Подсчет людей, используя данные, собранные с помощью сенсора Microsoft Kinect*

Я.А. Валуйская
yanavaluyskaya@mail.ru
Факультет вычислительной математики и кибернетики
Московский государственный университет им.М.В.Ломоносова

Проблема подсчета людей привлекает внимание многих исследователей из-за ее высокой практической актуальности и применимости в системах видеонаблюдения. Идея основного метода, предложенная в [6], основана на естественном предположении, что голова находится ближе к датчику Kinect Microsoft, чем плечи, когда датчик направлен вертикально вниз. В этой статье была предложена модификация базового метода: сегментация глубины, которая позволяет снизить ограничения основного метода на входные данные. Экспериментальная оценка предлагаемого метода показывает эффективность на данных, полученных под углом в плотном потоке, когда возникает проблема заслонения.

Ключевые слова: *подсчет людей, глубина карта, сенсор Kinect Microsoft, алгоритм заполнения водой, сегментация глубины*

People counting using data gathered via Microsoft Kinect sensor*

Iana Valuiskaia
Faculty of Computational Mathematics and Cybernetics
Lomonosov Moscow State University, Moscow, Russia

People counting captures the attention of many researches, because of its high practical relevance and applicability in the video surveillance systems. The idea of a basic method, proposed in [6], is based on the natural assumption that a head is closer to the Microsoft Kinect sensor than shoulders when sensor is directed vertically downward. In this paper, the modification of a basic method was proposed: depth segmentation, which allows to ease restrictions of the basic method on the input data. Experimental evaluation of the proposed method shows its efficiency on data shot at an angle in heavy traffic, when occlusion problem is raised.

Keywords: *People counting, depth map, Microsoft Kinect sensor, Water Filling algorithm, depth segmentation*

1. Introduction

People counting plays an important role in video surveillance systems. This area is in great commercial interest because of the high practical relevance and applicability of the results. Areas of application are very diverse and extensive. These systems are used on the entries and exits of the buildings, in the hallways and rooms, where a large crowd of people may be critical or dangerous.

Areas of application:

— Attendance monitoring
— Evaluation of tourists flow
— Market researches
— Security systems

Information about the number of people in conjunction with other information (for example, sales) allows to evaluate the effectiveness of promotions in stores. In addition, this information can be used to control the number of open cash desks and to plan stuff breaks.

In security systems, which are used in public places (like airports, train stations, department stores, stadiums) people counting also is of a great importance. People counting can prevent dangerous situation or help if an emergency has occurred. For example, people counting can be helpful in finding groups of people of abnormal size (this requires the collection of statistics on attendance and group behavior for quite a long period of time). In addition, in the case of evacuation knowledge about the number of people entered and left from the building is vital for the efficient rescue operation.

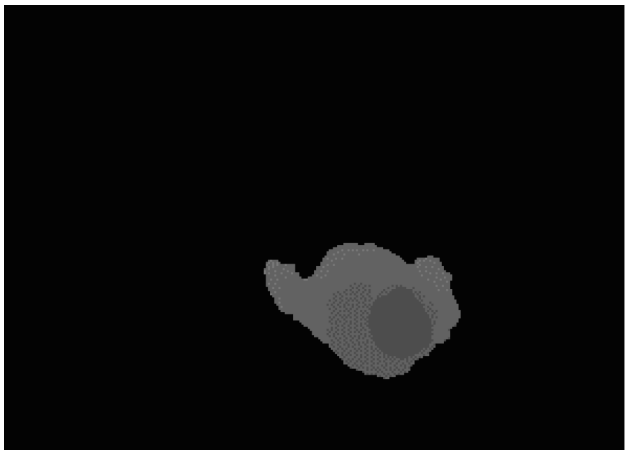

Figure 1: Data received from vertical Microsoft Kinect sensor

Работа выполнена и опубликована при финансовой поддержке РФФИ, гранты 14-01-00849, 15-07-20347

Different types of input data are used in people counting problem: data from temperature sensors, from infrared lasers, from rgb-cameras. Systems, based on video data, are the easiest to install and maintain. They create less problems to stuff and visitors, as opposed to, for example, mechanical turnstiles.
This paper describes a method that uses data, obtained from Microsoft Kinect sensor. Sensor is placed above people heads and directed either vertically downwards, as shown in Figure 1, or at angle to the vertical, Figure 2. When data is shot at an angle, occlusion problem is raised in heavy traffic.

