SCHOOL OF COMPUTER SCIENCE

Tiny: Tracking People using Multiple Kinects

Chi-Jui Wu

October 21, 2014

1 PROJECT

People detection and tracking are essential in personalized robotics, surveillance, interactive systems, and medical monitoring.

The current project proposes an algorithm for tracking people using multiple Kinects. The proposed algorithm would reliably track targets' appearance, position, and movement in realtime. [Tracking people using other technologies].

1.1 CONTRIBUTIONS

1.1.1 PROBLEM OF OCCLUSION

Real world environments are dynamic. Occlusion occurs when the tracked target is hidden by other objects in the field of view of one or more cameras. It increases the difficulty of people detection and tracking. There are two types of occlusions. Static occlusion refers to stationary objects, and it can be accounted by placing cameras at locations that encapsulate the entire world view. Dynamic occlusion arises from the interactions of targets in the environment, such as two person walking past each other. The project aims to resolve both types of occlusion using multiple Kinects.

The problem of occlusion is illustrated in Figure 1.1. The first camera

1.2 KINECT

Kinect is a low-cost sensor for motion capturing and tracking. The sensor provides infrared, RGB, and depth streams at high frame rates.

2 PREVIOUS WORK

Tracking moving targets is challenging. Because... The section will review the state of the art techniques in people tracking.

2.1 RGB-D

2.1.1 FEATURE-BASED TRACKING

Gavrila et al extract depth features from images and match them against a hierarchy of templates to detect pedestrians in realtime [0]. Mikolajczyk1 et al detect the human body as a collection of local feature leared from training examples using Adaptive Boosting, a machine learning technique that gradually improves the model based on the errors of the previous models [3]. Lowe presents the scale-invariant feature transform (SIFT) technique to extract highly distinctive interest points from images that can be used to identify objects from different views [6].

2.1.2 COLOR HISTOGRAM

In addition to feature detection, histogram based detection techniques are also used frequently. A color histogram based people tracking system is able to keep track of moving people after occlusion and changing illumination [4].

2.1.3 HOG: HISTOGRAM OF ORIENTED GRADIENTS

Moreover, Dalal and Bill Triggs show experimentally that Histograms of Oriented Gradient (HOG) descriptors outperform existing human detectors [1]. HOG is one of the most widely used techniques for people detection [7, 8], and subsequent work show that it can not only reliably track similar targets but also increase the performance of tracking algorithm [2, 9].

The method evalutes local histogram of image gradient orientations. An image gradient is a directional change in the color intensity in a fixed-size detection window. Each dection window is divided into cells. A histogram of gradient directions is created for the the pixels within the cell. Local features can be characterized by the distribution of the local gradients. The local histograms of a group of cells, or a block, are normalized. That is, differences among the local histograms of a block are small. The normalization procedure ensures the histograms are invariant to changes in illumination and contrast. HOG is represented as a collection of normalized descriptor blocks, each describing the local features of a region in an image. The descriptor is then used to train a linear Support Vector Machine(SVM) to do person and non-person classification. To detect multiple people, the detection window is scrolled over the image at different scales and the corresponding descriptors are computed to train the SVM.

2.1.4 HOD: HISTOGRAM OF ORIENTED DEPTHS

Inspired by HOG, Spinello and Arras develop a person detector called Histogram of Oriented Depths (HOD) for dense depth data [5]. Similarly to HOG, HOD divides a fixed-size detection window into cells. Instead of encoding changes in color intensity, it computes the local oriented depth gradients and creates a histogram for each cell. The histograms of a block, or four cells, are normalized. HOG captures local 2D shape, whereas HOD captures 3D shape. The descriptors are also taken for training a linear SVM.

Instead of encoding the local directional changes of color intensity in an image like HOG does, HOD encodes the local directional changes of depth.

2.1.5 Combo-HOD: RGB-D people detector

Spinello and Arras propose a novel person detector that probablistically combines HOG and HOD [5]. The method does not rely background segmentation nor a ground plane assumption. However, it scans the entire RGB frame, hence requires a GPU implementation to be executed in realtime.

2.1.6 Depth-based human model

2.1.7 Sub-clustering

Theme: Color normalization is applied in every algorithm which relies on differences in RGB data.

2.2 Multiple Kinects

2.2.1 Calibration

2.2.2 Merging the FOVs

3 Current Approach

[0] Vision-Based Pedestrian Detection: The PROTECTOR System [1] Histograms of Oriented Gradients for Human Detection [2] Exploring Context Information for Inter-Camera Multiple Target Tracking [3] Human Detection Based on a Probabilistic Assembly of Robust Part Detectors [4] A COLOR HISTOGRAM BASED PEOPLE TRACKING SYSTEM [5] People Detection in RGB-D Data [6] Distinctive Image Features from Scale-Invariant Keypoints

[7] Pedestrian detection: A benchmark [8] Monocular pedestrian detection: Survey and experiments [9] People tracking in RGB-D data with on-line boosted target models

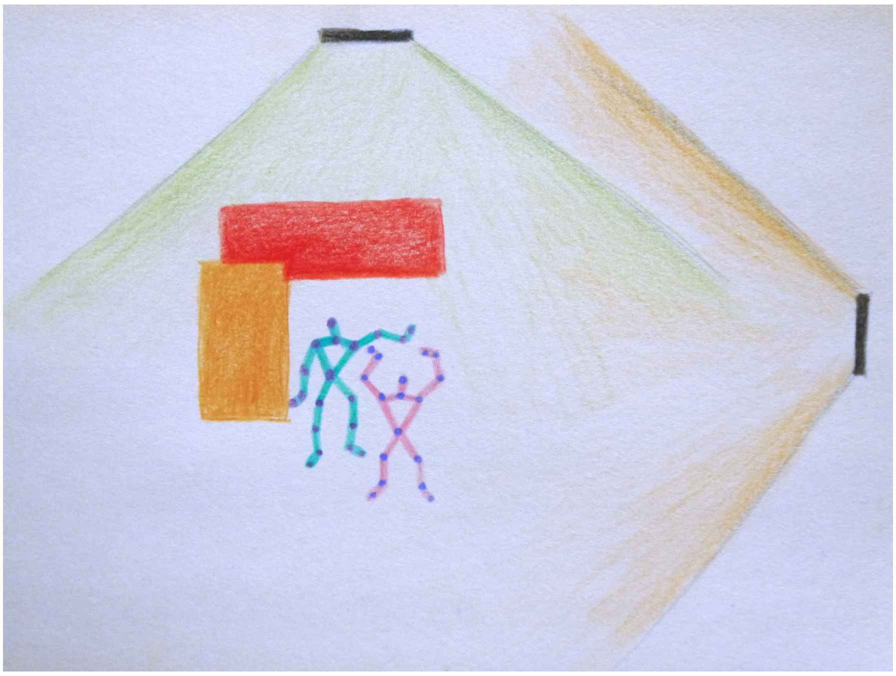

Figure 1.1: The problem of occlusion