

A method for counting people attending large public events

K. Kopaczewski · M. Szczodrak · A. Czyzewski ·
H. Krawczyk

Published online: 4 August 2013

© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract The algorithm for people counting in crowded scenes, based on the idea of virtual gate which uses optical flow method is presented. The concept and practical application of the developed algorithm under real conditions is depicted. The aim of the work is to estimate the number of people passing through entrances of a large sport hall. The most challenging problem was the unpredicted behavior of people while entering the building. The examined flow of people fluctuated between individual persons and dense crowd. A series of experiments during sport and entertainment events was made. The results of the experiments show a high efficiency of the elaborated algorithm.

Keywords Crowd · People counting · Crowd behavior

1 Introduction

Mass crowd gatherings such as sport games or concerts can be a source of various risks for individuals, particularly evoked by excessive number of people in a specific place. Exceeding the value regarded as the safe limit may cause that in some emergency situations people would suffer injuries or death [12]. The organizers should know the number of people that are gathered in the building or in the enclosed outdoor space. Similarly to many objects of this type, the building considered in this work is not equipped with the people counting systems such as mechanical gates. Besides, people often feel concerned crossing such an

K. Kopaczewski · M. Szczodrak (✉) · A. Czyzewski
Multimedia Systems Department, Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, ul. Narutowicza 11/12, 80-233 Gdansk, Poland
e-mail: szczodry@sound.eti.pg.gda.pl

K. Kopaczewski
e-mail: kkop@sound.eti.pg.gda.pl

A. Czyzewski
e-mail: andcz@sound.eti.pg.gda.pl

H. Krawczyk
Computer Architecture Department, Faculty of Electronics, Telecommunications and Informatics,
Gdansk University of Technology, ul. Narutowicza 11/12, 80-233 Gdansk, Poland

installation, as to what would happen in case of necessity of rapid leaving the building. Moreover, the behavior of the crowd while entering an object through wide doors makes it impossible to use other optical or mechanical means such as radiation beam systems. The infrared barriers are ineffective because they often count a group of people being close each to other as a single person [11]. A problem is also that such solutions usually are not able to recognize the pedestrian movement direction.

Aside of video processing algorithms commonly used for surveillance of public places, the audio analysis provides an important supplement [8]. Nevertheless, when a large group of people appears, it gets very hard to determine their number by the majority of image processing methods which commonly use object detection and tracking. Background extraction-based approaches, such as the ones developed at the Multimedia Systems Department of Gdansk University of Technology [5] cannot separate objects properly when people walk at very small distances or when their hands are connected. Other methods use multiple cameras to deal with the segmentation problem [10, 14, 15], or apply models of human figures obtained during observing the foreground of an image [7, 13]. Moreover, installing numerous cameras would be impractical in the considered building.

Other solutions which utilize camera image together with laser beams for tracking feet, require deploying sensors on the ground [4].

The commercial systems diverge in the technologies applied and the actual target to solve. A laser counter offered in the market today and visible light systems are very common. These kinds of products often achieve best performance in some specific conditions, only. Meanwhile, the situations found while gathering the experimental data proven to be difficult to interpret algorithmically.

The methods of counting people in crowded scenes can be found in literature.

Albiol et al. [1] describe a technique based on the analysis of the derivate image constructed from a time sampled section of original image. Another method [2] uses statistical analysis of object corners detected while people move. Both latter methods were investigated in some specific conditions of underground train doors. Bozzoli et al. [3] propose an approach based on the sparse optical flow method and pedestrians contours obtaining by edge extraction from the image. However, similarity of such contours is not guaranteed because people may have various hair or clothes colors.

The proposed algorithm for counting people in the crowd differs from described approaches. It uses dense optical flow for motion analysis employing the so called virtual gate. The algorithm is dealing with complex situations which occur while people entering large sport halls. Moreover the algorithm is designed to work in a system with centralized architecture, where the video signals gathered from multiple cameras are being processed by an efficient computer cluster. The aim of the system is to show the estimated number of people while they are incoming to the hall through several gates. The KASKADA supercomputing platform is the algorithm working environment [9].

2 Virtual gate algorithm

The Virtual Gate algorithm is based on the modified Optical Flow method. The method developed for counting people does not involve classifying modules because the aim of the algorithm is to detect size and direction of motion of objects in video sequences having dimensions similar to the size of an average human body. Moreover, the Virtual Gate is used

in places where human motion is expected, especially at entrances, passes, etc. Two parts of algorithm can be distinguished: the main module which performs image processing and the calibration module.

