Application of Computer Vision Systems for Passenger Counting in Public Transport

P. Lengvenis[1], R. Simutis[1], V. Vaitkus[1], R. Maskeliunas[1]

[1]Department of Process Control, Kaunas University of Technology,
Studentu St. 48–327, LT–51367 Kaunas, Lithuania, phone: +370 682 40371
pauliuslengvenis@gmail.com

Abstract—**This paper presents a passengers counting system based on computer vision. System prototype were created and installed in Kaunas public city transport. Four algorithms were created to calculate passengers on public transport and their advantages and disadvantages were analyzed. Qualitative detection algorithms analysis carried out. Promising results were obtained with the Algorithm of barrier simulation for zones (ABSZ) which has low false rate and it is effective for people–counting. Counting results information can be used for public transport optimization or service quality improvement.**

Index Terms—**Image analysis, image motion analysis, object detection, subtraction techniques.**

I. INTRODUCTION

Nowadays computer vision is implemented throughout entire world ranging from security solutions [1] and ending with passenger counting in public transport. Number of reports made in recent years showed that the popularity of video based automatic passengers counting systems (APCS) is increasing [2]–[4]. Passenger counting is a relevant problem for today's public transport in the whole world. Only knowing the flow of passengers, the public transport companies are able to rationally use their resources, improve service quality and lower the cost of transport [5]. A rational schedule of transport based on the passenger flow allows companies to avoid "empty routes" and to reduce environmental pollution.

Passenger counting is a complicated task [6], [7]. Bus passengers differ in their look, physical dimensions and outfit. Every stop has a different background. Shadows and solar position have a lot of influence on signal quality. There are two typical situations of how people can get in a bus: a) when one person gets on/off the bus or b) two people pass each other. The process is also complicated because a person getting in a bus covers from 20 to 50 percent of the image; and in some situations (when two or more people are getting in a bus) people compose up to 90% of the image, and a moment of getting in a bus is very short, from 1 to 5 seconds (2s average). The solution is to use a camera with a wide–angle lens or hang the camera higher.

Authors of this paper together with JSC "Kauno autobusai" made an investigation of APCS market and defined that generally APCS based on computer vision has an accuracy of 90– 95%, and the prices are starting from

35 000 Lt for one bus. Kaunas Bus Park has 200 buses, so huge investments would be needed. Therefore authors have suggested and investigated four methods for tracking bus passengers, capable of acceptable recognition accuracy for practical applications, while maintaining a low cost of the system hardware – about 2000 Lt for a bus.

II. INVESTIGATION OF METHODS FOR PASSENGER COUNTING

All methods were tested with a real–life video material witch was collected from a prototype installation (notebook and USB camera) in one of the Kaunas public transport buses.

A. Method of barrier simulation [ABS]

In Fig. 1 two areas are presented (pixels1 and pixels2), where a difference initiated by the person getting on/off the bus is studied [8]. The direction of a passenger was registered by IF…THEN logic (below): if pixels1 area is crossed first and then pixels2 is crossed – the passenger gets on otherwise gets off.

```
change=video(t)-video(t-1)
Intensity1= ∑change(pixels1)
If Intensity1>threshold1, then object=1
Else object1=0
Intensity2= ∑change(pixels2)
If Intensity2> threshold2, then object2=1
Else object2=0
```

Here *t* is time, *video* is data from a camera, *pixels1* and *pixels2* are two pixel zones near the entrance to a bus. As the experiment results showed it is rational to choose a *threshold* value of 30% of preprocessed zone pixel intensity sum. Selected zones *pixel1* and *pixel2* sizes are set to 220x3. Other view is not analyzed and this improves quick–acting aspect of the method and allows us to analyze the image in real time.
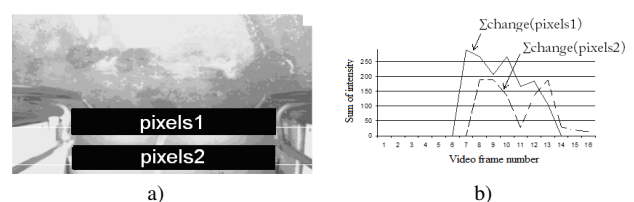

Fig. 1. Graph of the area variation during "getting in" phase.

