MARKERLESS HUMAN TRACKING FOR INDUSTRIAL ENVIRONMENTS

Mohanad Elshafie and Gary M. Bone
Department of Mechanical Engineering,
McMaster University, Hamilton, L8S 4L7, Canada,.
(E-mail: elshafm@mcmaster.ca and gary@mcmaster.ca)

ABSTRACT

A markerless multiple-camera vision-based 3D human tracking method for industrial environments is presented. The method can track humans in the vicinity of moving robots without using skin color cues or articulated human models. It is robust to self-occlusions and to partial occlusions caused by the robot. Foreground pixels corresponding to humans are found by background subtraction. A convex polyhedron enclosing the human(s) is generated online by bounding the foreground pixels in 3D space. Experimental results are included for a single person and multiple persons walking near a moving PUMA robot in a cluttered environment. Reliable tracking at 11.4 Hz is demonstrated using four cameras and a Pentium 4 PC. The tracking data may be used for online robot collision avoidance.

Index Terms— Human tracking, 3D person tracking, Robot safety, Human-Robot interaction.

1. INTRODUCTION

Robots in industry are usually separated from humans using physical barriers such as fences or sensor curtains to protect the workers. These approaches are inflexible and waste valuable factory space. To provide a more flexible approach for avoiding human-robot collisions, a fast system to track the human's 3D location in the proximity of the robot(s) is needed.

Research in the area of markerless vision-based human tracking has been an active topic for the last decade. In [1], a tracking method that depends on the human's skin color has been tested. Using skin color as a cue to find a human can be problematic, since self-occlusion of the limbs (*e.g.* a hand is hidden behind the torso) requires the system to do extensive calculations for it to be robust in finding the locations of the limbs. A solution to the self-occlusion problem is to have multiple cameras installed. This will give the system multiple views of the object as well as depth information at the cost of more computations. A system has been developed using multiple cameras to track people in a cluttered environment in [2]. The system takes the midpoints of matched segments from different camera views and projects the objects' locations onto the ground plane. Next it uses kernel estimation techniques to calculate the probability of the presence of an object on the 2D ground plane. Finding the probability of objects makes the system too slow for real time applications. This system also does not provide any information about the third dimension of the object.

More recent research approaches have focused on building articulated human models [3], [4] and [5]. The human model is fitted to data obtained from different sensors. Building a human model, where limbs are modeled as segments connected by joints, requires about 29 degrees of freedom, which means the system needs to measure the location of 29 parameters in each instant. This makes these markerless systems too complex and slow for real time applications. In addition, the human body is very diverse, so the system must be able to scale to people with different heights or even the absence of body parts. Human model methods also face a problem with occlusion of body parts due to loosely fitting clothing, e.g. a person wearing a dress.

A faster markerless human tracking system has been described in [6]. This system finds the 3D location of the hands and head using skin color segmentation. It then tracks the movements using particle filters. The system achieved a speed of 15 Hz using two cameras running at 25 Hz and an image size of 320 x 240 pixels. It suffers from the same limitations as other skin color-based approaches.

One of the common techniques used in human tracking is background subtraction. It relies on building a background model of the scene and then segmenting foreground objects by subtracting the current image from the background model. This technique has been implemented by several research groups (*e.g.* [2], [5], [7] and [8]). An accompanying problem is shadows being detected as foreground objects.

In this paper, we propose a novel method for markerless vision-based human tracking. The person/people in 3D space is/are modeled as a convex polyhedron. Our method may be implemented using two or more cameras. It is robust to self-occlusion and loose clothing. Moreover, it does not depend on skin color segmentation or fitting a human model. It is efficient enough for use in human-robot collision avoidance applications.

2. ASSUMPTIONS

Our method requires the following assumptions:

1. The robot has a distinct color from other objects in the area.
2. The human is not wearing a color similar the robot's color.

The first assumption is usually the situation or can be satisfied by painting the robot. The second assumption applies to all vision-based tracking systems.

CCECE/CCGEI May 5-7 2008 Niagara Falls. Canada
978-1-4244-1643-1/08/$25.00 © 2008 IEEE