Virtual Gate is devoted to counting people in crowd passing through the scene observed by the camera. The illustration of the sample setup of the virtual gate is presented in Fig. 1a. The Virtual Gate distinguishes two directions of people motion, namely “in” and “out” (see Fig. 1a).

The detailed structure of the Virtual Gate is depicted in Fig. 1b. It is composed of a set of rectangle regions (R_i) situated next to each other and overlapping. Rectangles have identical shape and their size is corresponding to the size of an average human body contour, with respect to its height and width (for a particular camera view). Rectangles are spaced evenly along x axis and they are 80 % overlapped.

The motion of objects is estimated in each region R_i using the Dense Optical Flow method [6]. Choice of Dense Optical Flow was made in order to obtain motion description per each pixel. The set of vectors representing the direction and the velocity of the motion detected is obtained as the result of the operation above. Displacement vectors can be expressed by the planar vector field:

$$\mathbf{V} = V_x(x, d)\mathbf{i} + V_d(x, d)\mathbf{j} = V_\rho(x, d)\mathbf{e}_\rho + V_\varphi(x, d)\mathbf{e}_\varphi \tag{1}$$

where:

\mathbf{i}, \mathbf{j} unit vectors of x and d axes

$\mathbf{e}_\rho, \mathbf{e}_\varphi$ unit vectors related to polar coordinates (the distance from the axis of symmetry, the angle measured counterclockwise from the positive x -axis).

Two directions of people motion through the virtual gate are considered, namely forward and backward (“in” and “out”, $+d$ and $-d$, as in Fig. 1c). Moreover, a small divergence of the direction ($\pm\alpha$) is allowed, because usually people do not maintain bearing while walking. The tolerance should not be too large, because of the need to discard those walking along the gate.

The block diagram of the algorithm (for individual region R_i) is presented in Fig. 2. For each input video frame, vectors representing motion speed and direction are calculated.

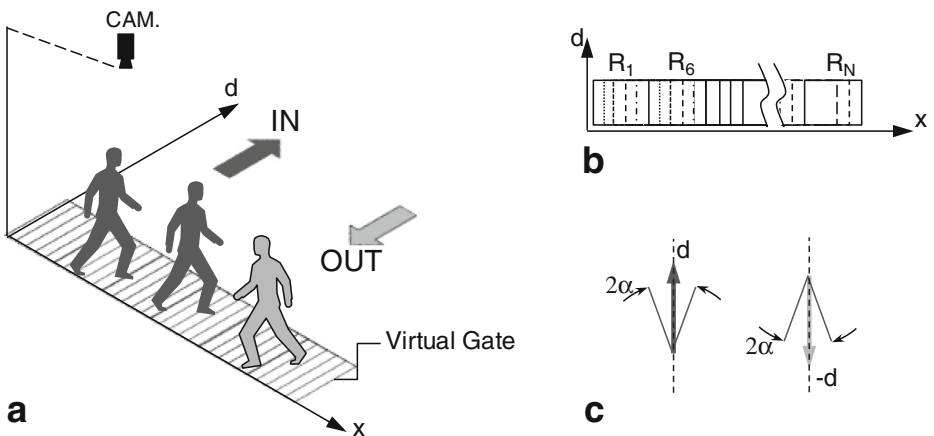


Fig. 1 Virtual Gate: setup for counting people (a) and details illustrating its principle of working (b, c)

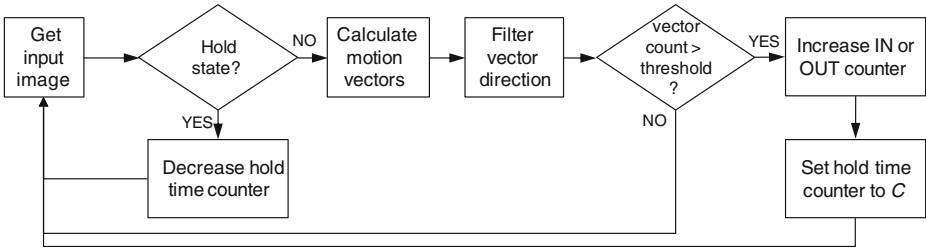


Fig. 2 Block diagram of the virtual gate main algorithm for a region R_i

Components of Eq. (1) can be written in a form of cylindrical (in this two dimensional case - polar) coordinates:

$$\begin{aligned} V_x &= V_\rho \cos\varphi + V_\varphi \sin\varphi \\ V_d &= V_\rho \sin\varphi + V_\varphi \cos\varphi \end{aligned} \tag{2}$$

