Real Time Hand Gesture Recognition Including Hand Segmentation and Tracking

Thomas Coogan, George Awad, Junwei Han, and Alistair Sutherland

Dublin City University, Dublin 9, Ireland
{tcoogan, gawad, jhan, alistair}@computing.dcu.ie

Abstract. In this paper we present a system that performs automatic gesture recognition. The system consists of two main components: (i) A unified technique for segmentation and tracking of face and hands using a skin detection algorithm along with handling occlusion between skin objects to keep track of the status of the occluded parts. This is realized by combining 3 useful features, namely, color, motion and position. (ii) A static and dynamic gesture recognition system. Static gesture recognition is achieved using a robust hand shape classification, based on PCA subspaces, that is invariant to scale along with small translation and rotation transformations. Combining hand shape classification with position information and using DHMMs allows us to accomplish dynamic gesture recognition.

1 Introduction

The primary goal of any automated gesture recognition system is to create an interface that is natural for humans to operate or communicate with a computerized device. Furthermore we aim to develop our system without using data gloves or colored gloves. Such a system could be used in, virtual reality, robot manipulation or gaming. In fact gesture recognition could be used to improve the intuitiveness of any Human-Computer Interaction (HCI). We routinely use hand gestures when communicating, describing and directing during our everyday activities. Incorporating gestures with HCI could be an extremely beneficial development.

In recent years various approaches to gesture recognition have been proposed, Gupta et al [1] presented a method of performing gesture recognition by tracking the sequence of contours of the hand using localized contour sequences. Chen et al [2] developed a dynamic gesture recognition system using Hidden Markov Models (HMMs). Patwardhan et al [3] recently introduced a system based on a predictive eigentracker to track the changing appearance of a moving hand. Kadir et al [4] describe a technique to recognize sign language gestures using a set of discrete features to describe position of the hands relative to each other, position of the hands relative to other body locations, movement of the hand, shape of the hand. While some of these approaches display impressive results, many exploit controlled environments and a compromise between vocabulary size and recognition rate.

To achieve accurate gesture recognition over a large vocabulary we need to extract information about the hand shape. Accomplishing this entails detecting the hands, segmenting them, differentiating them and classifying them. This requires using skin

G. Bebis et al. (Eds.): ISVC 2006, LNCS 4291, pp. 495 – 504, 2006.

detection techniques and handling occlusion between skin objects to keep track of the status of the occluded parts. We present a unified system for segmentation and tracking of the face and hands in a gesture recognition using a single camera. Unlike much related work that uses color gloves [5], we detect skin by combining 3 useful features: color, motion and position. These features together, represent the skin color pixels that are more likely to be foreground pixels and are within a predicted position range. Also, unlike other work that avoid occlusions entirely by choice of camera angle, sign vocabulary, or by performing unnatural signs [6,7], we handle occlusion between any of the skin objects using a Kalman filter based algorithm.

Once the hand is segmented classification is required. In hand shape recognition, transformation invariance is key for successful recognition. We propose a system that is invariant to small scale, translation and shape variations. This is achieved by using a-priori knowledge to create a transformation subspace for each hand shape. Transformation subspaces are created by performing Principal Component Analysis (PCA) on images produced using computer animation. We introduce our method that enables us to train this appearance based method using computer animation images. Also presented is the incorporation of this hand shape classifier into a dynamic gesture recognition system. Position information is combined with hand shape information to construct a feature vector that is passed to a DHMM for dynamic gesture recognition.

The remainder of this paper is organized as follows: The method used to segment the face and hands is reported in section 2. The gesture recognition technique including hand shape recognition is described in section 3. In section 4 we detail some experiments and finally we offer some conclusions in section 5.

2 Segmentation and Tracking

In gesture recognition we need to segment and track three objects of interest: the face and the two hands. The skin segmentation module is responsible for segmentation of skin objects, similarly the object tracking module is responsible for matching the resulting skin blobs of the segmentation component to the previous frame blobs while keeping track of the occlusion status of the three objects. In the next sections we will explain the details of these two components.

2.1 Skin Segmentation

In order to robustly detect the skin objects, we combine three useful features: color, motion and position. Color cue is useful because the skin has a distinct color that helps to differentiate it from other colors. The motion cue is useful in discriminating foreground from background pixels. Finally, the predicted position of objects using Kalman filter helps to reduce the search space.

2.1.1 Color Information

In order to collect candidate skin pixels, we use two simple classifiers. First, a general skin model (color range) is applied on small search windows around the predicted