Visual Tracking of Hand Posture in a Robot Control Application

Fabienne Lathuilière Jean-Yves Hervé
Groupe de Recherche en Perception et Robotique
École Polytechnique, Montréal, QC H3C 3A7
tel (514) 340-4711-5783, fax (514) 340-5802,
email: [fabienne,jyh]@ai.polymtl.ca

Abstract

The properties of versatility and dexterity of the human hand have raised a growing interest both in Human Computer Interaction applications and in gesture recognition. Emphasis is laid on ease of detection and reconstruction of the hand posture as well as on real-time computation. This paper proposes a visual hand tracking system and a hand posture estimation method where the position and the orientation are recovered from interest cues on the hand. A kinematic model of the hand and a reconstruction method to recover hand posture are provided. A skeletal model of the hand is built to model the kinematic properties of the hand. A single camera provides the frame sequence of the mobile hand while color segmentation is used to detect salient features on the glove adorned hand. After the pose of the wrist is computed, the value of the finger joint angles are obtained by inverse kinematics. This method enables the successful recovery of the hand posture, opening the door to applications such as 3D input devices or powerful 6-dof control tools. Human hand skill can thus be exploited to control a robot gripper in a master-slave system taking advantage of the information provided by the human guide.

1 Introduction

Recent developments in Human Computer Interaction (HCI) have led to different approaches as regards the use of the hand as an input device. Sensor-equipped devices such as input gloves or magnetic sensors hinder the user's natural movements and furthermore lack precision. In order to avoid such cumbersome and expensive devices, a lot of work has been dedicated to visual interfaces with one or more cameras tracking the hand. A recent survey about HCI [1] highlights the requirements of an efficient visual interface device. In particular, the ease of use, the absence of tedious calibration steps, and real-time reaction of the system are of much importance, hence the current efforts to create efficient vision-based interface applications. Two main categories stand up in vision-based applications. The first class of work places the emphasis on gesture recognition, where only a finite amount of hand gestures are recognized. For instance, the system presented in [2] recognizes several American Sign Language (ASL) gestures by tracking fingertip movement. Each gesture is defined by the vector list of the fingertip motion, and recognition is achieved by look-up table list matching. [3] uses a view-based representation of the hand to learn and recognize dynamic hand gestures by statistical matching. In both cases, the palm remains in a constant location and no 3D tracking is addressed. The second approach emphasizes gesture reconstruction, possibly through the definition of some specific mapping between hand posture and actions on the device. In this case, a 3D model of the hand is needed and hand posture is generally recovered by a fitting process. Among these applications, [4] uses a Kalman filter to determine 3D hand position, and [5] proposes a tracking system of the index finger by 2D image fitting for a virtual gun interface, but both systems require some part of the operator to remain still. [6] successfully recognizes the posture of an unadorned hand by model state estimation and residual vector minimization, but needs two cameras to successfully position in space characteristic lines and points on the fingers. Finally, the system described in [7], based on color markers to designate the fingertips, leads to a precise reconstruction of the hand posture but spends much time in iterative procedures. The approach chosen in this article aims at providing enough information on an adorned hand to recognize its posture in space with a single camera and little com-

The support of NSERC (RGPIN 171210-97) for this work is gratefully acknowledged.