Let φ_0 denote the angle corresponding to the direction pointed by d . New functions V_x^I, V_d^I, V_x^O and V_d^O are calculated as given in Eq. (3) in order to obtain desired direction vectors (I-means “in”, O-“out”):

$$\begin{aligned} V_x^I &= \begin{cases} V_x & \text{if } \varphi_2 \leq \varphi \leq \varphi_1 \\ 0 & \text{otherwise} \end{cases} \\ V_d^I &= \begin{cases} V_d & \text{if } \varphi_2 \leq \varphi \leq \varphi_1 \\ 0 & \text{otherwise} \end{cases} \\ V_x^O &= \begin{cases} V_x & \text{if } \varphi_2 + \pi \leq \varphi \leq \varphi_1 + \pi \\ 0 & \text{otherwise} \end{cases} \\ V_d^O &= \begin{cases} V_d & \text{if } \varphi_2 + \pi \leq \varphi \leq \varphi_1 + \pi \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{3}$$

where: $\varphi_1 = \varphi_0 + \alpha,$
 $\varphi_2 = \varphi_0 - \alpha.$

Subsequently, the number of origins of vectors directed towards “in” (L^I) or “out” (L^O), enclosed in each region R_i is obtained according to the following Eq. (4):

$$L^{(*)} = \sum_{i=1}^I \sum_{j=1}^J \tilde{V}_{i,j} \tag{4}$$

where: $\tilde{V}_{i,j} = \begin{cases} 1 & \text{if } |\mathbf{V}_{i,j}^{(*)}| > T_M \\ 0 & \text{otherwise} \end{cases},$

T_M vector magnitude threshold
 $\mathbf{V}_{i,j}^{(*)} = \mathbf{V}^{(*)}(x_i, d_j)$ vector with origin at (x_i, d_j) , refer to Eqs. (2) and (3)
 I, J number of points in region R_i along x and d axes, respectively.

In the next step, L^I and L^O are compared to a threshold value T_S which is proportional to the average area of human silhouette at a given camera view, the value being obtained experimentally. If $L^{(*)}$ is greater than the threshold value T_S , the “in” or “out” people counter is increased and rectangle R_i enters an inactive (“hold”) state for the period of C frames. The inactive state means that calculations are not performed for region R_i . This operation is done

Fig. 3 View of entrance in recording 6, frame 13580. Ticket checker position marked with yellow ellipse. Compare placement of litter bin (orange contour) to e. g. Fig. 4



in order to avoid counting errors and to allow leaving the area of R_i by the moving object while not being counted more than once. The period of the “hold” state is obtained experimentally during the calibration.

The aim of calibration is to find the optimal counting threshold (T) in order to improve the effectiveness of the algorithm. Searching is based on the bisection method which approaches the optimal result in sequential iterations. The input data for the calibration process are:

- video sequence presenting passage of individuals and groups of people,
- the Virtual Gate geometry and parameters (i.e. dimensions of regions R_i , number and distance between regions),
- number of people passed through the gate in selected time moments.

The error of counting is calculated in successive iterations as presented in Eqs. (5) and (6):

$$S_n = \begin{cases} w_n \cdot (y_n - p_n) & \text{if } n = 0 \\ w_n \cdot [(y_n - y_{n-1}) - (p_n - p_{n-1})] & \text{if } n > 0 \end{cases} \quad (5)$$

$$S = \sum_{n=1}^N S_n \quad (6)$$

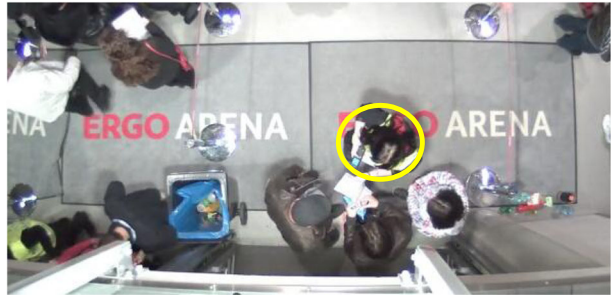
where:

- N number of selected time moments
- $p_n = p_0, p_1, \dots, p_N$ real number of people passed through the gate during the period between the instants $n-1$ and n (the Ground Truth)
- $w_n = w_0, w_1, \dots, w_N$ weight coefficient.

