Thermal Imaging for Active Evacuation Routes

Nikolaos Doulamis
National Technical University of Athens
Rural & Survey Engineering
9th Heroon Polytechniou str. Athens
+302107722664
ndoulam@cs.ntua.gr

Anastasios Doulamis
National Technical University of Athens
Rural & Survey Engineering
9th Heroon Polytechniou str. Athens
+302107722676
adoulam@cs.ntua.gr

Konstantinos Makantasis
Technical University of Crete
Production Engineering & Management
Technical University Campus
Chania Crete
+302821037367
konst.makantasis@gmail.com

Konstantinos Karantzalos
National Technical University of Athens
Rural & Survey Engineering
9th Heroon Polytechniou str. Athens
+302107721673
konstantinos.karantzalos@gmail.com

Konstantinos Loupos
Institute of Communication and Computer Systems
Athens, Greece
+30210772 2467
kloupos@iccs.gr

ABSTRACT
Evacuation is a complex process influenced by multiple parameters that have significant impact on the design and execution of an efficient Active Evacuation Route (AER). Computer vision algorithms are critical for an effective AER, since it indicates the current situation awareness of the environment. These tools usually integrated with other pervasive technologies, such as RFIDs, smart spaces for a real time execution of a successful egress. Visual analysis is concentrated into two main directions; micro-scale and macro-scale behavior recognition. Micro-scale refers to detection of individuals dynamics, whereas macro-scale to crowd dynamics properties. In all cases, however, thermal imaging is an alternative effective computer vision mechanisms for the analysis of the crowd behavior either at the micro or macro scale. Thermal imaging allows efficient determination of people from the background even if highly dynamic scenes, illumination, occlusions or content alterations. However, visual analysis from thermal cameras presents new research challenges mainly due to the fact that thermal cameras are of floating point precision. In this paper, we provide an image segmentation method from the background using thermal visual data. Then, these data are used for evacuation purposes.

Categories and Subject Descriptors
I.4 **[IMAGE PROCESSING AND COMPUTER VISIO]:** Segmentation – pixel classification, scene analysis- tracking. I5 **[Pattern Recognition]:** Models-statistical.

General Terms
Algorithms, Performance.

Keywords
Tracking, Thermal imaging, evacuation, behavior analysis.

1. INTRODUCTION
The evacuation is a complex process influenced by multiple parameters that have significant impact on the design and execution of a successful egress. This renders the whole evacuation process inherently complex. Thus, design and adoption of effective plans that should be executed so as to achieve a safe evacuation are very challenging tasks for the responders (e.g. civil protection agencies). For example, there is a large number of interactions between people that are in a state of egress, (such as people-to-people/people-to-building physical forces, crowd density and friction etc) whose impact on the overall process is extremely hard to be assessed [1] [2].

In real life situations, it is usually the responsibility of specific commanding persons to comprehend the status of an incident and to give the appropriate orders for the execution of an evacuation process. However, the efficient simultaneous real-time perception of all relevant factors that may influence such a venture is extremely complex and relies mainly on the experience and training of the person responsible for the coordination of the evacuation actions. Moreover, specific human response under severe stress imposed by certain situations can easily lead to inefficient/wrong perceptions of the situation and subsequently inefficient or even fatal decisions [3]. To showcase this, statistically, the average human, during an evacuation process, a person may select the fastest path for getting out of the building or head back to the place from which they entered the building. However, such a collective decision may lead to blockage of the emergency exit becoming a bottleneck and thus driving to inefficient evacuation process.

Computer vision tools are critical for an efficient active evacuation process, since it indicates the current situation awareness of the environment. Crowd flows are monitored using sophisticated cognitive surveillance tools operating in multi-camera configurations able to learn from the monitored environment and adapt their monitoring process to the ever-changing conditions characterizing people-dense situations. By combining cognitive surveillance tools and crowd dynamics models we are able to apply: a) reasoning methods able to learn isolated and multiple events and b) stochastic prediction methods to estimate the evolution of such events. Such estimates of crowd flows/states evolution will be used for suggesting critical-time effective plans for safe evacuation.

Computer vision toolkits are incorporated with other pervasive tools such as 3D audio sensors, RFIDS and smart-spaces for providing an efficient Active Evacuation Route (AER). Crowd understanding the likely psychology of crowd engagement (or collective behaviour) is performed at various types of mass-

gathering types and events. This will be necessary to adopt for providing more context-awareness to multi-scale crowd behaviour recognition tools with deep learning capabilities. Furthermore, the prediction of the psychology of crowd engagement will enable better preparedness and planning by emergency services for adopting rapid and strategic crowd spatial evacuation from dangerous situations to safer areas [4].

Micro-scale Behaviour Recognition: The analysis of crowd is concerned with individuals or smaller leading groups within a crowd ("seeds"). Such analysis consists of understanding such individuals or group interactions with the crowd; group formation, split or merger of individuals; and event detection for groups. The usual approach to solving these processes consists of the following tasks which lead to seeds behaviour recognition [5]: object-detection, object-tracking, object-classification and behaviour-recognition.