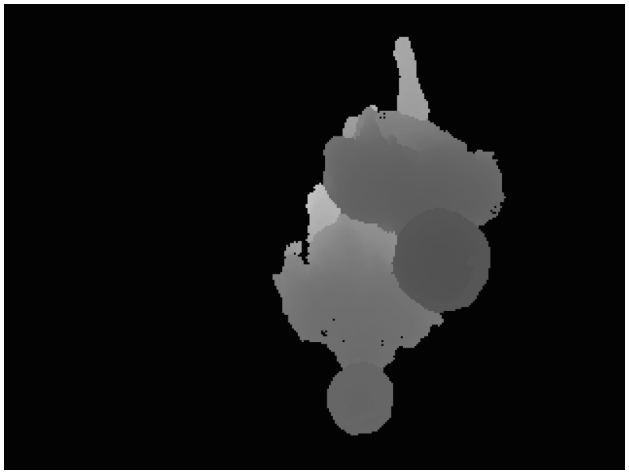

Figure 2: Data shot at an angle

Example of such system is shown in Figure 3. White horizontal lines mark region of interest. Depending on the direction of people movement, the number of people entered the ROI (IN) and the number of people left the ROI (OUT) is counted. People who haven't yet left the ROI are considered in counter STAY.

2. Review of existing methods

The main approaches for solving people counting problem are:

— Counting by people detection
— Counting by regression

In the first approach people detection problem is solved, for example, detection of peoples' heads. Then detected people are tracked by calculating movement trajectory. This approach is used in [1,4,6]. In the paper [1] authors used block-wise background subtraction and then k-means clustering is used to enable the segmentation of single persons in the scene. For tracking, they used "greedy algorithm". In the paper [4] authors extracted blobs and found peaks value for each one. Then Extended Kalman Filter is used. In the paper [6] head detection problem is reduced to finding suitable local minimums: Water Filling Algorithm is used. This approach demonstrates high accuracy when people flow is not too crowded. RGB-data and depth information are used as input.

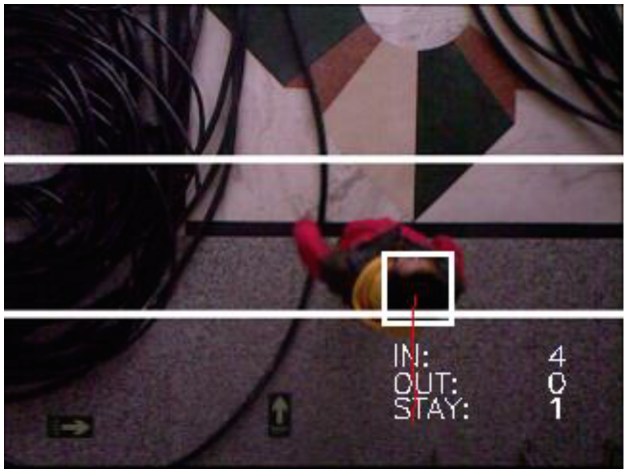

Figure 3: The result of people counting system

In the second approach some features of an input frame are extracted (for example, area, perimeter, HOG, etc). Sometimes, an input frame is segmented into several parts, for example, flows of people who go into different directions. In the paper [2] a set of simple holistic features is extracted from each segmented region, and the correspondence between features and the number of people per segment is learned with Gaussian Process regression. In the paper [5] the number of people is estimated in a set of overlapping sliding windows, using a regression function that maps from local features to a count. Then integer programming method is used.
This approach shows high results in crowded environment when people tracking is not possible. The major drawback of this approach is a necessity in a big amount of training data.
In addition, in paper [7] authors proposed a hybrid method. They use counting by detection if density of a flow is low. Otherwise, they analyze optical flow: segmentation by DBSCAN [3] algorithm is used. Accuracy of proposed method is 55% when evaluated on data with high density of a flow.

