Automatic people density maps generation with use of movement detection analysis

M. Parzych[1], A. Chmielewska[1], T. Marciniak[1], A. Dąbrowski[1], A. Chrostowska[2], M. Klincewicz[2]
[1]Poznan University of Technology, Poznan, Poland,
{marianna.parzych, agata.chmielewska, tomasz.marciniak, adam.dabrowski}@put.poznan.pl
[2]BranchBrothers, Poznan, Poland,
{agnieszka.chrostowska,michal.klincewicz}@branchbrothers.pl

Abstract. **The paper presents technical aspects related to the video computer analysis of people movement in the sales rooms. Automatic generation of the people density maps (also referred to as the "heat maps") is helpful for the analysis of customers' and employees' behavior in sale and office rooms advertising effectiveness research or tracking movements. Generation of the density maps was realized with the video recordings acquired by typical video surveillance cameras. Data processing was done off-line but the achieved processing time allows the real-time implementation. Experimental tests were carried out at several points of the customer services.**

***Keywords*: density (heat) maps, crowd analysis, behavior analysis, people tracking**

I. INTRODUCTION

Despite the development of online shopping, physical distribution sale network remains the most important channel of the customer service. Analysis of psychology of the consumer behavior in the sale room aims to increase the sale success. Based on even short-time observations, consulting companies officers can draw conclusions that can very effectively improve functioning of the sale area.

Conventional solutions for the customer behavior analysis in a store or an office require the use of manual analytical tools like questionnaires or physical traffic counting by interviewers. Such methods are cumbersome and expensive, thus impossible to operate in a large number of rooms over a long time. An interesting and attractive option is the use of the CCTV (*closed circuit television*) system, since most commercial spaces are equipped with a video monitoring. By this means manual customer behavior observation can be supported by advanced systems for automatic analysis of video sequences.

Automatic systems, which are subject of this paper, can be equipped with intelligent image analysis facilities that allow motion detection, classification of moving objects, and behavior analysis. For example, detection of people crossing a line in one or both directions can be used for counting them.

In case of the commercial sale space usage analysis, it is useful, from perspective of the consumer behavior analysis, to generate people density maps (referred also to as the "heat" maps). A density map is a graphical representation of a table data with individual values represented as particular colors. Very popular are web page heat maps used for displaying areas of web pages most frequently scanned by visitors.

Generation of density maps based on monitoring data is offered by several companies engaged in CCTV systems e. g. MxAnalytics [1]. It, however, turns out that, due to the nature of registration (camera placement, type of the room, lighting, conditions, number of the analyzed individuals, etc.), a proper interpretation of the image and the corresponding real world is a relatively difficult task, often impossible for users with the commercial software only. Proper selection of the optimal parameters for the image processing and the density map generation is typically not available in the commercial software. Therefore just this is the subject of our research.

II. RELATED WORK

The people density maps generation is mainly related to the crowd analysis. Recently Davies, Yin, and Velastin tested elements of the image processing for an estimation of the crowd density and crowd motion [2].

Estimation of crowd density based surveillance monitoring is provided in [3] with the use of the texture information based on grey level transition probabilities.

Motion heat maps can be built by accumulation of binary BLOBs (*binary large objects*) of moving objects [4]. Foreground objects can be extracted by fusing the classification results from both stationary and moving pixels [5]. Generation of motion heat maps applied to the crowd monitoring on the escalator is presented in paper [6].

III. VIDEO ACQUISITION

Proper localization of cameras for video sequences acquisition is very significant for quality of the collected data. During the recording of test videos, which are used in our experiments with the *Heat Maps Generator* program, it is important that the camera should be placed in the room at an appropriate height. The best place is close to the ceiling, thus the people in the room are not behind each other and every person is clearly visible. It has a positive impact on the quality of the moving object detection algorithm.

If angle of the camera view encompasses the entire room properly, the usage of a single camera is possible. In case of long and narrow rooms, two cameras should be placed close to the ceiling in two opposite corners of the room. Such a camera arrangement allows well observation of people location and position from both sides.

