Human Detection Using Depth Information by Kinect

Lu Xia, Chia-Chih Chen and J. K. Aggarwal
The University of Texas at Austin
Department of Electrical and Computer Engineering
{xialu|ccchen|aggarwaljk}@mail.utexas.edu

Abstract

Conventional human detection is mostly done in images taken by visible-light cameras. These methods imitate the detection process that human use. They use features based on gradients, such as histograms of oriented gradients (HOG), or extract interest points in the image, such as scale-invariant feature transform (SIFT), etc. In this paper, we present a novel human detection method using depth information taken by the Kinect for Xbox 360. We propose a model based approach, which detects humans using a 2-D head contour model and a 3-D head surface model. We propose a segmentation scheme to segment the human from his/her surroundings and extract the whole contours of the figure based on our detection point. We also explore the tracking algorithm based on our detection result. The methods are tested on our database taken by the Kinect in our lab and present superior results.

1. Introduction

Detecting human in images or videos is a challenging problem due to variations in pose, clothing, lighting conditions and complexity of the backgrounds. There has been much research in the past few years in human detection and various methods are proposed [1, 2, 6, 13]. Most of the research is based on images taken by visible-light cameras, which is a natural way to do it just as what human eyes perform. Some methods involve statistical training based on local features, e.g. gradient-based features such as HOG [1], EOH [8], and some involve extracting interest points in the image, such as scale-invariant feature transform (SIFT) [9], etc. Although lots of reports showed that these methods can provide highly accurate human detection results, RGB image based methods encounter difficulties in perceiving the shapes of the human subjects with articulated poses or when the background is cluttered. These will result in the drop of accuracy or the increase of computational cost. Depth information is an important cue when human recognize objects because the objects may not have consistent color and texture but must occupy an integrated region in space. There has been research using range image for object recognition or modeling in the past few decades [12, 14]. Range images have several advantages over 2D intensity images: range images are robust to the change in color and illumination. Also, range images are simple representations of 3D information. However, earlier range sensors were expensive and difficult to use in human environments because of lasers. Now, Microsoft has launched the Kinect, which is cheap and very easy to use. Also, it does not have the disadvantages of laser so it can be used in human environment and facilitate the research in human detection, tracking and activity analysis.

In recent years, there is a body of research on the problem of human parts detection, pose estimation and tracking from 3D data. Earlier research used stereo cameras to estimate human poses or perform human tracking [3, 4, 15]. In the past few years, a part of the research has focused on the use of time-of-flight range cameras (TOF). Many algorithms have been proposed to address the problem of pose estimation and motion capture from range images [5, 7, 11, 16]. Ganapathi et al. [5] present a filtering algorithm to track human poses using a stream of depth images captured by a TOF camera. Jain et al. [7] present a model based approach for estimating human poses by fusing depth and RGB color data. Recently, there have been several works on human/parts detection using TOF cameras. Plagemann et al. [10] use a novel interest point detector to solve the problem of detection and identifying body parts in depth images. Ikemura et al. [6] proposed a window-based human detection method using relational depth similarity features based on depth information.

In this paper, we present a novel model based method for human detection from depth images. Our method detects people using depth information obtained by Kinect in indoor environments. We detect people using a 2-stage head detection process, which includes a 2D edge detector and a 3D shape detector to utilize both the edge information and the relational depth change information in the depth image. We also propose a segmentation method to segment the figure from the background objects that attached to it and extract the overall contour of the subject accurately. The method is evaluated on a 3D dataset taken in our lab using the Kinect for Xbox 360 and achieves excellent results.