UDK 621.313.33:519.876.5

Antoschuk S., ScD.,
Brovkov V. PhD.,
Kovalenko N.

DISTRIBUTED MULTI–VIDEOCAMS SECURITY SYSTEM NOT INFRINGING THE PRIVATE LIFE RIGHTS

Abstract. *The article exposes a wireless distributed video-surveillance networks for efficient human behaviour recognition using cameras with and without controlled areas overlapping. A.м. system application would allow effective solution of existing dilemma between the need in increasing security systems' level through video-surveillance and the individual confidentiality rights to be ensured.*

Keywords: *security systems, individual confidentiality right, videocontext automated evaluation, behaviour recognition, video camera, distributed systems, Gaussian distribution, movements' segmentation, motion detection, background image, Kalman filter*

Антощук С., д-р техн. наук,
Бровков В., канд. техн. наук,
Коваленко Н.

РАСПРЕДЕЛЕННАЯ МУЛЬТИКАМЕРНАЯ СИСТЕМА БЕЗОПАСНОСТИ БЕЗ НАРУШЕНИЯ ПРАВА ЛИЧНОСТИ НА КОНФИДЕНЦИАЛЬНОСТЬ

Аннотация. *Представлена беспроводная распределенная сеть видеонаблюдения для эффективного распознавания человеческого поведения с помощью камер с перекрытием или без такового. Применение представляемой системы позволит разрешить противоречие между необходимостью повышения эффективности систем безопасности на основе видеонаблюдения и обеспечением права человека на конфиденциальность.*

Ключевые слова: *системы безопасности, право человека на конфиденциальность, автоматизированная оценка видеоконтекста, распознавание поведения, видеокамера, распределенные системы, гауссовское распределение, сегментация движения, отслеживание движения, фоновое изображение, фильтр Калмана*

Антощук С., д-р техн. наук,
Бровков В., канд. техн. наук,
Коваленко М.

РОЗПОДІЛЕНА МУЛЬТИКАМЕРНА СИСТЕМА БЕЗПЕКИ БЕЗ ПОРУШЕННЯ ПРАВА ОСОБИ НА КОНФІДЕНЦІЙНІСТЬ

Анотація. *Наведено безпровідну розподілену мережу відеоспостереження для ефективного розпізнавання поведінки людей за допомогою камер з перекриттям або без нього. Застосування такої системи дає змогу усунути протиріччя між потребою підвищення ефективності систем безпеки на основі відеоспостереження і забезпеченням права людини на конфіденційність.*

Ключові слова: *системи безпеки, право людини на на конфіденційність, автоматизована оцінка відеоконтексту, розпізнавання поведінки, відеокамера, розподілені системи, гаусівський розподіл, сегментація руху , відслідковування руху, фонове зображення , фільтр Калмана*

The dilemma between the need in increasing security systems' level and the requirement of individual confidentiality rights to be ensured is exists nowadays. For that reason, in EU countries is practiced the method when at subway and railway stations , in airports the surveillance systems are activated by clicking the "alarm" button, that significantly reduces the effectiveness of security system, still being consistent with the principles of non-interference in the private life of citizens.

To resolve the contradictions it is advisable to create the intelligent video surveillance systems, processing the video stream and identifying the conflict situations in real time by the built-in video-surveillance system function, thus eliminating the need for transmission, distribution and storage of video data and thereby ensuring compliance with the privacy principles.

The separate autonomous video surveillance device's for credible recognition capabilities are limited due to a number of factors: the camera view field, view angle, system resources.

© Antoschuk S., Brovkov V., Kovalenko N., 2013

These factors imposed limitations can be significantly reduced through establishing a network of intellectual video surveillance devices interaction, as well as through creation of distributed security control systems.

At the system alike the cameras are exchanging not with the video stream but with reports on possible (probable) public order breach.

The cameras being adjusted to different vision field and respectively different angles for observed area control is possible to analyze the situation associated with unlawful acts of individuals by pooling data from several (numerous) network elements. This approach opens the possibility of constructing adaptive algorithms for individual nodes as well as for the network thus increasing the efficiency of the whole system.

This allows to resolve the problem of entity's motion identification across cameras in the environment.

One more problem herein relates to the increased flow of raw video data to be processed by the node. However, in the last few years a new type of mini computers appeared, having a considerable processing power and a large RAM size. Applying a.m. hardware in tandem with surveillance cameras allows the video data processing on the fly, which results activation of emergency messages only for the server.

In this paper, we present a wireless distributed video surveillance network for efficient human behavior recognition using cameras with overlapping or non-overlapping fields of view.

As first step we consider a single-camera scenario, after which we work up to the multi-camera system.

Single-camera surveillance

A single-camera surveillance system consists of two main subsystems: preprocessing subsystem, which includes segmentation and tracking components, and a recognition subsystem, comprised of an object type classification and a behavior recognition components.