978-1-4673-5637-4/13/$31.00 ©2013 IEEE

The acquisition rate in the *Heat Maps Generator* program is irrelevant. The algorithm (in the offline mode) processes data two times faster than the recording speed, which is 25 frames/sec (for 704×288 pixels resolution, processed by Intel®Core™i7-2620M CPU@2,7GHz). The image resolution is also not very significant. However, the size of the moving objects (people) is important. According to the *PN-EN 50132-7: CCTV and alarm systems* norm [7], the object in the motion detection process should occupy at least 10% of the height of the image (for the CCTV systems with the minimum image height of 480 lines – Fig. 1).

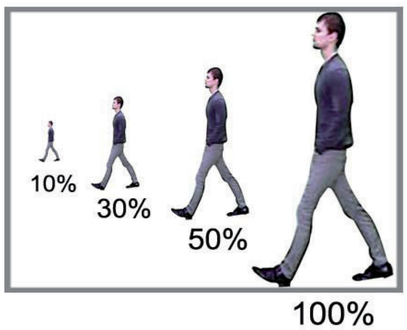


Fig. 1. Illustration of the recommended minimum size of the observed object according to PN-EN 50132-7: CCTV & alarm systems norm

During recordings the *KTC3510* color camera from *KT-VISION* with *TAV308DC 3-8mm SIRIUS F* lens was used. The most important camera features are: 1/3" SuperHAD CCD (*charge coupled device*), movement detection algorithm, definitions of privacy zones, DNR (*dynamic noise reduction*) and strong light compensation. For the video signal recording *Falcon DDE04 MINI KPL* was used. Four cameras can be connected to this DVR (*digital video recorder*). The recordings can be saved in the *H.264* format at the maximum processing speed of 100 frames/sec.

IV. Moving Objects detection techniques

Density map generation requires a simple and accurate algorithm for moving objects detection. Below we shortly describe basic techniques for video sequence processing.

A. Background subtraction

Background subtraction is a popular method for motion detection in video sequences, used, among others, by Liying [8] and Azeb [9]. This involves the determination of the difference between the current frame and the reference frame (scene without moving objects). Due to possible changes in the lighting conditions and geometry settings, regular updating of the background image is necessary.

Fast and low memory-cost background estimation is the running Gaussian average [10]. This method is based on fitting a Gaussian probability density function on *n* last images. The running average is computed as

$$\mu_t = \alpha I_t + (1-\alpha)\mu_{t-1}, \tag{1}$$

where I_t is an actual pixel value, μ_{t-1} is a previous computed average and α is an empirically chosen weight.

B. Gaussian Mixture Model

Another background estimation technique is the Gaussian mixture modeling [11]. It is based on an assumption that a pixel has no correlation with other pixels in an image [12]. Each of the background image pixels is estimated by a mixture of k Gaussian distributions. The probability of pixels X at time t can be defined as

$$P(X_t) = \sum_{i=1}^{k} \omega_{i,t} \cdot \eta(X_t, \mu_{i,t}, \Sigma_{,t}), \tag{2}$$

where k is the number of Gaussians (usually k is an integer between 3 and 5 [13]), X_t is a current pixel value vector, $\omega_{i,t}$ is the weight of *i*-th Gaussian distribution and

$$\sum_{i=1}^{k} \omega_{i,t} = 1 \;. \tag{3}$$

The *i*-th Gaussian distribution at time t is $\eta(X_t, \mu_{i,t}, \Sigma_{,t})$, where $\mu_{i,t}$ is the mean, and $\Sigma_{,t}$ represents covariance matrix.

For each pixel model parameters are adjusted. The Gaussians are ordered by the value of $\omega/|\Sigma|$ and the first distribution of Gaussians is selected as the background model [12]

$$B = \arg\min_b \{\sum_{i=1}^{b} \omega_{i,t} > T\}, \tag{4}$$

where T is the threshold, and B is a component responsible for the background.