The analysis of seeds behaviour in a crowd relies predominantly on their detection and tracking [6], [7]. Once detected, their relative and cumulative motion characteristics may represent the overall state of crowd behaviour when in relatively less dense conditions. However, this will not be robust in the behaviour analysis of highly crowded or complex scenes. eVACUATE will overcome such challenge by adopting optical flow and stochastic approaches using Scale Invariant Feature Transform techniques (SIFT) [8]. These follow the generic tasks for seeds behaviour recognition in densely crowded scenes: motion-description, scene-segmentation, and behavior recognition.

The optical flow and stochastic methods, when put under a generic behaviour analysis framework, shall defer from the macroscopic analysis of crowd too [9]. In such a case the analysis of motion and behavior of crowd is based on the aggregation and accumulation of its inner parts. The aim is to have an organic reconstruction of recognised behaviour of the seeds at micro-scale and leading to crowd behaviour recognition at macro-scale. Furthermore, this leads to the enablement of a generalised learning process with relevant feedback on seeds abnormal behaviour recognition under various crowd types and emergency situations [10].

Macroscale Crowd Behaviour Recognition: Crowds from a distance exhibit a variety of high level fluid flow motions, which include both steady and turbulent flows. These flows are motions between large scale groups acting together within a crowd [11]. This type of behaviour is defined as macroscopic behaviour of the crowd. The techniques employed to study and model macro-scale crowd behaviour depend on a number of factors such crowd density, crowd motion (velocity) and dominant direction), environment (presence of bottle-necks and no go areas) etc.

However, from the perspective of computer vision research, there are two classes of approaches that are employed [12]. The first class is to use techniques to isolate individuals, track them and estimate the crowd density and typical motions. The second class is ignoring the individuals and modelling the crowd motion by using fluid flow theory [13].

The selection of crowd density estimation algorithms depends on the number of people in the scene, presence of occlusions and whether crowd is stationary or moving. In uncluttered scene with not high density of crowd, the most suitable approaches for density estimation are based on counting the individuals directly [14]. Chebychev moments or gray level dependency matrix are used. However, texture-based approaches do not take into consideration temporal information of the crowd. Therefore, for moving crowd dynamic texture analysis method based on the sparse spatiotemporal local binary descriptor is selected for density estimation [15].

This paper presents a method for analysis of thermal imaging as regards human tracking and localization purposes. The paper introduces a statistical model that allows localization and tracking of humans from thermal visual content. Then, the results are exploited for evacuation purposes. For this reason, human detection are projected onto orthogonal 2D maps in order to be determined crowd movement parameters such as density and direction. These parameters are used for an efficient evacuation policy in a real time constrained environment.

2. Thermal Vision

In the framework of this paper thermal imagery is used for individuals detection. The task of detection can be thought as an image segmentation task, which aims to divide an image into regions belonging to two disjoint classes corresponding to objects of interest and background, respectively. Image segmentation can be handled through background subtraction / foreground extraction techniques.

Background subtraction consists of a major pre-processing step in many computer vision applications and it aims to identify moving objects from the portion of a video frame that differs significantly from the background model. In general, there are many challenges in developing a good background subtraction technique. The algorithm must

1. be robust in dynamically changing background conditions, e.g. illumination changes
2. avoid detecting as foreground non stationary background objects, e.g. swinging leaves
3. react quickly to changes in background and update its internal background model.

Although there are a lot of background subtraction techniques in the literature [16]- [18] concerning RGB video (a review can be found in [19]) that corresponds to visual spectrum, this problem was not extensively studied for thermal imagery. Unfortunately, when the goal is to extract the foreground from a thermal video stream, background subtraction techniques that has been applied on RGB data and perform well cannot be applied, out of the shelf, on thermal data too.

In order to develop a robust background subtraction algorithm appropriate for application on thermal imagery, the peculiarities of thermal data must be taken into consideration. To be more specific, RGB data that correspond to visual spectrum imagery, usually consist of three channels, each one contains integer values ranging from zero to 255. Furthermore, conventional RBG sensors can produce high resolution images with low Signal-to-Noise ratio. In contrast, thermal sensors produce single channel images that contain floating-point elements (pixels) and due to low resolution infrared sensor the produced data present high Signal-to-Noise ratio. The differences between RGB data correspond to visual spectrum imagery and thermal data are summarized in the Table 1..

Table 1: Differences between thermal and RGB data

	# of channels	pixel values	SNR	Resolution
RGB visual spectrum	3	Integers	Low	High
Thermal imagery	1	Floating points	High	Low

Although, RGB sensors seem to produce higher quality data, it is not true. Thermal imagery presents some properties that are not present in data that correspond to visual spectrum. These properties can add value to thermal data. Concretely, due to the fact that thermal video cameras detect the amount of thermal radiation emitted/reflected from objects in the scene, produced thermal data can be used for both day and night scenarios. Therefore, they are a prime candidate for a persistent (24-7) video system for surveillance and monitoring. As long as the thermal properties of a person are slightly different (higher or lower) from the background radiation, the person region is detectable in thermal imagery. Furthermore, shadows appear smoother in thermal imagery unless the person.