Movement segmentation and tracking

Moving object detection and segmentation is carried out using a basic background subtraction algorithm. The background frame is constructed by statistical modeling of pixel intensity probability densities, using a mixture of Gaussians [2].

The background image is modeled using three Gaussian distributions. The first Gaussian distribution shows the persistent pixels and represents the background image. The repetitive variations and the relative static regions are updated to the second Gaussian distribution. The third Gaussian represents the pixels with quick changes.

In order to predict the location of the segmented object in the next frame, matching the objects in the consequent frames and resolving the object occlusion problem, we use a Kalman filter. This results in the object's track throughout the camera's field of vision being received (Fig.1).

a

b

Fig. 1. Results of movement segmentation (a) and tracking (b)

For the following use in high-level analysis and behavior recognition, visual descriptors are computed based on the segmentation and tracking results [7, 10]:

- *cvSpeed*: the current speed of an object, an estimated displacement of the object's bounding box expressed in pixels per second,

- *cvFlow*: a recent history of the object's motion,
- *cvLifetime: whether or not the object has been newly instantiated or is about to be terminated due to no longer being detectable in the image,*
- *cvOccstat: whether the object is estimated to be unoccluded or occluded,*
- *cvDistance: vector of normalized relative distances between each pair of objects in a frame,*
- *cvOrientation: a vector of relative movement orientation between each pair of objects (an angle between their directions of movement),*
- *cvHistdist: interframe object shape variation estimated as a difference of Histograms of Oriented Gradients* (*DHoG*).

Object type classification

The classification component determines the segmented object type and discerns humans from other types of objects (vehicles, for example). This is done by using the previously computed *cvHistdist* descriptor that shows the deformation of the object's shape as time passes [3]. It is believed, that humans exhibit a repeating change in shape while walking or running, while other rigid bodies, like vehicles, do not. Thus, a simple nearest neighbor classifier is used to determine the type of object.

High-level video analysis

To recognize human behavior we employ a modified Bayesian network, constructed using a domain ontology, organized by decomposing the scene into several levels of abstraction: overall scenarios, situations which the objects take part in, their roles and their attributes.

Building an ontology

An ontology encompasses a structure of concepts relations, which can be used to describe the aspects of the world and perform logical inference. The supposed content of a scene is organized in a hierarchical ontology by decomposing it into several levels of abstraction.

The lowest level of abstraction contains the visual descriptors, computed from the video data: object speed, track, occlusion state etc.

The object *state* (attributes) is an object dynamics property which shows the state or states in which the object is currently in. Suppose, a man can simultaneously be in three states: '*Active*', *'Running'* and '*Waving hands'*.

Object *role* is the human's overall line (model) of behavior in a frame, which consists from the combination of states the object is currently in. This means, a role can be comprised from several states, For example, the states *'Active'* and *'Running'* mean a role of a *'Runner'*, while *'Running'* and *'Waving hands'* means *'Panicking'*.

The *situation* to which an individual is involved depends on a human itself and therefore the state/position in a scene and it shows how the objects interact with each other. For example, a *'Fighting'* situation can be two object close enough to each other and moving towards each other, while playing the role of *'Runners'* and having a *'Waving hands'* states.

Finally the highest level of abstraction has *scenarios* – the overall context, general circumstances in a frame, which is comprised of set of situations.

Constructing a Bayesian network

The ontology serves as a basis for constructing a network of Bayesian inference. The structure of a Bayesian network consists of a directed acyclic graph (DAG) G whose connectivity matrix defines the conditional dependence relations among its constituent nodes X and hence defines the form of the conditional probability tables [5].

Learning the network structure requires a means of searching the space of all possible DAGs over the set of nodes X and a scoring function to evaluate a given structure over the training data D.

For a network structure learning we chose a *K2* algorithm [5] — a greedy search technique which starts from an empty network but with an initial ordering of the nodes. In order to compute the score of a candidate network over the training data while avoiding over fitting we used a Bayesian Information Criterion (BIC) [6] which approximates the marginal likelihood using a Minimum Description Length (MDL) approach.

After constructing and learning the Bayesian network is ready for human behavior recognition in a real-time video.

Multiple-camera system

We assume for multi-camera tracking that all cameras are viewing the same ground plane. In order to track people across cameras we first need to discover the relationship between the FOVs of the cameras [4]. When the tracking is initiated there is no information about the FOV lines of the cameras. The system can, however find this information by observing motion in an environment. This is done during a training phase in which a single person walks in the environment [1].