C. Subtraction of consecutive video frames

A technique based on subtraction of consecutive frames is a very fast, low memory, and low calculation cost motion detection algorithm. It has strong adaptability. However, this method cannot extract all relevant characteristics of pixels. This method is often used as a supplement to other motion detection algorithms [14].

D. Optical Flow

Optical flow is a motion detection method based on extracting the movement of pixels in the image. a principle of the optical flow algorithms is as follows: by comparing consecutive frames of the video sequence, the correlation between them is found. Then a vector table called optical flow field is created. These vectors define the shift of pixels or regions caused object motion in relation to the camera.

There are several methods to estimate the optical flow fields [15]: region-based, energy based, phase-based and differential. The most common are: the Lukas-Kanade (LK) and Horn-Schunck (HS) methods [16, 17].

The LK method is based on comparing consecutive frames of the video, assuming brightness constancy between them. The optical flow equation is as follows [18]:

$$I_t + I_x u + I_y v = 0\,, \tag{5}$$

where $I(x,y,t)$ represents the brightness at pixel location (x,y) and time t, I_x, I_y and I_t are the partial derivatives of $I(x,y,t)$, which respect to x,y and t, (u,v) is the velocity corresponding to the optical flow of $I(x,y,t)$ and can be computed as

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \dfrac{S_{yy}S_{xt} - S_{xy}S_{yt}}{S_{xx}S_{yy} - S_{xy}^{2}} \\ \dfrac{-S_{xy}S_{yt} + S_{xx}S_{yt}}{S_{xx}S_{yy} - S_{xy}^{2}} \end{pmatrix}, \quad (6)$$

where S_{xx}, S_{xy}, S_{yy}, S_{xt} and S_{yt} are the product sum i.e. $S_{ab} = \sum_{\Gamma} I_a I_b$ over the small range Γ.

V. People Density Maps Generator - software

A. Algorithm

The software to generate the density maps were prepared with the use of the Microsoft Visual Studio 2010 development environment and OpenCV library version 2.3. The program is written in C/C++. The algorithm processes the input video sequence frame-by-frame. a simplified block scheme of the program stages is shown in Fig. 2.

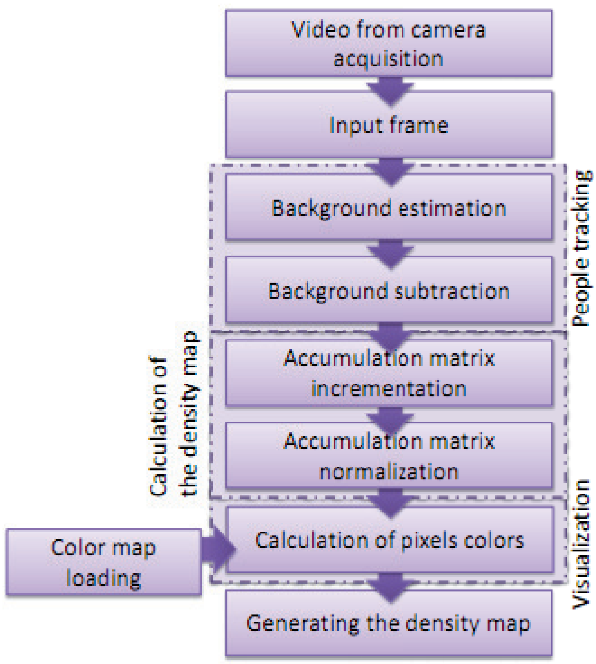


Fig. 2. Schema of the program processing stages

People tracking is realized by the background subtraction. Consecutive video frames subtraction method was rejected because it does not detect a person who has not changed its position in relation to the previous frame. Background estimation uses the moving average [19] of a variable number of frames. Subtraction of constant background (set at beginning of the algorithm) would be sensitive to changes in background due to e.g. moving shadows and streaks of light. Gaussian mixture and the optical flow methods require a more complex implementation than the background estimation using moving average, which provides a satisfactory high-performance motion detection. The result of the background subtraction is then converted to the grayscale and thresholded. At the next step the image is dilated and erosed [19], to ensure that every person is detected as a single object. It is not necessary to check additional geometric parameters in order to distinguish humans from others objects (such as vehicles). The study included indoor commercial spaces, where were no moving objects other than humans. The tracking algorithm is shown in Fig. 3.