Fig. 4 View of entrance in recording 6, frame 59955. Position of ticket checker marked with yellow ellipse



Fig. 5 View of entrance in recording 6, frame 63650. Position of ticket checker marked with yellow ellipse



The variables are defined as follows:

- $y_n = y_0, y_1, \dots, y_N$ the number of moving objects counted during the period between the instants $n-1$ and n by the Virtual Gate algorithm
- $S_n = S_0, S_1, \dots, S_N$ counting error in selected time moment
- S total counting error.

The weight coefficient is added in order to favor either passage of individuals or groups of people. In case of crowded scene value of w_n is decreased and in case of individuals, it is increased. In each step, partial errors (Eq. (5)) and the total error (Eq. (6)) are minimized and then the counting threshold is increased or decreased respectively according to Eqs. (7) and (8). The calibration process stops when $T_{corr} \leq 1$.

$$T = \begin{cases} T + T_{corr} & \text{if } S > 0 \\ T - T_{corr} & \text{if } S \leq 0 \end{cases} \quad (7)$$

$$T_{corr} \leftarrow \frac{T_{corr}}{2} \quad (8)$$

where:

- T counting threshold (value within range 0, 1...100)
- T_{corr} counting threshold correction.

Fig. 6 View of entrance in recording 6, frame 101185. Position of ticket checker marked with yellow ellipse

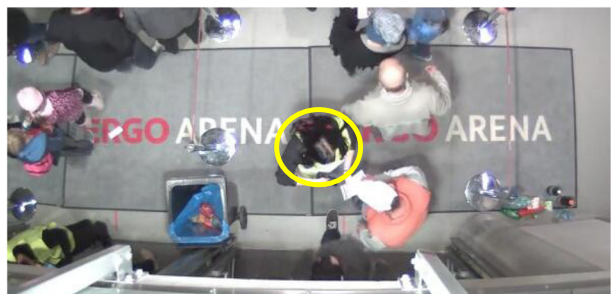
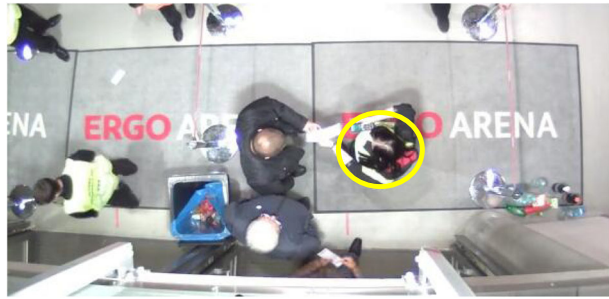


Fig. 7 View of entrance in recording 6, frame 110505. Position of ticket checker marked with yellow ellipse



3 Experimental results

The experiments were made in the sports and entertainment hall which maximum capacity is 15.000 of people (“Ergo Arena” located in Gdansk). The image acquisition hardware was deployed at the main entrance which consists of 6 symmetrical doors. The cameras were set to observe only 3 doors because the others were not being used during the events, frequently. The cameras were installed at the height of 6.5 m above the floor, whereas the width of the door is 2.9 m. Image acquisition speed was 30 frames per second, resolution 640×360 points.

Total number of 11 recordings have been gathered during sport and entertainment events. Each represents a real situation of people and crowd entering the sports and entertainment hall, whereas the length of each is about 1.5 h. The total length of the experimental material is about 16.5 h. The number of people in each test recording was counted manually and treated as Ground Truth reference value, which was later compared to the Virtual Gate algorithm output.

The people counting system based on Virtual Gate algorithm can count in real time people passing through numerous entrances. The image gathered from multiple camera is being analyzed simultaneously on a supercomputer with the support of the KASKADA platform. Such a centralized architecture provides simplicity of changing the counting system scale.