3. Proposed method

In this paper the method proposed in [6] is reviewed and modified. This basic method showed the best results among all the reviewed methods of first approach. In addition, it is capable of handling with tight groups of people (more than 3 people in a group). However, a major disadvantage of the basic method is a strict limitation on the input data: the sensor should be directed vertically downwards. If data is shot at an angle, occlusion problem appears. In this section, a modification of the basic method is proposed. It will ease the restrictions on the input data.

3.1 Basic method

The brief scheme of the basic method is shown in Figure 4.

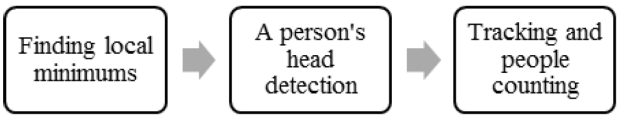

Figure 4: The brief scheme of the basic method

The input to the algorithm is a sequence of depth maps , obtained by Microsoft Kinect sensor. Depth map is an image, which contains information about the distance between objects and sensor's position. Vertical Microsoft Kinect sensor solves the occlusion problem naturally, that's way authors of the basic method use the fact that a person's head is a closest object to the sensor compared to some of it neighborhood. Thus, the problem of detecting peoples' heads is reduced to finding suitable local minimums of depth map.

3.1.1 Finding local minimums of the depth map

In order to find suitable local minimums of the depth map in [6] it was proposed to use Water Filling algorithm. Function

$$f(x,y) = \begin{cases} I_j(x,y), (x,y) & \in \text{Fore ground} \\ 0, (x,y) & \in \text{Back ground} \end{cases}.$$

Additional function $g(x,y)$ "measures" function $f(x,y)$. It can be used to infer $f(x,y)$. The definition is below.

Definition: $g(x,y)$ is a measure function of $f(x,y)$ if and only if, $\exists \quad \varepsilon > 0, \forall (x_1,y_1),(x_2,y_2)$, such as

$$\|(x_1-x_2)^2+(y_1-y_2)^2\| < \varepsilon \quad \text{and}$$
$$\text{if} \quad f(x_1,y_1) \leqslant f(x_2,y_2) \quad \text{then}$$
$$f(x_1,y_1)+g(x_1,y_1) \leqslant f(x_2,y_2)+g(x_2,y_2)$$
$$g(x_1,y_1) \geqslant g(x_2,y_2)$$
$$g(x_1,y_1) \geqslant 0, g(x_2,y_2) \geqslant 0$$

Finding proper measure function is about simulation rain process. Function $f(x,y)$ can be visualized as a land with hills and hollows. When it is raining, a raindrop in a hill will flow directly to the hollow under force of gravity. Measure function $g(x,y)$ reflects the number of water drops in a point with coordinates (x,y).

Figure 5: 2D situation

Example is shown in Figure 5: green color is used for depth function and blue color for measure function.

Results of Water Filling algorithm is shown in Figure 6.

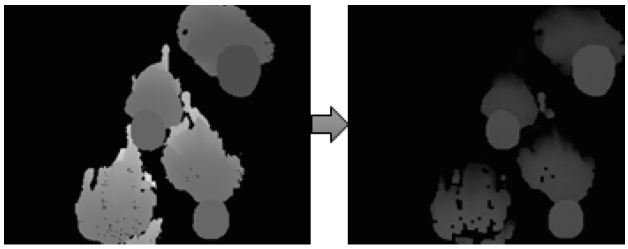

Figure 6: The result of Water Filling algorithm

3.1.2 Further analysis and tracking

At this stage, the resulting measure function g (x, y) is analyzed to find regions that correspond to people's heads. This stage consists of several steps:
1. Thresholding of function $g(x,y)$
2. Region analysis

First step allows to get rid of noise regions, as well as to separate people who walk very close to each other, so only "shoulders-head" regions will remain.
On second step separated regions $\text{Region}_i, i = 0,\ldots,m$, in function $g(x,y)$ are calculated. It is known that the region "shoulders" is located below the region "head" of each person, so

$$g(x_h,y_h) > g(x_{sh},y_{sh}), (x_h,y_h) \in \text{Head}_i, (x_{sh},y_{sh}) \in \text{Shoulders}_i$$

Secondly, function $g(x,y)$ modified as follows:

$$g(x,y) = \begin{cases} g(x,y), & \text{when} g(x,y) \geqslant M_i - \delta, (x,y) \in \text{Region}_i \\ 0, & \text{when} g(x,y) \leqslant M_i - \delta, (x,y) \in \text{Region}_i \end{cases},$$

where $M_i = \max\big(g(x,y)\big), (x,y) \in \text{Region}_i, \delta > 0$ – parameter. A size of remaining regions is analyzed: too small regions (total area $<$ 400 pixels) are removed.