People detection accuracy was 86% for a single passenger getting on/off the bus, however it could not detect people

who were passing each other or getting on a bus together, also it was sensitive to environmental variations (shadows and lighting changes).

B. Method Based On Intensity Maximum Detection [ABIMD]

This method detection is performed evaluating total intensity change [9] with respect to X and Y axis and by recording their maxima. Such method allowed us to observe the motion trajectory [10, 11] of the object and to evaluate duration of "getting in". Total projection on X axis only helps us to locate a person with respect to X axis. Total projection on Y axis varies with a motion towards/from the bus, therefore continuity of Y axis variation is much more important than continuity of X axis variation.
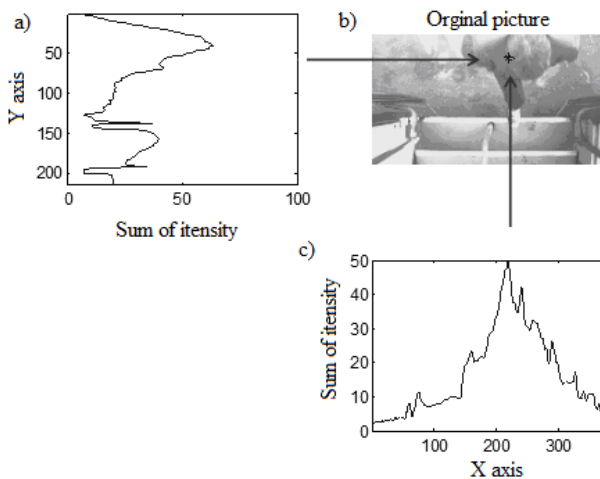

Fig. 2. Image of person detection: a – Y axis intensity graph; b – analyzed image; c – X axis intensity graph.
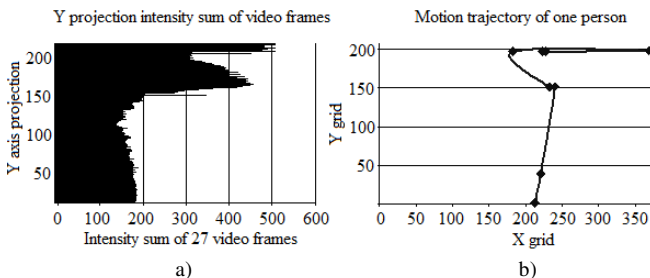

Fig. 3. Graph of total intensity in Y projection and obtained motion trajectory: a – Y projection intensity sum graph; b – identified motion trajectory of one person.

It was defined that large inaccuracies prevail in places where total intensity jumps occur (the reason being a steel hardware of entrance stairs, i.e. Fig. 3 on Y axis (image lines 130–214). Therefore this factor should be reduced by 50–70% (taking in account the average value of variations during boarding the bus) [12]. After lowering it by 50% a graph of total intensity with respect to Y axis projection was obtained.

The analysis of motion trajectory showed that an improvement is obtained, although due to several flaws inaccuracy still prevails. In attempt to solve this linear filter of moving average was implemented

$$y(n)=ax(n)+ax(n–1)+ax(n–2)+ax(n–3), \qquad (1)$$

where a is weight coefficient, n is intensity sum of Y axis projection row number. As the experiment results showed, it is rational to choose filter parameter a to 1/4.

This filter allowed us to reduce the influence of background noise and to correctly indicate the trajectory of passenger movement (Fig. 5).
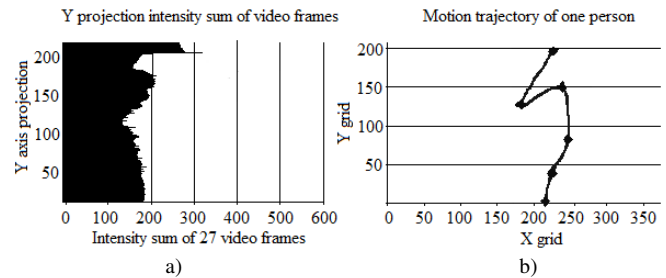