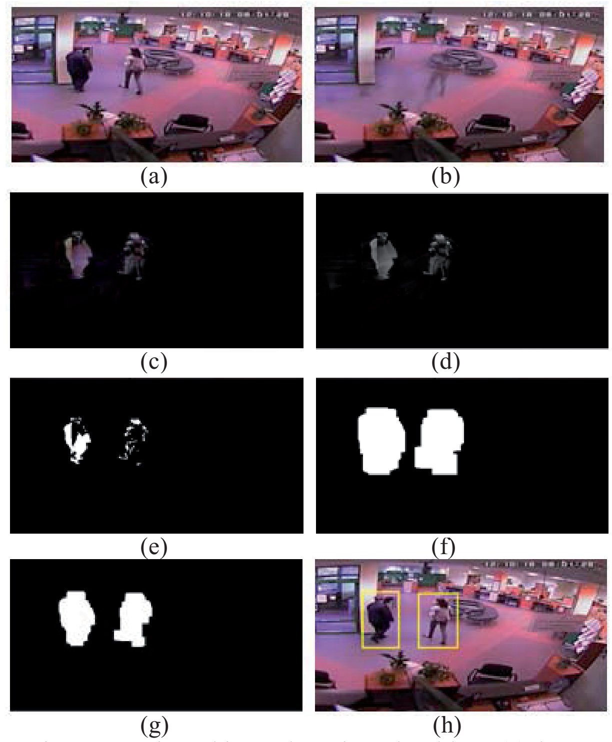


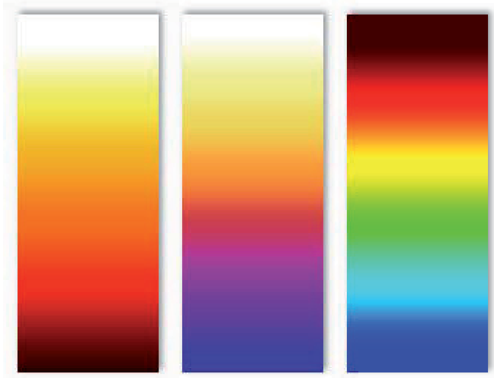
Fig. 3. Moving objects detection algorithm: (a) input, (b) estimated background, (c) subtracted background, (d) grayscale, (e) thresholding, (f) dilation, (g) erosion, (h) final result.

Each detected person is described by the coordinates of the vertices of a rectangle surrounding it. Then the center of the rectangle is calculated. In the two-dimensional accumulation matrix the corresponding cell is incremented. To prevent the datatype overflow, the values in the accumulation matrix are divided by 2 after achieving given threshold. This has no negative impact on the final result, because the algorithm does not calculate the absolute number of people of movements during the sequence, but the differences in people density in different parts of the analyzed spaces. After the processing of the entire video sequence the density map is generated based on the values in the accumulation matrix. During the accumulation process the maximum values is searched for. The constant A is calculated according to equation

$$A = 255 / Maximum\_value \quad (7)$$

This constant is used for scaling the accumulation matrix (zero value excluding) to values (1;255).

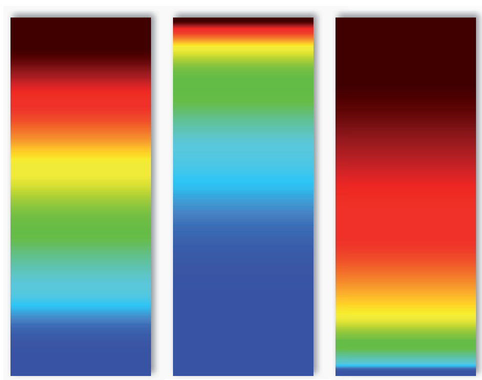
The next stage is the reference image loading (Fig. 4 and 5), which has dimensions 255×100 pixels. Each value in the accumulation matrix (after scaling) corresponds to color in the adequate line in the reference image. Thanks this operations we can visualize people density maps using different references in an easy way.