After head detection it is possible to track people. Nearest Neighbor Tracking Algorithm is used for tracking people.

3.2 Modification of the basic method

If the sensor is directed at an angle to the vertical, then the first layer of people will be closer to the sensor than second. In case of occlusion problem, Water Filling Algorithm will not detect peoples' heads.

To solve this problem, in the preprocessing stage the input frame is splitted into segments according to depth. As a result, we receive groups of people located at the same distance from the sensor.

The depth segmentation consists of two steps:
1. Edge detection based on Canny edge detector
2. Identifying regions of people

First step allows to avoid situations when all foreground pixels will be allocated in one segment because of the occlusion problem.

Regions of people are identified by taking into account the edges obtained in step 1. The edges are not included in the resulting regions and used as delimiters.

Then, Water Filling Algorithm is applied not to the whole image, but separately to each segment found at the preprocessing stage. The result of depth segmentation is shown in Figure 7. Each segment is colored in its own color.

Figure 7: Depth segmentation result

4. Results

Implementation of the basic method and its modification were evaluated and compared with each other. Two datasets – Vertical and Angle - were used.

Vertical dataset consists of 3384 frames 320×240. It was received from authors of the basic method [6]. This dataset was taken by Microsoft Kinect sensor, which was directed vertically downwards.

Angle dataset consists of 1102 frames 320×240. It was taken from sensor, which was directed at an angle to the vertical. Both datasets were taken in natural environment. They consists of separated people and dense groups. In addition, people of different heights and sizes present in datasets.

Following metrics were used for evaluation and comparison of methods:

- Accuracy = TP/(TP+FP), where TP – true-positive decisions, FP – false-positive decisions
- Recall = TP/(TP+FN), where FN – false-negative decisions
- F-score = $2 * (\text{Accuracy} * \text{Recall})/(\text{Accuracy} + \text{Recall})$

Results are shown in Table 1. As the authors of basic method did not evaluate their algorithm on angle dataset, the comparison on angle dataset was carried out between the implementation of the basic method and modified algorithm.

	Vertical dataset			Angle dataset		
Method	Accuracy	Recall	F-score	Accuracy	Recall	F-score
Results of authors [6]	0.992	0.991	0.992	-	-	-
Basic method without modification	0.986	0.929	0.957	0.974	0.682	0.802
Modified method	0.986	0.929	0.957	0.985	0.779	0.87

Table 1: Experimental Evaluation

5. Conclusion

In this paper people counting method based on Water Filling Algorithm was reviewed. The analysis shows that its main disadvantage is a strict limitation on the input data: the efficiency of the basic algorithm decreases when a direction of Kinect sensor changes from vertical to angle. Modification of the basic method was proposed in order to ease this restriction.

Experimental evaluation of the modified algorithm showed its efficiency and improvement of the results compared with the implementation of the basic method.

Bibliography

[1] Antic B. et al. K-means based segmentation for real-time zenithal people counting //Image Processing (ICIP), 2009 16th IEEE International Conference on. – IEEE, 2009. – pp.2565-2568.

[2] Chan A. B., Liang Z. S. J., Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking //Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. – IEEE, 2008. – pp.1-7.

[3] Ester M. et al. A density-based algorithm for discovering clusters in large spatial databases with noise //Kdd. – 1996. – V.96. – No. 34. – pp.226-231.

[4] Hernandez D., Castrillon M., Lorenzo J. People counting with re-identification using depth cameras. – 2011.

[5] Ma Z., Chan A. B. Crossing the line: Crowd counting by integer programming with local features //Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on. – IEEE, 2013. – pp.2539-2546.

[6] Zhang X. et al. Water filling: Unsupervised people counting via vertical kinect sensor //Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. – IEEE, 2012. – pp.215-220.

[7] Harebov P., Salimzibarov R. Zenithal people counting (In Russian) //The 22nd International Conference on Computer Graphics and Vision – 2012. – pp.158-162