Fig. 4. Graph of total intensity in Y projection and obtained motion trajectory: a – Y projection intensity sum graph; b – identified motion trajectory of one person.
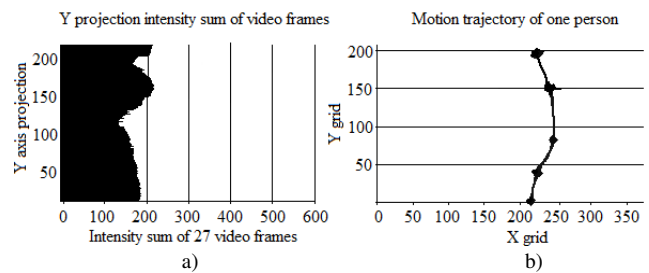

Fig. 5. Graph of total intensity in Y projection and obtained motion trajectory: a – Y projection intensity sum graph; b – identified motion trajectory of one person.

After performing a qualitative evaluation of the method in 70 different situations, the accuracy of 90% was observed for a single person's getting on/off a bus. This method was not suitable for situations when more than one person was getting on/off. It was possible to indicate the stops where passengers have difficulties in getting on/off the buss. This information would help to evaluate the quality of the driver's work, for example, approaching the pavement. If the driver approaches the pavement inconveniently, the average duration of the passenger boarding will be greater than one with the other drivers doing this correctly (comparing buses of the same type and mark). This would allow improving a service quality.

C. Method of barrier simulation for zones [ABSZ]

Good accuracy was observed while utilizing the method of barrier simulation, but it could not detect passengers who were passing each other or getting on/off at the same time. Therefore method for different zones was created. Method structure was the same as in [ABS] only with more zones. Detection is performed by differing 4 zones: pixels1, pixels2, pixels3, pixels4 (see Fig. 6). In the area of getting on/off (for the door which allows a passing of 2 persons at maximum) and evaluating independently, thus in the total intensity value in the image of each zone's area, detection was registered using IF…THEN logic.

Fig. 7 illustrates a variation caused by passengers going on the right and the left sides. A complicated situation was analysed: passengers passing each other, several people getting on/off (on the right side: 3 people getting off and 1

getting in, on the left: 2 getting in and 2 getting off). Arrows with zone names indicate in which zone a passenger was observed first. Time zones of Pixels2 and pixels4 were chosen four times smaller than zones of pixels1 and pixels3 (eight pixels wide for the best accuracy as determined by our experiments) to achieve a shorter calculation time. Because pixels1 and pixels3 were used only to determine whether a person is getting in or off, therefore their total variations value were smaller. The variation caused by the passenger flow was considerable enough; therefore this method was suitable for counting the passengers.

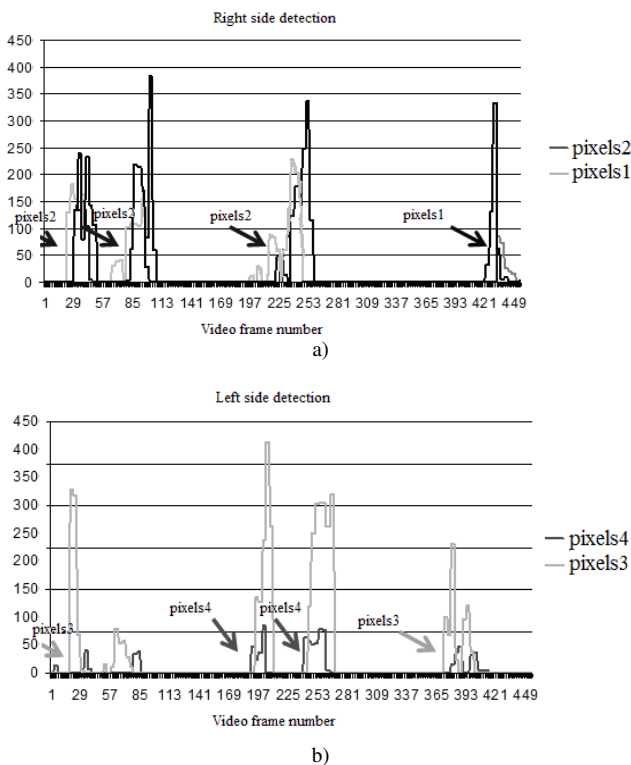Fig. 6. 4 zones are separated in the area of getting on/off.

a)

b)

Fig. 7. Graphs of 8 persons detection.