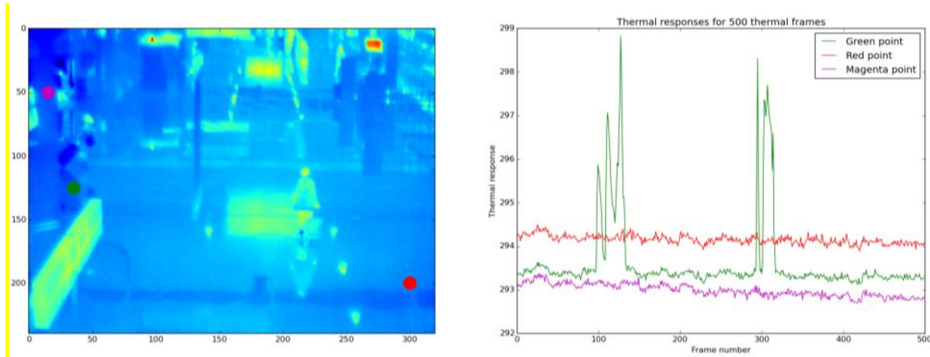


Figure 1: Responses of three different pixels to infrared spectrum

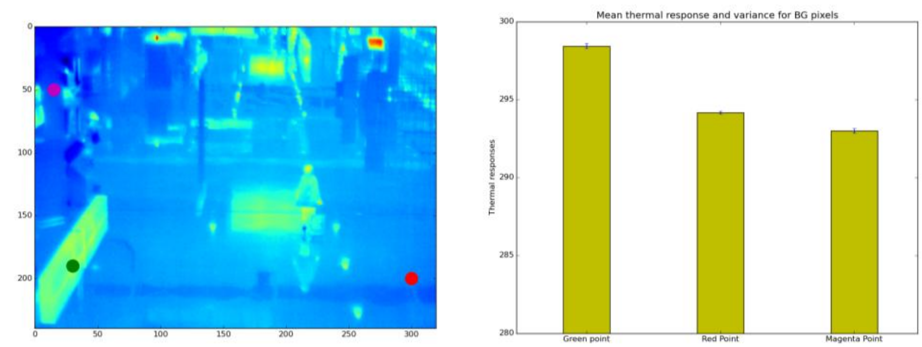


Figure 2: The mean infrared responses along with their variances of three different pixels belonging to the background

2.1 Statistical Analysis

The background model on the one hand must be robust i.e. be able to successfully describe responses of pixels belonging to the background and on the other it must be dynamic i.e. be able to automatically adapt to dynamically changing conditions. Furthermore, background model must be estimated in a fast and cost effective way in order to fulfil requirement for real-time operations.

This suggests that background subtraction must take place using “simple” and lightweight techniques that produced good results and avoid sophisticated and computational intensive algorithms. As shown in [20], while complicated techniques often produce superior performance, simple techniques such as adaptive mean filtering can produce good results with much lower computational complexity. In order to develop a method for estimating the background model, firstly the statistical properties of pixels, concerning their responses to infrared spectrum, are examined.

Figure 1 resents the responses of three different pixels to infrared spectrum. Their responses for 500 frames range from 292.8-Kelvin degrees (19.65-Celsius degrees) to 298.7 Kelvin degrees (25.55-Celsius degrees). During 500 frames, the red and magenta pixels correspond to background i.e. during the frames no foreground object passed through those pixels. In contrast, during the same frames, the green pixel corresponds mainly to background but during sort time periods foreground objects are passing through this pixel. The responses of red and magenta pixels to infrared spectrum present small variations around an average value. These variations are mainly caused by the high Signal-to-Noise ratio. Regarding the green pixel, it is easy to be seen that when a foreground object is passing through this pixel the infrared responses are greatly varying due to the large temperature difference between the background and the foreground objects (foreground objects correspond to individuals are much hotter than the background). Furthermore, in Figure 2 the mean infrared responses along with their variances of three different pixels belonging to the background are presented. Examining this figure we can conclude that the background model for each one of the image pixels can be estimated using a unimodal distribution described by its mode and variance.

In order to create a dynamic background model, the estimated unimodal distribution must be adapted to dynamically changing background conditions. For fulfilling this requirement we developed and tested three different background subtraction algorithms, which can automatically re-estimate parameters that control the background model. Namely, these algorithms are: i) background subtraction using running average, ii) statistical background modeling and iii) background subtraction using mixtures of Gaussian distributions. These algorithms are analytically presented in the next section.

3. Image Segmentation through Background Modeling

The running average (RA) algorithm includes two different steps: the differencing step and the background-modelling step. During the differencing step, the algorithm extracts the foreground objects by computing the difference between current frame and the background model, while during the background modelling the model of background is estimated using single modality assumption.

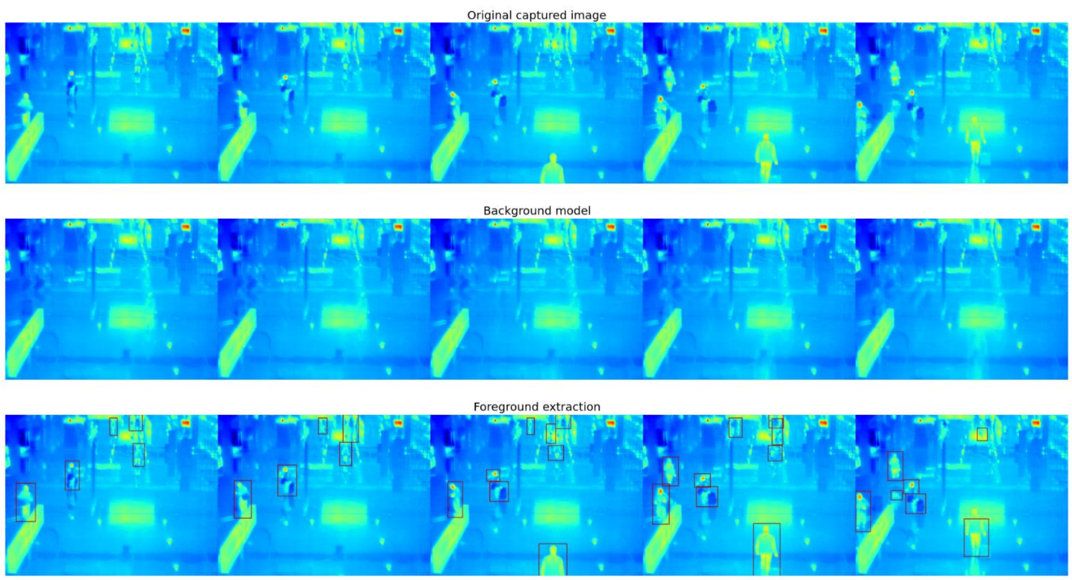
The following relation can describe the differencing step