3.1 Real situations during people counting

The algorithm was examined in real conditions, while crowds have been entering the sport and entertainment hall. Practically some situations posed considerable difficulties for people

Fig. 8 View of entrance in recording 6, frame 115544. Position of ticket checker marked with yellow ellipse

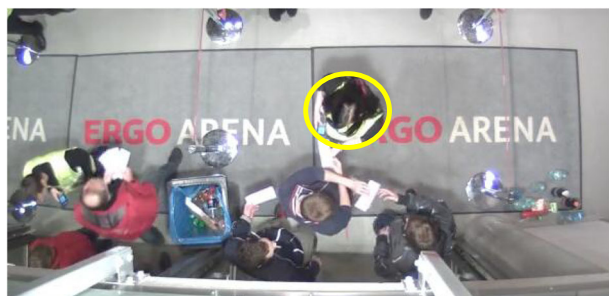


Fig. 9 View of entrance in recording 7, frame 92211. People staying or loitering in counting area



counting algorithms. Below, we present some examples of people behavior and encountered problems.

Counting errors were caused by changing the position of crowd control barriers by ticket checkers. Placement of a litter bin has effected people moving back and a unpredictable behavior in the area of Virtual Gate constituting the counting region. This situation is shown in Fig. 3: the ticket controller (wearing a yellow vest) has moved the litter bin to position that disturbs people flow.

Other conditions which certainly might cause counting errors were fluctuations in the position of a person who checks tickets. This person should stand in the area seen in the bottom part of Fig. 4. The ticket checking in this location evokes some crowd congestions and stand still people (including the ticket checker) in the counting area. The behavior of ticket checker and crowd is presented in Figs. 4, 5, 6, 7, and 8.

Another example of difficult situation met in practice is that people were staying in the counting area. Such a behavior causes counting errors, because the Virtual Gate may count persons more than once. People seen in the photo (Fig. 9) were chatting and frequently changing position (about two steps in diverse directions).

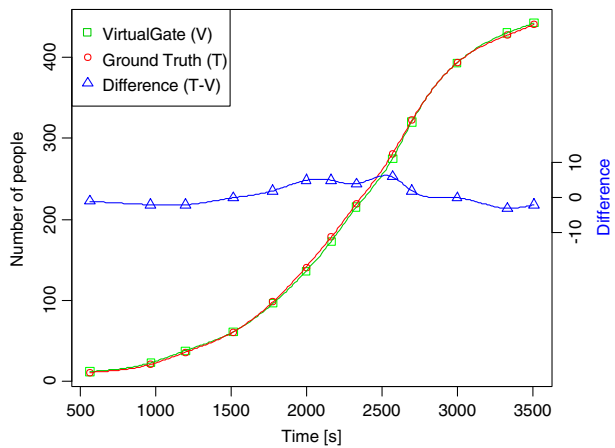
3.2 People counting results

During the offline test procedure of people counting system a set of 11 Virtual Gate algorithms was initialized according to the number of recordings. In the first phase of experiments, the test was conducted with identical parameters of each Virtual Gate, without any calibration. The parameters included: size of region, distance between adjacent regions, counting threshold (see Section 2). In the second phase of experiments parameters were tuned in order to improve the accuracy of people counting. Results obtained for recordings 3, 5, 8, 9, 10 were satisfactory in the first phase, thus corrections were not applied. The optimal counting threshold found for each test recording is presented in Table 1. Counting threshold had to be verified individually for each recording. Dissimilarities are caused by two main reasons. The first one arises from the differences in camera placement and changes the geometry of the virtual gate. The size of each rectangle region has to be fit according to the camera parameters. The second one results from the character of people motion. For example in recording 4, people were entering the hall steadily and mainly individuals passed

Table 1 Counting threshold (T) parameter of Virtual Gate for test recordings

Recording	1	2	3	4	5	6	7	8	9	10	11
T	65	65	68	62	68	65	70	68	68	68	65

Fig. 10 Result of people counting for recording 3



through the virtual gate. In the recording 7, the prevailing motion of two or three people simultaneously through the gate was observed.

The detailed results of people counting obtained by the Virtual Gate algorithm compared to true number of people, for test recordings 3, 6 and 7, are presented in Figs. 10, 11 and 12. Outcomes of Virtual Gate algorithm obtained for the recording 3 represent the best accuracy of counting.

Recordings 6 and 7 represent cases of the worst achieved counting error values (overestimation and underestimation of number of people, respectively). The error value shown in Fig. 11 is rising rapidly between 33rd and 35th minute of recording. In this period a very dense crowd has pushed towards the hall and groups of people moved chaotically in the area of virtual gate. Moreover, the position of the ticket checker was not constant. In case of chart depicted in Fig. 12, the error is increasing significantly between 35th and 50th minute of the experiment. The error is mainly caused by stopping people in the counting area and by unpredicted movements (i.e. moving back and partially re-entering the counting area).

