Real Time Crowd Counting with Human Detection and Human Tracking

Xinjian Zhang and Liqing Zhang

Key Laboratory of Shanghai Education Commission for Intelligent Interaction and Cognitive Engineering,
Department of Computer Science and Engineering,
Shanghai Jiao Tong University, China
{zha,lqzhang}@sjtu.edu.cn

Abstract. Real-time crowd counting is of many potential applications, such as surveillance, crowd flow control in subway. In this paper, we propose a fast and novel method for estimating the number of people in crowded surveillance scenes. This method is able to count people in real time and is robust to changes of illumination and background. The combined rectangle features and cascade of boosted classifier are employed to train a multi-scale head-shoulder detector. The detector can detect human in every frame with a high accuracy. Then human tracking is used to track the detected people and remove duplicates in successive frames. Experiments on a real-world video show the proposed method can give an accurate estimation in real time.

Keywords: Crowd Counting, Human Detection, Human Tracking.

1 Introduction

Real-time crowd counting in videos becomes more and more important for public area monitoring for the purpose of safety and security. The goal of crowd counting is to estimate the number of people passing through a given line or a given area. It has many valuable real-world applications, such as controlling the number of people in the venues, estimating the people flow in the subway station, counting people entering and exiting. There are still many challenges to be solved in this task. First, in crowded scenes, the occlusion between people is serious. Second, the resolution of video in surveillance camera is relatively low, detailed information is lost. In real-word places, such as in subway stations and libraries, we find that most monitors are above the front of people's heads. This is because at this position, monitors can capture faces, dresses and other characteristics of pedestrians passing through. So in our scenario, we assume that the cameras are installed at a high place, facing the crowd flow direction.

Most previous approaches can be divided into two categories. The first type is based on the counting-by-regression framework, where extracted features are directly regressed to the number of people. The second type is mainly based on multi-target human detection.

C.K. Loo et al. (Eds.): ICONIP 2014, Part III, LNCS 8836, pp. 1–8, 2014.

The counting-by-regression methods extract low level features, such as the foreground pixels, HOG features, and such local features are transformed to the number of people using regression. Cong et al. [1] used 1-D flow velocity estimation to extract dynamic mosaic. Then they did regression between the features of dynamic mosaic (pixels number, edge pixels number, width of the mosaic) and number of people. Ma and Chan [2] used local HOG features as the low level feature and did regression between the meaningful local descriptors and the number of the people. These methods only used low level features, so most of them can run in real time. But these systems can't be large-scale deployed because extracting these low level features is highly dependent on the prior knowledge of the background and the position of cameras. When the scene changes, these low level features will change a lot and the pre-trained regression function will fail to work.

Multi-target human detection can be used to count people in crowd scenes. Lin et al. [3] used Harr wavelet transform to extract the area with head-like contour, then used Support Vector Machine to determine whether these areas are heads or not. Their method can count people in a single image. Li et al [4] used a foreground segmentation algorithm and a HOG based head-shoulder detection algorithm to estimate the number of people in images. But these methods can't work in real time; the heavy computational cost limits their applications to real-world problems.

In recent years, there are great progresses in object detection. It is possible to implement robust pedestrian detection in real time. In this paper, we propose a novel real-time crowd counting method. Fig. 1 shows the block diagram of the whole system. We use combined rectangle features and cascade of boosted classifier to train a head-shoulder detector. The detector can detect human in every frame with a high accuracy and then human tracking is used to track the detected people. The number of detected pedestrians is equal to the number of the tracks.

The rest of this paper is organized as follows: Section 2 introduces the training process of the head-shoulder detector; Section 3 proposes how to remove duplicate detected pedestrians by human tracking, followed by experimental results and discussions in Section 4. Finally conclusions are given in section 5.

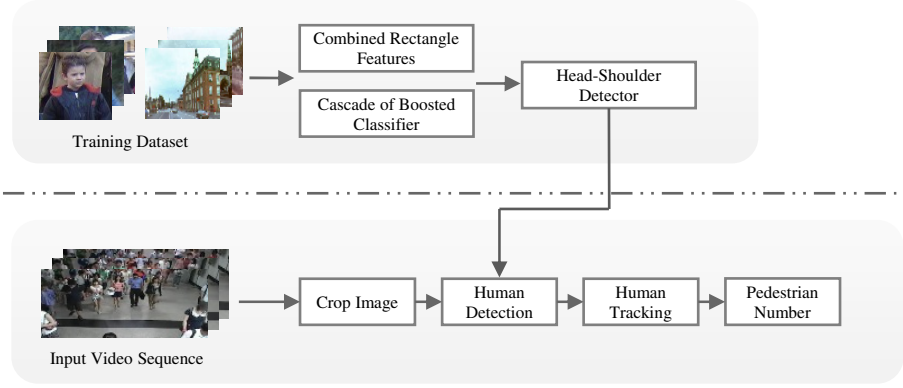

Fig. 1. Framework of the Proposed Crowd Counting System