(a) (b) (c)

Fig. 4. Illustrative color scales for people density visualization

The last step of the algorithm is the overlay of the previously generated people density map and display of the legend on the left hand side on the input image .



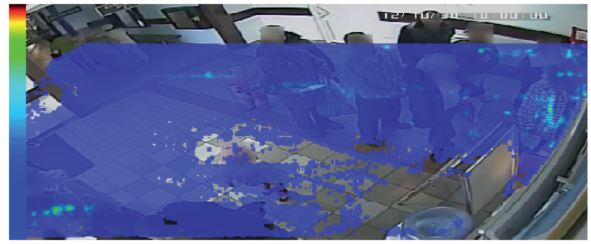
(a) (b) (c)

Fig. 5. Scale intervals used: (a) equal, (b) logarithmic, (c) inverse logarithmic

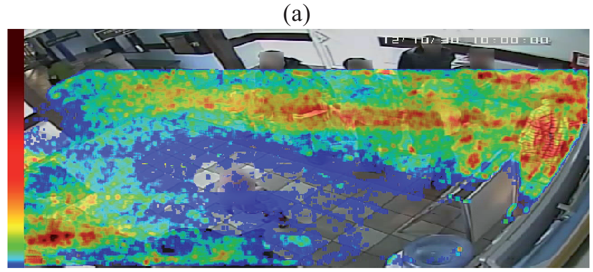
B. People density map visualization

In order to chose the most convenient density map visualization standard (color scale) three sets of reference images in different colors were tested. Fig. 4 shows three typical color scale images with equal intervals. For further research we decided to use the third scale shown in Fig. 4 (c). It illustrated the results in the best way and was the most intuitive. Examples of the use of scale images in the blue-green-yellow-red colors was shown in Fig. 6 and 9.

For the previously chosen reference image, the images were prepared (Fig. 5), in which the color ranges are equal (a), logarithmic (b) and inverse logarithmic (c). Each reference images are used for different types of video sequences. a scale with equal intervals is suitable for short videos (several minutes duration), in which the motion distribution is uniform. In the case of queues or narrow passage ways in short clips we used logarithmic color intervals, while for long sequences (hours or days long) we used inverse logarithmic ranges of colors.



(a)



(b)

Fig. 6. Results of the *Heat Maps Generator* program with the use of: (a) logarithmic and (b) inverse logarithmic scale

An issue of the appropriate choice of the reference image intervals for long video sequences is illustrated in Fig. 6. Especially for needs of protection of personal data, (e.g. faces) the faces in illustrative images are blurred. As shown in Fig. 6 (a) the use of the logarithmic reference image for long video badly affects on the visibility of the maxima, which in this case are located in the bottom left corner of the image. When we use the inverted logarithmic reference image (Fig. 6 (b)), the queue in the long recordings is more visible. Thus, we can accurately determine where the largest people density was.

C. Interface for Heat Maps Generator program

The *Heat Maps Generator* application is equipped with the *graphical user interface* (GUI) as shown in Fig. 7. The program allows to load a video sequence from a file, choose the proper reference map depending on the input video length and size of the color point depending on the input video resolution.

Recordings used during the test were equipped with the timestamp. It was necessary to exclude timestamps from the processing area of the frame. There were also other areas, like e. g. sliding doors, escalators that had to be excluded. Program allows to define the processing area. *Upper timestamp* and *Lower timestamp* options should be used for excluding the appropriate marker. *User-defined area* option enables to exclude any other area from processing. After choosing this option the sample frame is displayed and the user can indicate the excluded area by selecting 10 limiting points (Fig. 8).

The *Generate Heat Maps* button starts the processing. The *Open the last area coordinates* button opens the text file with the coordinates of the last excluded area, which can be used if multiple videos from one localization are tested. The *Open the destination folder* button can be used to open the output density maps destination folder.