The measurement of accuracy of Virtual Gate algorithm was obtained by comparing its counting result to the true (reference) number of people. Accuracy of Virtual Gate algorithm (A) is calculated according to the following equation:

Fig. 11 Result of people counting for recording 6

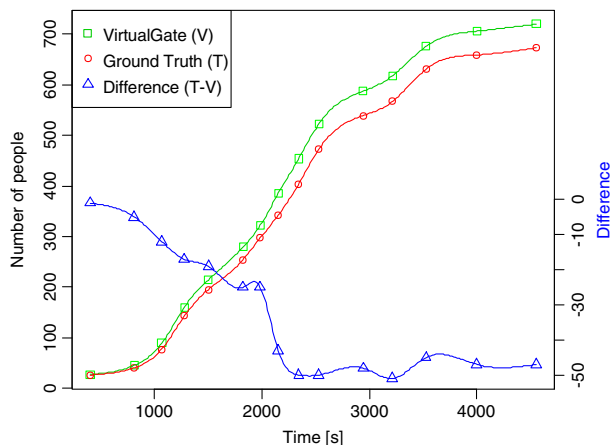
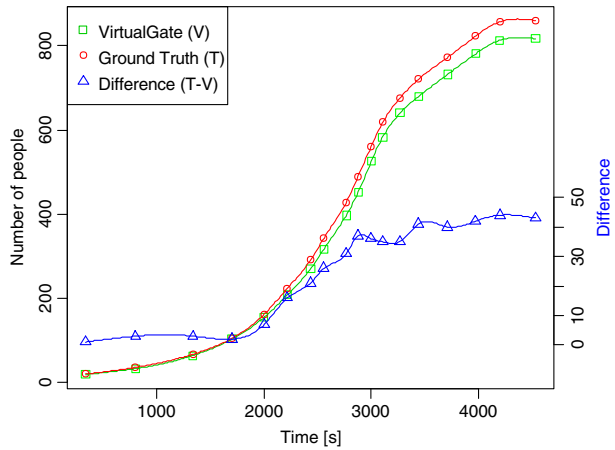


Fig. 12 Result of people counting for recording 7



$$A_i = \frac{N_i - |N_i - V_i|}{N_i} \cdot 100\% \tag{5}$$

where:

- V number of people obtained by Virtual Gate
- N reference (true) number of people
- i denotes recording number.

The resulting obtained accuracy of the algorithm is presented in Table 2. The obtained accuracy of counting is very good, since the highest achieved value is 99.7 % and the lowest is 93.1 %.

Table 2 Accuracy of Virtual Gate algorithm

Recording	Count (Virtual Gate)	Count (Ground Truth)	Accuracy of V.G. algorithm (A) [%]
1	644	642	99.69
2	995	1017	97.84
3	440	442	99.55
4	555	573	96.86
5	739	771	95.85
6	720	673	93.47
7	817	860	94.74
8	324	336	96.43
9	402	420	95.71
10	420	433	97.00
11	445	478	93.10
Total	6497	6649	97.71



4 Conclusions

The concept and practical application results of the algorithm for people counting in a crowd passing through the gates of large sport object were presented in this paper. We described the approach to the analysis of motion data obtained with the dense optical flow estimated with the devised Virtual Gate algorithm. The algorithm efficiency was examined using some real videos recorded at a large sport and entertainment hall (Ergo-Arena in Gdansk). The counting precision achieved of 97.6 % on an average at total 6,649 persons, is satisfactory considering the real people behavior which was far from the organized movement. A further work will focus on extending the system to operate at other entrances to the hall of the Ergo-Arena object being used less frequently. Moreover, changes of organization of the process of crowd entering the building would be necessary in order to improve the algorithm accuracy. The future work may also focus on a practical application of the system to other large public objects.

