demonstrative example: the depth image that was used as input is shown in the first image of Figure 5. We assume that only three degrees of freedom were allowed, corresponding to BAP’s describing 3 flexions of the index, and that the 3D position of the joint between the index and the palm has been accurately estimated. The resulting 3D mesh after 35 iterations of the optimisation method is presented in the second image of Figure 5.

Having adapted the 3D model successfully, it is now an easy task to map static hand con-
Fig. 5: 3D mesh adaptation

configurations and dynamic hand gestures to trajectories. Of course, in the case of trajectories, time warping techniques, analysis into sequences of static landmark configurations or other techniques will be used in future work.

Hand tracking. tracking of hand position and orientation is done if it is required by application. Still gesture recognition usually does not employ hand tracking. A two level of tracking is planned. First tracking level is applied to hand and fingers positions. Tracking fingers positions can be of help in optimising segmentation module reducing computation efforts from whole image to some regions of interest. Second tracking level is applied to extracted features.

Gesture recognition: recognition problem implicitly answers the question about what kinds of gestures are suitable. Recognition approaches use hand model information to match with pre-acquired single gesture models.

Static gestures are recognized using single static images or image sequences. In the first case segmentation, feature extraction and hand modeling and classification are involved. In the case of image sequences static gestures have to be hold for a while and system loops on recognition/movement detection.

3. CONCLUSION

A compact and affordable real-world gesture recognition system is being developed, which uses an analysis by synthesis approach internally, and which effectively exploits the 3D data that is available from novel structured light camera. The final results seem very promising, and justify our conjecture that depth information can significantly aid to simplifying the implementation and enhancing the performance of a gesture recognition system as a whole.

The development of HISCORE gesture recognition system is still at initial phase. The design of system architecture and modules is research issue, so it could be modified as result of the ongoing studies. The prospective applications of HISCORE gesture recognition system ranges from gesture recognition SW for 3D Color images (Virtual reality, task training, medicine, CAD, sign language recognition, video games,) to Gesture tracking software for 3D Color images (Telerobotics, medicine, CAD, graphical character animation,). Finally the design of 'Six dimension mouse kit' can be considered, to be used in CAD, video games or graphical animation contexts.

4. REFERENCES