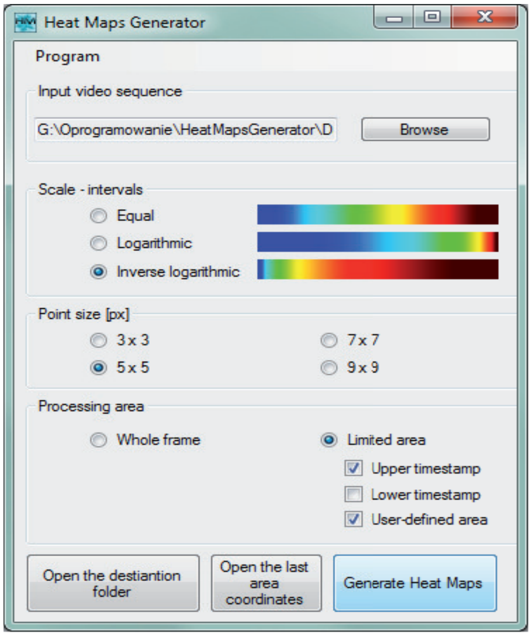


Fig. 7. *Heat Maps Generator* program interface

VI. Results

During the tests 64 hours of video sequences (saved in *H.264* format) from 4 localizations were analyzed by our program with the following options:

- Inverse logarithmic reference image due to long recordings - an hour length.
- Point size: 5×5 pixels, because input image had a 704×288 pixels resolution. In case of choosing a smaller point size, unaesthetic holes may occur in the density maps.
- Limiting the processing area enable options: *upper timestamp* and *user-defined area*. Figure 8 shows an example of using the *Limited user area* option in case of cash registers and customer service points, which should not be taken into account in the determination of the density map.

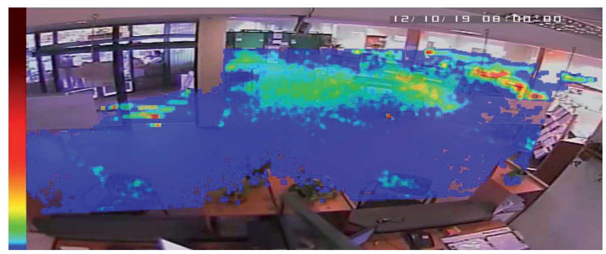
During the tests Intel(R) Core(TM) i7-2620M CPU @ 2,7 GHz was used. The video sequence processing time is two times smaller than the real video time.

Fig. 8. Example of using *Limited user-defined area* option

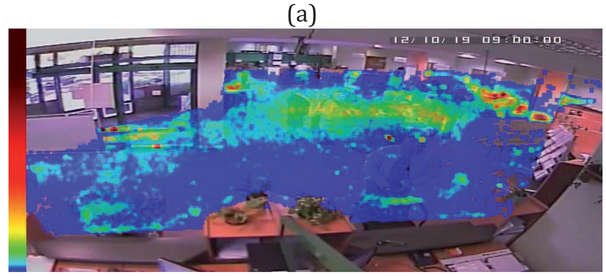
During the tests the video recordings were divided into one-hour long sequences and full-day long sequences. Density maps analysis with the one-hour long videos allows to designate busy hours as we can observe changes in the people distribution in the room during the day: movement of queues, density of people at service points, etc.

Fig. 9 and 10 show the density maps computed for the one-hour long time intervals during the working hours from one day at one location. Thanks to one-hour intervals during tests we could find that the busy hour is between 12:00 and 13:00. At that time the density map shows a long queue of people extended across the room. By means of these results we could quickly and precisely determine, which point of the customer service handled the largest number of people.

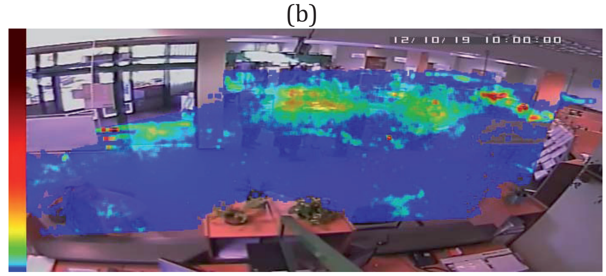
Simultaneously, video sequences have been manually analyzed in terms of the people density and distribution. The authors have prepared summary describing the location and time of increased traffic, the formation of queues and occupancy of customer service points. Comparison of the data obtained during the analysis of watched sequence to automatically generated heat maps has shown a clear convergence of results and same conclusions.