$$M^{t+1}(x,y) = |I^{t+1}(x,y) - BG^{t}(x,y)| \qquad (1)$$

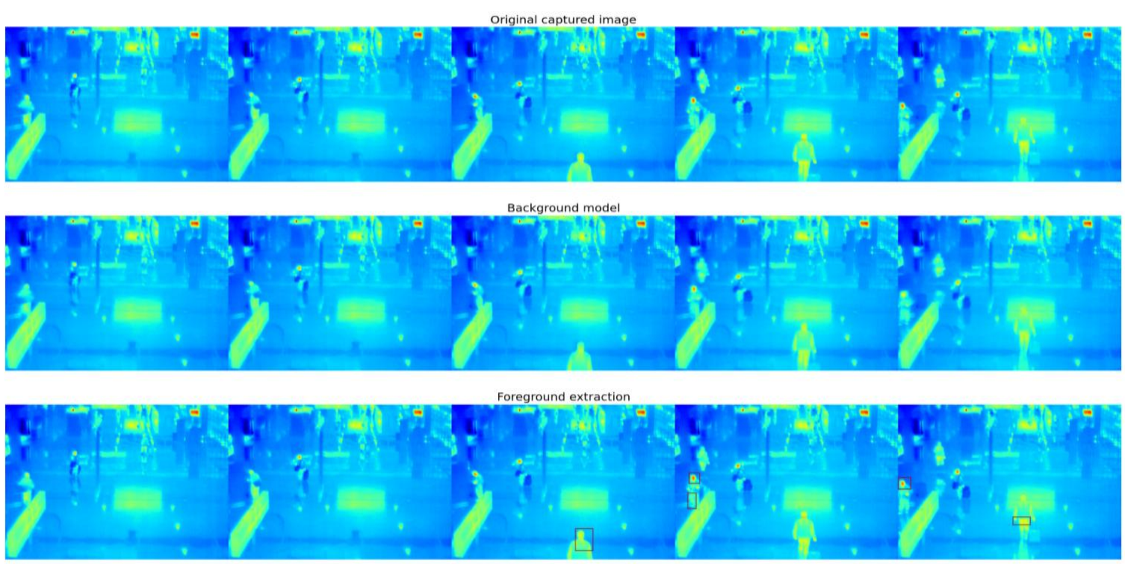
where $I^{t+1}(x,y)$, $BG^{t}(x,y)$ and $M^{t+1}(x,y)$ stand for the pixel value of current frame at time *t+1*, the background model at time *t* and the subtraction result respectively. The background model is assumed to be a unimodal distribution, which mode value is recursively updated by new captured frames. Base on this assumption computational efficiency is improved while at the same time memory resource allocation is reduced. The recursive form of the mode value for the unimodal distribution, $BG^{t}(x,y)$, is described as

$$BG^{t}(x,y) = (1-a)BG^{t-1}(x,y) + aI^{t} \qquad (2)$$

parameter a corresponds to the learning parameter and controls the adaptation speed of the background model. In order to extract the foreground objects the foreground mask FG^{t+1} is calculated



α = 0.013



α = 0.5

Figure 3: Performance of Running Average (RA) using different values for parameter α

by thresholding the difference image $M^{(t+1)}$using the following relation

$$FG^{(t)}(x,y) = \begin{cases} 1, & if M^{(t+1)}(x,y) > t \\ 0, & otherwise \end{cases} \quad (3)$$

Flow diagram in Figure 3 presents the process for estimating the background model, while **Error! Reference source not found.** visually presents the performance of RA when applied on thermal data with different values for the parameter a.