This method can count the passengers which are passing each other, getting on/off the bus together. Method works when the door is adjusted for 2 passengers maximum. This type of buses is the most popular one in our public transport sector (70% of the buses in Kaunas are of this type).

D. Method based on correlation of the object form [ACOF]

All projections of people are similar in a video signal; therefore a method was created to search for correlation in typical forms of the people. The necessity to separate the edges in the image was implemented using a MATLAB function EDGE [13]–[15] allowing 7 different methods for distinguishing the edges. A default "sobel" setting was chosen with no detailed qualitative investigation, because

with the default parameters it gave visible edges near a head or on other body parts, while other methods were too noisy or with lost links. Also "sobel" allowed achieving a shorter calculation time than others methods.

25 sub images of the heads of different people were collected. The passenger's head sub images were used as templates. A few handmade head shape correction were necessary, because some inaccuracies were observed in head edges. Then cross–correlation was calculated by using this formula

$$\gamma(u,v) = \frac{\sum_{x,y}\left[f(x,y) = \overline{f}_{u,v}\right]\left[t(x-u,y-v)-\overline{t}\right]}{\left\{\sum_{x,y}\left[f(x,y)-f_{u,v}\right]^2 \sum_{x,y}\left[t(x-u,y-v)-\overline{t}\right]^2\right\}^{0,5}}, \quad (2)$$

where f is the image, $\overline{t}$ is the mean of the template, $\overline{f}_{u,v}$ is the mean of $f(x,y)$ in the region under the template.

The peak of the cross–correlation matrix is registered for a video frame within each template. For experimental evaluation we used a video image where 3 different people were boarding the bus one by one. Results are presented in Fig. 8.
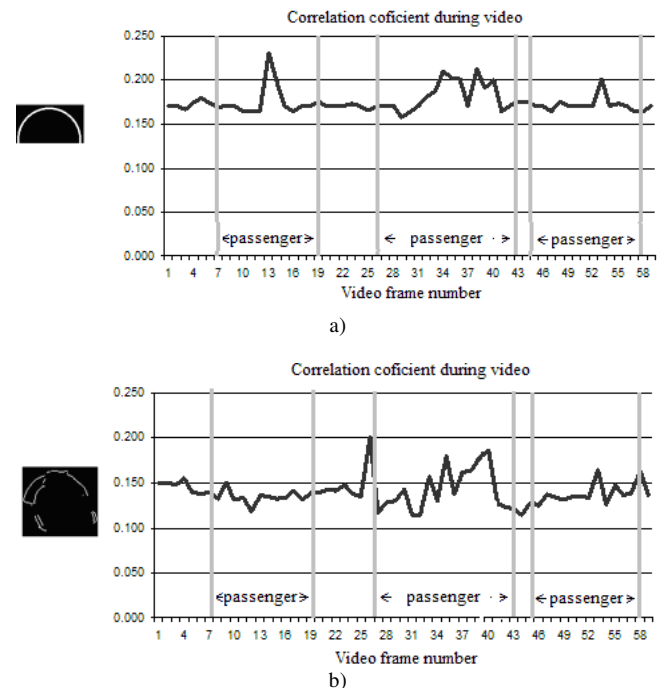

a)

b)

Fig. 8. Correlation coefficient during an analysis of video with 3 different passengers.

Correlation coefficient for head forms was in range from 0.15 to 0.4. The detection of a passenger was registered by IF …THEN logic:

```
If correlation_coefficient > threshold
Then object=1
Else object1=0
```

As the experiment results showed it is rational to use the data consisting of typical head forms and set a *threshold* to 0.2. Unfortunately using this setting the detection accuracy was 60% and only for the 46% of these recognized the direction was correctly identified. Therefore the other situation (when two people enter a bus) wasn't tested. A

comparison of detection accuracy of all the methods analyzed is given in a Table I.

This evaluation will be repeated in near future when more video data will be gathered and processed.

TABLE I. ACCURACY COMPARISON OF ALL INVESTIGATED METHODS.