(a)



(b)

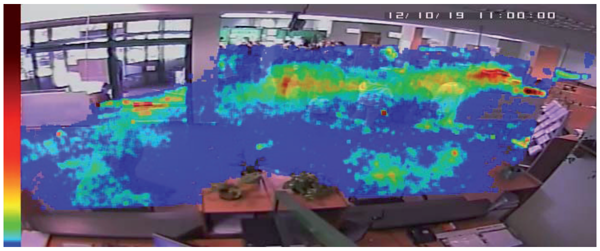


(c)

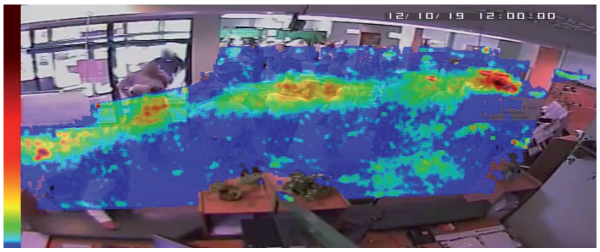
Fig. 9. Results for one-hour long time intervals: (a) 8:00-8:59 am, (b) 9:00-9:59 am, (c) 10:00-10:59 am

VII. Conclusions

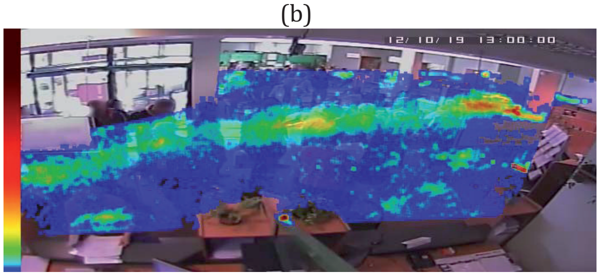
The developed software for generation of the people density maps entirely fulfills expectations defined by the specialists from the consulting company. During the research a synergy effect among specialists with the engineering knowledge related to the imaging and video sequences technologies and people with the marketing competencies proved to be visible and very valuable. A combination of these experiences resulted in obtaining the optimum density maps.



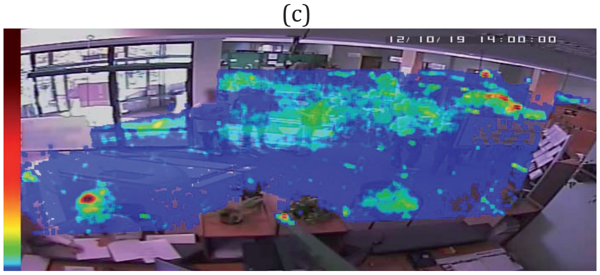
(a)



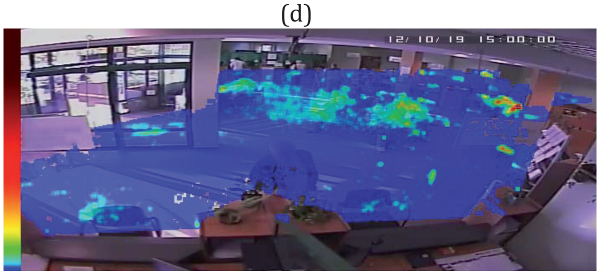
(b)



(c)



(d)



(e)

Fig. 10. Results for one-hour long time intervals: (a) 11:00-11:59 am, (b) 12:00-12:59 pm, (c) 1:00-1:59 pm, (d) 2:00-2:59 pm, (e) 3:00-3:59pm

A helpful element in the implementation of the algorithms and preparing the program was the use of the OpenCV libraries that contain basic functions related to the image processing. The current software version developed by the authors is designed to operate off-line but the processing speed allows to use it also for the on-line applications. This can be used, e. g. in the dynamic allocation at a cash registers and operating stations at the customer service points.