3.1.1 Statistical background modelling

Statistical background modeling aims to estimate a background mean and variance model for each one of image pixels. For doing so, this algorithm uses a history of N frames. As foreground objects could be present in the history of frames, the statistical background model for each pixel is created by computing weighted means and variances from the N frames. Mean and variance values for a pixel located at *(x,y)* are given by the following relations:

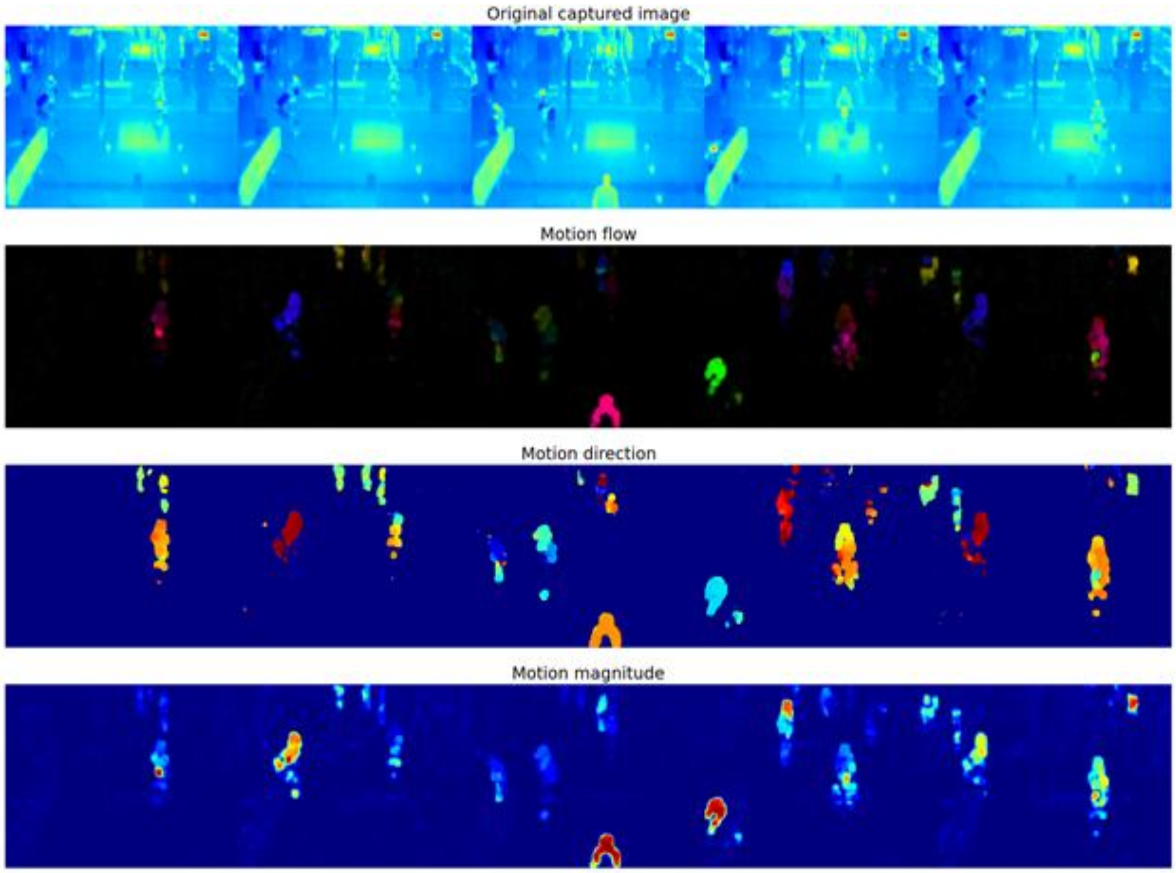


Figure 4: Performance of Farneback's method

As can be seen in Figure 3 when the value of α is small the background model is being adapted slower to background changes. When the value of α is getting larger the background model is being adapted faster, and results to a foreground objects mask that contains a lot of false negatives.

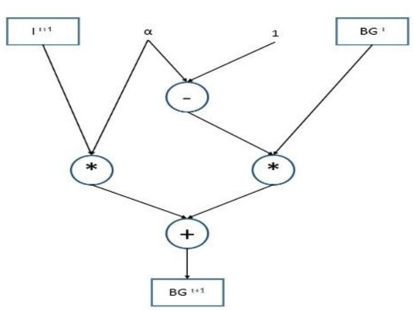


Figure 4: Flow diagram of the process for estimating the background model

$$\mu(x,y) = \frac{\sum_{i=1}^{N} w_i(x,y) \cdot I_i(x,y)}{\sum_{i=1}^{N} w_i(x,y)} \quad (4)$$

$$\sigma^2(x,y) = \frac{\sum_{i=1}^{N} w_i(x,y) \cdot (I_i(x,y) - \mu(x,y))^2}{\frac{N-1}{N}\sum_{i=1}^{N} w_i(x,y)} \quad (5)$$

where the weights $w_i(x,y)$for each pixel location are used to minimize the effect of outliers, which in our case are the pixels belonging to the foreground objects. For estimating the weights a representative image, I_{rep}, for the history of the N frames is used. As representative image, the median or the the mean image of the N frames is used. Having estimate the representative image, the weight for a pixel located at *(x,y)* is computed from a Gaussian distribution centered at $I_{rep}(x,y)$

$$w_i(x,y) = \exp\left(\frac{\left(I_i(x,y) - I_{rel}(x,y)\right)^2}{-2\hat{\sigma}^2}\right) \quad (6)$$

the parameter $\hat{\sigma}$ is set by the user. During experiments we set this parameter to the value of 5.

