Top-view People Counting in Public Transportation using Kinect

Filip MALAWSKI

AGH University of Science and Technology
Faculty of Computer Science, Electronics and Telecommunication
Department of Computer Science
al. Mickiewicza 30, 30-059 Krakow, Poland
e-mail: fmal@agh.edu.pl

Abstract. This article describes a method for people counting in public transportation. In this particular scenario, various body poses corresponding to holding handrails must be accounted for. Kinect sensor mounted vertically has been employed to acquire a database of images of 1-5 persons and an algorithm based on maxima detection and head candidates filtering has been devised for robust people counting, resulting in 90% accuracy.

Keywords: people counting, Kinect, public transportation

Introduction

The problem of people counting is an important aspect of video surveillance. Possible applications include, among others, security on mass events, analysis of pedestrian traffic, tourist flow estimation, analysis of marketing efficiency in shops or malls. One particular application, that is the focus of the paper, is people counting in public transportation. Information on how many people travel in public transportation, depending on different routes and hours, is crucial for proper allocation of buses, trams and subways. In the big cities, with hundreds of thousands, or even millions of residents, and hundreds or thousands public transportation vehicles, proper management of the transportation network poses a difficult challenge. Obtaining information on traffic patterns by manual counting is expensive, time-consuming and inconvenient. On the other hand automatic people counting devices could be installed in the vehicles and propagate the data in a real-time manner.

People counting is a well-know problem in the computer vision area and multiple solutions have been proposed so far. Although several methods employ RGB, side-view cameras [1][2], using top-view depth cameras proved to be much more robust to illumination changes, occlusions and crowded environments. In [3] authors propose people counting on top-view depth images by employing a water filling method, in [4] a new feature descriptor is devised for detection of human heads. In [5] and [6] methods based on fusion of depth and color information are explored.

Although presented methods have very high efficiency (over 99%), they do not consider a scenario, in which people may be holding handrails, which is a typical situation in the public transportation. It has a considerable impact on the depth map and therefore on the efficiency of the people counting algorithms. In this paper a novel method is presented, based on top-view depth-camera images, which is robust to different arrangements of people in public transportation. The method employs local maxima detection with subsequent filtering of non-head objects, resulting in 90% accuracy.

Methods

The study was conducted in two stages. First a database of top-view depth images with varying number of people in different body poses has been acquired. Secondly a method has been devised for people counting and verified on the database.

Database

Kinect sensor has been installed for top-view image acquisition, approximately 2,60m from the ground, therefore the images have been taken about 80-100cm from the subjects heads. The database contains 50 images, with 1-5 persons in each image - 10 images per each number of people. The size of the images is 640x480 pixels. In half of the images there is at least one person with body pose of holding a handrail. Data acquired from the Kinect sensor has been calibrated to record images in such a range that only head, shoulders and arms of the subjects are visible. Everything that is further than that is not relevant in subsequent processing. A sample image from the database, containing 5 persons, is presented in Fig. 1.

Figure 1: Sample top-view depth image of 5 people

People counting

Proposed method is based on a parameter of expected size of human head. Given a setup with Kinect sensor installed on a known height it is possible to measure a minimal expected area that will be occupied by a single head. Once a local maxima is found it can be assumed that no part of other head may be present in the neighboring area of the selected range. In the study, based on series of experiments, the parameter of
minimal distance between two local maxima was set to 60 pixels.

The method of people counting consists of 3 steps: 1) local maxima candidates are found in a block-processing manner 2) local maxima are selected which are potential head candidates 3) the candidates are filtered to remove non-head objects such as stretched arms.

In the first step pixels with highest values (closest to the sensor) are found in square blocks with side size equal to the minimal allowed distance between maxima. Blocks of this size guarantee that all potential maxima will be found. In the next step the potential maxima are verified. The area around the potential maxima, with the range of the minimal allowed distance, is scanned in order to check if this is in fact the pixel with the highest value in its neighborhood. Once a maxima is verified, the area around it is marked as occupied by a head, therefore other maxima areas may not intersect with this one.

Selected maxima are the head candidates. Non-head maxima, particularly stretched arms, have to be filtered out in the last step. Heads are verified by scanning the area around the maxima and computing the ratio of proper pixels to improper pixels. Proper pixels are the ones with value similar to the maxima, while improper ones have distinctly different value - mostly zero value, indicating a pixel out of the range of interest. The ratio distinguishes the heads well, as maxima corresponding to stretched arms and hands have usually much more improper pixels around them. The ratio threshold providing best classification of head and non-head maxima has been selected based on the training dataset, by iterative search with adaptive step.

Results

Proposed algorithm has been verified on the acquired database. Table 1. presents the results grouped according to the different number of people in the image. Two cases have been compared - in the first one the final filtering step is omitted and all the head candidates are counted. In the second case, non-head maxima are removed. Significant difference in the results can be observed for both cases, indicating that in this scenario it is crucial to handle the non-head maxima, as they can have great influence on the efficiency of people counting.

Conclusions

In this paper a robust method of people counting in public transportation have been presented, with overall efficiency of 90%. The method has been proved to handle well the problem of different body poses, such as stretched arms, which are typical in this scenario.

Table 1: Accuracy of the proposed method, comparison of cases with and without final non-head filtering

<table>
<thead>
<tr>
<th># people in the image</th>
<th>accuracy [%] without filtering</th>
<th>accuracy [%] with filtering</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td>40</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>60</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>20</td>
<td>60</td>
</tr>
<tr>
<td>average</td>
<td>46</td>
<td>90</td>
</tr>
</tbody>
</table>

BIBLIOGRAPHY