Acknowledgements Research funded within the project No. POIG.02.03.03-00-008/08, entitled “MAYDAY EURO 2012—the supercomputer platform of context-dependent analysis of multimedia data streams for identifying specified objects or safety threads”. The project is subsidized by the European regional development fund and by the Polish State budget.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

1. Albiol A, Mora I, Naranjo V (2001) Real-time high density people counter using morphological tools. *IEEE Trans Intell Transp Syst* 2(4):204–218
2. Albiol A, Silla J (2009) Statistical video analysis for crowds counting, In: 16th IEEE International Conference on Image Processing (ICIP), 2009, pp 2569–2572
3. Bozzoli M, Cinque L, Sangineto E (2007) A statistical method for people counting in crowded environments, In: 14th International Conference on Image Analysis and Processing, 2007. ICIAP 2007, pp 506–511
4. Cui J, Zha H, Zhao H, Shibasaki R (2007) Laser-based detection and tracking of multiple people in crowds. *Comp Vision Image Underst* 106(2–3):300–312
5. Czyzewski A, Dalka P (2008) Moving object detection and tracking for the purpose of multimodal surveillance system in urban areas. In: *New directions in intelligent interactive multimedia. Studies in computational intelligence.* Springer, Berlin, pp 75–84
6. Farneback G (2003) Two-frame motion estimation based on polynomial expansion. In: *SCIA13: Gothenburg, Sweden, 2003*, pp 363–370. *Lecture Notes in Computer Science*, 2749
7. Ge W, Collins RT (2009) Marked point processes for crowd counting, *Proc. CVPR*, 20–25 June 2009, 2913–2920
8. Kotus J, Lopatka K, Czyzewski A (2012) Detection and localization of selected acoustic events in acoustic field for smart surveillance applications, *Multimedia Tools and Applications*, published online 31 July 2012, pp. 1–17 (2012), doi:[10.1007/s11042-012-1183-0](https://doi.org/10.1007/s11042-012-1183-0)
9. Krawczyk H, Proficz J (22–24 July 2010) The task graph assignment for KASKADA platform, *Proc. 5th International Conference on Software and Data Technologies*
10. Lim J, Kim W (2012) Detecting and tracking of multiple pedestrians using motion, color information and the AdaBoost algorithm, *Multimedia Tools and Applications*, published online 16 June 2012, pp 1–19, doi:[10.1007/s11042-012-1156-3](https://doi.org/10.1007/s11042-012-1156-3)
11. Mathews E, Poigne A (2009) Evaluation of a “Smart” pedestrian counting system based on echo state networks. *EURASIP J Embed Syst* 1:1–9
12. Mollen M (Jan. 1992) A failure of responsibility - report to Mayor David N. Dinkins on the December 28, 1991 tragedy at City College of New York



13. Szczuko P (2012) Genetic programming extension to APF-based monocular human body pose estimation, *Multimedia Tools and Applications*, published online 13 June 2012, pp 1–16, doi:[10.1007/s11042-012-1147-4](https://doi.org/10.1007/s11042-012-1147-4)
14. Wang Y, Velipasalar S, Gursoy MC (2012) Distributed wide-area multi-object tracking with non-overlapping camera views, *Multimedia Tools and Applications*, published online 13 November 2012, pp 1–33, doi:[10.1007/s11042-012-1267-x](https://doi.org/10.1007/s11042-012-1267-x)
15. Yang DB, Gonzalez-Banos HH, Guibas LJ (2003) Counting people in crowds with a real-time network of simple image sensors, *Proc. of the Ninth IEEE International Conference on Computer Vision*, 13–16 Oct. 2003, 122–129



Krzysztof Kopaczewski completed his studies at the Mathematics, Physics and Informatics department of University of Gdańsk in 2008 acquiring M.Sc. degree. The subject of his thesis was measurement of noise level from light and heavy vehicles in theory and practice. His experience is: calculation of “digital acoustic maps”, influence of noise on the environment, acoustics environment measurement.



Maciej Szczodrak Received his M.Sc. degree from the Gdansk University of Technology in 2006. His thesis concerned implementation of algorithms for open air sound propagation analysis. He is interested in audio and video processing, environmental acoustics. He is also concerned about programming. So far he has published, as author or co-author, 30 publications including international journal conference papers and journal articles.





Prof. Andrzej Czyzewski Head of the Multimedia Systems Department is author of more than 400 scientific papers in international journals and conference proceedings. He has led more than 30 R&D projects funded by the Polish Government and participated in 5 European projects. He is also author of 8 Polish patents and 4 international patents. He has extensive experience in soft computing algorithms and sound & image processing for applications among others in surveillance.



Prof. Henryk Krawczyk Rector of Gdansk University of Technology, four times he was the Dean of the Faculty of Electronics, Telecommunications and Informatics. He is an author of 284 publications, including 52 abroad, 6 monographs. He promoted 15 doctors, including three foreign (Lithuania, Iraq, Yemen). He managed 12 grants and 12 European national grants. He realized 12 deployments and built 7 complex information systems.