frame (timestamp)	seeds_coordinates (X,Y)		A_density	A_velocity	A_direction	B_density	B_velocity	B_direction	C_density	C_velocity	C_direction	D_density	D_velocity	D_direction
1	37.93648529	23.94783401	0	0	0	0	0	0	2	0.34461835	0.21659837	5	0.13238002	0.15880385
	37.93656158	23.9480381												
	37.93657303	23.94801903												
	37.93658066	23.9479866												
	37.93658066	23.94784164												
	37.9365921	23.94792557												
	37.9365921	23.94811249												

frame (timestamp)	seeds_coordinates (X,Y)		A_density	A_velocity	A_direction	B_density	B_velocity	B_direction	C_density	C_velocity	C_direction	D_density	D_velocity	D_direction
2	37.93648529	23.94784164	0	0	0	0	0	0	1	0.71661093	0.19067035	5	0.17998885	0.16901365
	37.93656158	23.9480381												
	37.93657303	23.94802094												
	37.93658066	23.94799042												
	37.9365921	23.94792557												
	37.9365921	23.94811249												

frame (timestamp)	seeds_coordinates (X,Y)		A_density	A_velocity	A_direction	B_density	B_velocity	B_direction	C_density	C_velocity	C_direction	D_density	D_velocity	D_direction
3	37.93647003	23.94817352	0	0	0	0	0	0	2	0.20972238	0.33227638	4	0.25706368	0.28815566
	37.93648529	23.94782448												
	37.93657303	23.94801712												
	37.93658066	23.9480114												
	37.9365921	23.94792747												
	37.9365921	23.94811249												

frame (timestamp)	seeds_coordinates (X,Y)		A_density	A_velocity	A_direction	B_density	B_velocity	B_direction	C_density	C_velocity	C_direction	D_density	D_velocity	D_direction
4	37.93648529	23.94782639	0	0	0	0	0	0	1	0.47611763	0.27458333	5	0.25738105	0.54651199
	37.93657303	23.94801903												
	37.93658066	23.9480114												
	37.93658829	23.94791031												
	37.9365921	23.94811249												
	37.9365921	23.94799995												

frame (timestamp)	seeds_coordinates (X,Y)		A_density	A_velocity	A_direction	B_density	B_velocity	B_direction	C_density	C_velocity	C_direction	D_density	D_velocity	D_direction
5	37.93647003	23.94817352	0	0	0	0	0	0	2	0.31935793	0.40891586	5	0.23730318	0.45616263
	37.93648529	23.94782639												
	37.93657303	23.94801903												
	37.93657684	23.94801331												
	37.93658829	23.94791031												
	37.93659592	23.94811249												
	37.9365921	23.94800186												

Figure 5: The extracted information regarding the detected people (seeds), crowd density, velocity and direction in the four different sub-regions of the test site.

Once the background model has been estimated, the foreground pixels for a new input image can be obtained using the squared Mahalanobis distance

$$D(x,y) = \begin{cases} 1, & \dfrac{(I_i(x,y) - \mu(x,y))^2}{\sigma(x,y)^2} > T \\ 0, & otherwise \end{cases} \quad (7)$$

During experiments, we set T parameter equal to 5. Figure 3 visually presents the performance of statistical background modeling when applied on thermal data with different values for the parameter T. As shown in Figure 3 when the value of T is small the output of the algorithm contain a lot of false positive detections of foreground objects. Contrary, when the value of T is large the output of the algorithm contains a lot of false negative detections.

3.1.2 Background subtraction using mixtures of Gaussian distributions

Background modeling using Gaussians Mixture Model [9] is one of the most common used techniques for subtracting the background of RGB videos, because it is i) fast, ii) robust to small periodic movements of background and iii) easy to parameter. For these reasons, we used it with thermal videos and evaluated its performance, although it is not using a unimodal distribution to model pixels' responses.

The goal of Mixtures of Gaussians (MOG) background modeling is to decide if a pixel $x^{(t)}$at time t, belongs to background (BG) or to foreground (FG). A reliable decision can be made if the probability of pixel $x^{(t)}$to belong th BG is bigger than a threshold. This can be expressed as

$$p(x^{(t)} \mid BG) > c_{thr} \quad (8)$$

The probability $p(x^{(t)} \mid BG)$consists the background model, which can be estimated via a training set $X_T = \{x^{(t)}, x^{(t-1)}, x^{(t-T)}\}$defined over a time span T. The estimated background model is denoted as $\hat{p}(x^{(t)}|X_t, BG)$. For each new pixel sample the training set X_T is updated and $\hat{p}(x^{(t)}|X_t, BG)$ is re-estimated. Fitting the data of X_T using a mixture model of M Gaussian component we have:

$$\hat{p}(\, x^{(t)} \mid X_T, BG + FG \,) = \sum_{m=1}^{M} \hat{\pi}_m \; N(x^{(t)}; \hat{\mu}_m, \hat{\sigma}_m^2 I)s \quad (9)$$

where $\hat{\mu}_1 \ldots \hat{\mu}_M$ are the estimates of the means and $\hat{\sigma}_1 \ \ldots \hat{\sigma}_M$ are the estimates of the variances that describe the Gaussian components. The non-negative mixing weights are denoted as $\hat{\pi}_m$ and sum up to one. In conditional probability, *BG*+*FG* is used because among the samples of X_T could be some values that belong to the foreground objects. Due to the fact that the samples of X_T that correspond to background consist the majority in the set

and their values change gradually, Gaussian components associated with background samples will present larger mixing weights compared to the samples that are associated with foreground samples. Therefore, the background model can be approximated as:

$$\hat{p}\left(x^{(t)} \mid X_T, BG \right) \sim \sum_{m=1}^{B} \hat{\pi}_m \; N\left(x^{(t)}; \hat{\mu}_m, \hat{\sigma}_m^2 I\right) \qquad 10)$$

If the components are sorted to have descending weights, then B is estimated as

$$B = \arg\min_b \left(\sum_{m=1}^{b} \hat{\pi}_m > \left(1 - c_f\right) \right) \qquad (11)$$

where c_f is a measure of the maximum portion of the data that can belong to foreground objects without influencing the background model.