Different situations		ABI			ABIMD			ABSZ			ACOF		
		Detected	Accuracy	T	Detected	Accuracy	T	Detected	Accuracy	T	Detected	Accuracy	T
1 passenger OUT	38	32	84%		34	89%		37	97%		18	47%	
1 passenger IN	32	28	88%		29	91%		30	94%		14	44%	
2 passengers IN (same time)	20	-	-		-	-		19	95%		-	-	
2 passengers OUT (same time)	22	-	-	0.06	-	-	0.06	20	91%	0.07	-	-	0.38
2 passengers bidirectional	30	-	-		-	-		27	90%		-	-	

*T time in seconds (s) need to proceed one video frame (Intel C2D P8600 CPU)

III. CONCLUSIONS

Total 4 methods for passenger detection were reviewed in this paper. Real life video data information were collected maintaining real–life scenarios in one of the Kaunas public transport buses and was used to test these methods. 214 passengers entered or leaved the bus during the experiment. When one passenger was getting on/off the bus the image should best be analyzed using ABIMD method (useful with a bus type where doors allow only one passenger at a time), otherwise it is better to use ABSZ method.

There were 32 in and 38 out situations for testing the detection of a single passenger, 20 simultaneous in, 22 simultaneous out and 30 bidirectional situations for testing the detection of two (same time in/out) passengers. ABI method allowed achieving 86% accuracy of recognition, while ABIMD method allowed achieving an increased 90% accuracy but this method only worked properly when a single person was present and no significant variation in lighting was present. ABSZ method showed over 90% accuracy in really complicated situations.

ACKNOWLEDGMENT

REFERENCES

[1] M. Petkevičius, A. Vegys, T. Proscevičius, A. Lipnickas, "Inspection system based on computer vision", *Elektronika ir Elektrotechnika (Electronics and Electrical Engineering)*, no. 10, pp. 81–84, 2011.

[2] I. J. Amina, A. J. Taylor, *Automated people–counting by using low–resolution infrared and visual cameras*. MIT press, 2007.

[3] D. Lefloch, *Real–time people counting system using a single video camera*. MIT press, 2008.

[4] D. Roqueiro, V. Petrushin, *Counting people using video cameras*. MIT press, 2007.

[5] P. White, *Public transport – its planning, management and operation,* 5th ed., UK, 2008.

[6] M. Rossi, A. Bozzoli. "Tracking and Counting Moving People" // *IEEE Proc. of Int. Conf. Image Processing*, vol. 3. pp. 212–216, 1994.

[7] J. W. Kim, K. S. Choi, D. B. Choi, *Real–time vision–based people counting system for the security door*. MIT press, 2004.

[8] J. A. Richards, X. Jia, *Remote sensing digital image analysis*, Germany, 2006.

[9] A. J. Lipton, H. Fujiyoshi, R. S. Patil, *"Moving Target Classification and Tracking from Real–time Video"*, in *4th IEEE Workshop on Applications of Computer Vision*, 1998, pp. 8–15.

[10] S. Gupte, O. Masoud, R. F. K. Martin, N. P. Papanikolopoulos, "Detection and Classification of Vehicles", *IEEE Transactions on Intelligent Transportation Systems*, vol. 3, no. 1, pp. 524–531, 2002. [Online]. Available: http://dx.doi.org/10.1109/6979.994794

[11] P. L. Rosin, T. Ellis, "Image difference threshold strategies and shadow detection", in *Proc. of the 6th British Machine Vision Conference*, 1994.

[12] R. Cucchiara, M. Piccardi, A. Prati, "Detecting moving objects, ghosts, and shadows in video streams", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1337–1342, 2003. [Online]. Available: http://dx.doi.org/10.1109/TPAMI.2003.1233909

[13] I. Hanisah, B. Zulkafli, *Object detection by using edge segmentation techniques,* 2009. [Online]. Available: http://dspace.fsktm.um.edu.my/handle/1812/815

[14] O. Vincent, O. Folorunso, "A Descriptive algorithm for Sobel Image Edge Detection", in *Proc. of the Informing Science & IT Education Conference (InSITE)*, 2009.

[15] W. Jianxin, C. Geyer, "Real–time human detection using contour cues", in *Proc. of the IEEE International Conference of Robotics and Automation (ICRA)*, 2011, pp. 860–867.