References

[1] MxAnalytics Camera-Integrated Video Analysis, http://www.mobotix.com/eng_AU/Company/News/Recent-News (2013)

[2] Davies A, Yin JH, Velastin S (1995) Crowd monitoring using image processing, Electronics & Communication Engineering Journal, vol. 7(1), pp 37–47

[3] Marana AN, Velastin SA, Costa LF, Lotufo RA (1997) Estimation of crowd density using image processing, Image Processing for Security Applications, Digest No.: 1997/074, pp 11/1–11/8

[4] Marciniak T, Dabrowski A, Chmielewska A, Nowakowski M (2012) Real-Time Bi-Directional People Counting Using Video Blob Analysis, Proc. of IEEE NTAV/SPA 2012, IEEE New Trends in Audio and Video / Signal Processing Algorithms, Architectures, Arrangements and Applications, Poznan, Poland, pp. 161-166

[5] Li L, Huang W, Gu IYH, Tian Q (2003) Foreground object detection from videos containing complex background, Multimedia '03: Proceedings of the eleventh ACM international conference on Multimedia, pp 2–10

[6] Sharif MH, Ihaddadene N, Djeraba C (2008) Crowd behaviour monitoring on the escalator exits, ICCIT 2008, 11th International Conference on Computer and Information Technology, pp 194-200

[7] PN-EN 50132-7 norm: Alarm systems – CCTV surveillance systems for use in security applications – Part 7 (2003)

[8] Lijing Z, Yngli L (2010) Motion human detection based on background subtraction, 2010 Second International Workshop on Education Technology and Computer Science (ETCS), vol.1, pp 284-287

[9] Azab MM, Shedeed HA, Hussein AS (2010) a new technique for background modeling and subtraction for motion detection in real-time videos, 17th IEEE International Conference on Image Processing (ICIP) pp 3453-3456

[10] Piccardi M (2004) Background subtraction techniques: a review, 2004 IEEE International Conference on Systems, Man and Cybernetics, vol.4, pp 3099- 3104

[11] Liu B, Hui X (2010) Object detection based on Gaussian mixture background, 2010 International Conference on Computational Problem-Solving (ICCP), pp 215-217

[12] Yunchu Z, Zize L, Zengguang H, Hongming W, Min T (2005) An adaptive mixture Gaussian background model with online background reconstruction and adjustable foreground mergence time for motion segmentation, IEEE International Conference on Industrial Technology 2005, pp 23- 27

[13] Rui Y, Xuehua S, Shu Y (2009) Moving object detection based on an improved gaussian mixture background model, International Colloquium on Computing, Communication, Control, and Management, vol.1, pp 12-15

[14] Linkai C, Pinwei Z, Guangping Z (2010) Moving objects detection based on background subtraction combined with consecutive frames subtraction, 2010 International Conference on Future Information Technology and Management Engineering, vol.2, pp 545-548

[15] Barron JL, Fleet DJ, Beauchemin SS, Burkitt TA (1992) Performance of optical flow techniques, 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp 236-242

[16] Pathirana PN, Lim AEK, Carminati J, Premaratne M (2007) Simultaneous estimation of optical flow and object state: a modified approach to optical flow calculation, 2007 IEEE International Conference on Networking, Sensing and Control, pp 634-638

[17] Pawlowski P, Borowczyk K, Marciniak T, Dabrowski A (2009) Real-time object tracking using motorized camera, Proc. of IEEE SIGNAL PROCESSING SPA'2009, Signal Processing Algorithms, Architectures, Arrangements and Applications, Poznan, Poland, pp. 110-115

[18] Lei C, Hua Y, Takaki T, Ishii I (2011) Real-time frame-straddling-based optical flow detection, 2011 IEEE International Conference on Robotics and Biomimetics (ROBIO), pp 2447-2452

[19] OpenCV v2.3 documentation, OpenCV API Reference (2011)

The paper has been partly financed with the INDECT project means.