3.1.3 Optical flow techniques – Dense and sparse

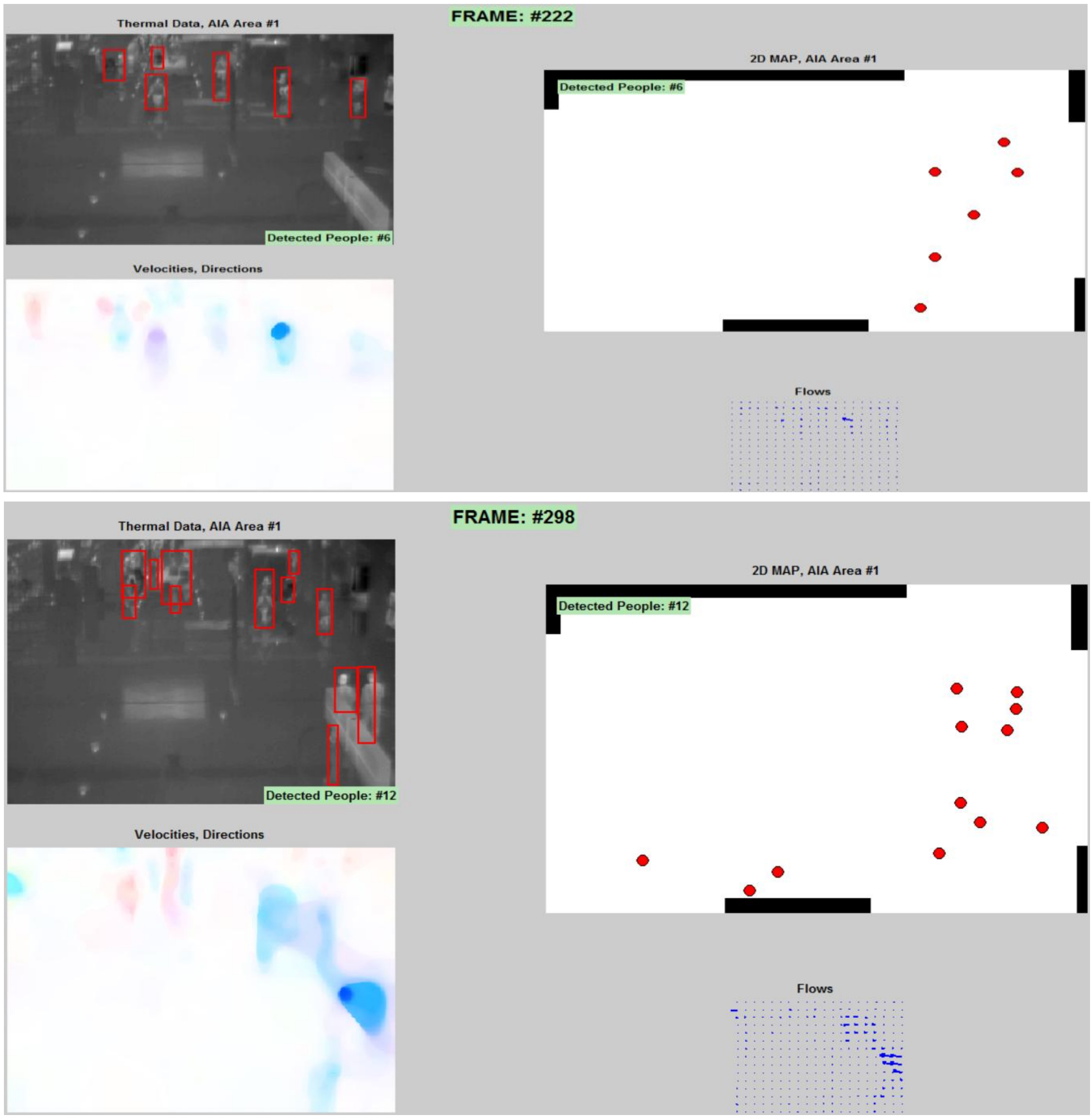


Figure 6: Experimental results after the application of the developed algorithm for people detection in thermal video sequences.

Another common for estimating motion and thus detecting moving objects in a scene is to use optical flow. Optical flow techniques are divided into two different categories: sparse and dense. Sparse techniques detect a few almost "unique" points, or good features to track according Shi and Tomasi, in an image and then track these points from frame to frame. In contrast dense optical flow techniques detect the presence of motion for each one of the image's pixels. Although, dense techniques are more computationally intensive than dense, they produce more accurate results. Regarding thermal imagery, due to low resolution images and high Signal-to-Noise ratio sparse techniques con not be used as it is very difficult to detect good features to track.

In the framework of this work package we used a dense optical flow technique, namely Farneback's algorithm. Farneback algorithm uses Polynomial Expansion to approximate the neighbors of a pixel. The Expansion could be seen as a quadratic equation with matrices and vectors as variable and coefficients. This dense optical flow analysis produces a displacement field from two successive video frames. Each displacement vector in the field is estimated by solving a minimization problem subject to constraints derived from the polynomial expansion. The error during minimization is the weighted sum of differences in the pixel neighborhood between the images. Image Pyramids is used to detect large displacements and Gaussian filter to smooth out the neighboring displacements.

Figure 4 visually presents the performance of Farneback's method when applied on thermal data. The top row presents the original captured frames, the second row presents the presence of dense optical flow in the scene, which is a fusion of motion intensity and motion direction. The third row presents the direction of motion. Different colours represent different directions, and finally the last row represents the intensity of the motion. The blue colour correspond to no motion, while the red colour represents motion of high intensity.

4. Experimental results

The detected information in every video frame consisted of the number of the detected people (seeds), the crowd density, velocity and direction at each sub-region of the test site. (see Figure 5).Regarding the detection of people (seeds) the preliminary experimental results on the thermal dataset indicated that the background estimation procedure can ease significantly the detection and at the same time decrease the complexity of the required detection algorithms. Experimental results are shown in Figure 6.

5. ACKNOWLEDGMENTS

This paper is supported by the 3D ORO Projected "Pervasive 3D Computer Vision for increasing the efficiency of 3D Digitalization" funded under PAVET Programme of the Greek Secretary of Research and Technology and the "eVacuate: A holistic scenario –independent, situation-awareness and guidance System for sustaining the Active Evacuation Route for large Crowds" European Union funded projected under security call.

6. REFERENCES

[1] J. Boin, Arjen and A. McConnel, “Preparing for Critical Infrastructure Breakdowns”, Journal of Contingencies and Crisis Management, 15, 1 pp. 1-4, 2007.

[2] S. Rose-Pehrsson, J. Owrutsky, S. Wales, F.W. Williams, J.P. Farley, D. Steinhurst and C. Minor, D.T. Gottuk, J. Lynch, “ Multi-Sensory approach to improved situational awareness,” 14th International Conference On-Site Analysis & Homeland Security, 2006, Arlington, Virginia U.S.A.

[3] R.L. Hughes, “The flow of human crowds,” Annual Rev. Fluid Mech., Vol. 35, pp. 169-183, 2003.

[4] A holistic scenario –independent, situation-awareness and guidance System for sustaining the Active Evacuation Route for large Crowds, European Integrated Project, [on line available at] http://www.evacuate.eu/.

[5] G. J. Brostow, R. Cipolla, “Unsupervised Bayesian Detection of Independent Motion in Crowds,” CVPR, 2006.

[6] M. C. Chang, N. Krahnstoever, S. Lim, T. Yu, “Group Level Activity Recognition in Crowded Environments across Multiple,” IEEE 7th Conference on Advanced Video and Signal-based Surveillance, 2010, Boston, USA.

[7] W. Ge, W. Ge, R. T. Collins, R. B. Ruback, “Vision-based Analysis of Small Groups in Pedestrian Crowds,” IEEE Transactions on Pattern Analysis and Machine Intelligence, to appear, on-line available, 2011.

[8] C. Lalos, A. Voulodimos, A. Doulamis, and T. Varvarigou, “Efficient Tracking using a Robust Estimation Technique,” Multim. Tools and Applications, Springer, pp. 1-16, 2012.

[9] D.-Y. Chen, P.-C. Huang, “Motion-based Unsual Event Detection in Human Crowds,” Journal of Image Communication and Image Representation, Vol. 22, No. 2, pp. 178-186, 2011.

[10] Doulamis, Nikolaos, and Anastasios Doulamis. "Evaluation of relevance feedback schemes in content-based in retrieval systems." Signal Processing: Image Communication 21.4 (2006): 334-357.

[11] V. Coscia and C. Canavesio, “First order macroscopic modelling of human crowds,” Math. Models Methods Appl. Sci., Vol. 18, pp. 1217-1247, 2008.

[12] A. Marana, S. Velastin, L. Costa, R. Lotufo, “Automatic Estimation of Crowd Density using Texture,” Safety Science, Vol. 28, pp. 165-175, 1998.

[13] S. Pellegrini, A. Ess, K. Schindler, L. van Gool, “You’ll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking,” IEEE International Conference on Computer Vision, Japan, 2009.

[14] S. Saxena, “Crowd Behavior Recognition for Video Surveillance,” Computer, pp. 1-12, 2008.

[15] Hua Yang, Hang Su, Shibao Zheng, Sha Wei, Yawen Fan, “The large-scale crowd density estimation based on sparse spatiotemporal local binary pattern,” IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, Barcelona, Spain, July, 2011.

[16] A. Elgammal, D. Harwood, and L. Davis, “Non-parametric Model for Background Subtraction,” in Computer Vision — ECCV 2000, Springer Berlin Heidelberg, 2000, pp. 751–767.

[17] D.-S. Lee, “Effective Gaussian mixture learning for video background subtraction,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 5, pp. 827–832, 2005.

[18] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, 2004, vol. 2, pp. 28–31 Vol.2.

[19] M. Piccardi, “Background subtraction techniques: a review,” in 2004 IEEE International Conference on Systems, Man and Cybernetics, 2004, vol. 4, pp. 3099–3104 vol.4.

[20] S. S. Cheung and C. Kamath, “Robust techniques for background subtraction in urban traffic video,” 2004, vol. 5308, pp. 881–892.