Suppose, without loss of generality, that $L_l^{i,j}$ and $L_r^{i,j}$ are the projections of the left and right FOV lines of camera C_i on camera C_j such that $i, j \in \{1..n\}$, where n is the total number of cameras. Suppose, the object being tracked in camera C_j enters or exits camera C_i from its left side. The point in C_j, at which the bottom of object touches the ground plane, actually lies on the projection of FOV of C_i on camera C_j. A least squares method is used to obtain $L_l^{i,j}$ from multiple such observations. In case of non-overlapping cameras, suppose an object exits from C_j. We keep on predicting the position of the object for a certain time interval T. The prediction is made by using a linear velocity model. If the object enters C_i from the left side, within interval T, the predicted position in C_j provides a constraint to determine $L_l^{i,j}$. Basically, we are extending the coordinate space of C_j to obtain a virtual overlap between FOVs of C_i, and C_j. All correspondences are known during the training phase since there is a single object in the environment. Thus, by using the abovementioned method, we can find the relationships between the FOVs of all pairs of cameras in which transition of objects is possible within time T. An example of the line generation process is shown in Fig. 2.

Establishing object correspondence across multiple cameras

The correspondence problem occurs when an object enters the *FOV* of a camera. We need to determine if the object is already being tracked by another camera or it is a new object in the environment. Suppose an object *O* enters camera C_i from the left side. Let *S* be the set of the cameras, which contains the projection of left *FOV* line of C_i. Let $L_l^{i,j}$ denote the projected line in camera $C_j \in S$. Let *P* be the set consisting of objects that are currently visible in C_j or have exited the camera within time *T*. For each object $P_k \in P$, where *k* being the object's label, a Euclidean distance, $D(P_k^j, L_l^{i,j})$, is computed from line $L_l^{i,j}$. Note that the positions of exited objects are continuously updated by linear velocity prediction. If the object *O* is present in any C_j, its distance from $L_l^{i,j}$ should be small. Therefore the object *O* is assigned a label based on the criteria of the minimal of the distances $D(P_k^j, L_l^{i,j})$.

Finally, the object *O* is given a label as described above. Note that the distance from the line only puts a spatial and temporal constraint for label assignment. If an object exits the environment while in the non-overlap area and a new person enters in the same time frame then it will be assigned the wrong label. To cater for this situation an appearance based distance measure can be combined with the distance function [8,9].

Fig. 2. An example of the FOV line generation process

Testing results

The initial testing of the proposed algorithms was performed using a CAVIAR dataset, which provides a set of videos, showing the same scenes from two different angles, and also a PETS 2006 dataset, which in turn, shows the scene from three angles. The initial testing showed that establishing camera correspondence and human behavior recognition is performed with fairly good results (table).

Initial testing results

Type of activity	Total number	Correctly classified	% Success
Standing	4	2	50 %
Walking	5	5	100 %
Running	4	4	100 %
Browsing	4	3	75 %
Fighting	3	3	100 %

It should be noted that the types of activities characterized by active movements (such as walking and running) are recognized much better that activities that incorporate being in a static positions and lack of movement due to proposed algorithms' heavy dependence on movement segmentation approaches [9, 11].

For experimental performance evaluation of proposed solution a wireless network of two cameras was created, with both cameras overlooking different non-overlapping sections of a hallway. Each camera was accompanied by an ARM based Raspberry PI minicomputer [Rasp], connecting them to a wireless network using USB Wi-Fi modules.

Video stream from each camera is processed on a local mini-computer, which includes moving objects segmentation, object tracking and visual descriptors estimation. Then video processing results from each mini-computer as well as the visual descriptors are sent to the main server via the wireless network, where they serve as a basis for camera FOV correspondence establishment, object classification and behavior recognition.

The tests demonstrated that the proposed system is capable of real-time human behavior recognition in a multi-camera scenario with satisfactory results. The advantages of the investigated system include the simplicity of deployment and flexibility, as well as a possibility of a remote access from an Internet-terminal or a mobile device.

Conclusion

Performance evaluation showed, that using a wireless network of cameras and a software, which implements our algorithms, not only allows to cut the expenses on the system itself and it's maintenance, but also provides an efficient and accurate human behavior recognition, when a network of overlapping or non-overlapping cameras is used. Furthermore, local video data processing excludes the sensitive (private) data transfer to the server and guarantee the person privacy together with the high security system efficiency. The proposed network can be viewed as a specialized computer vision system, since it doesn't include human identity recognition. It is only supposed to detect a certain behavioral situation and alert the operator about the situation happening in a certain place and a certain time. From this point of view moral aspects of using our system should not be a concern due to its non-intrusive and impersonal nature.

A system like this may be used in places and territories where the security of citizens and accident prevention, and for instance, terrorist attacks is of the first priority, for example subway stations or airports.

References

1. Khan S., Javed O., Rasheed Z., and Shah M. Human Tracking in Multiple Cameras, (2001), *Proceedings IEEE Computer Vision,* Vol. 1, pp. 331 – 336, url: http://visionnas2.cs.ucf.edu/papers/handoff-iccv.pdf.

2. Stauffer Chris, and Grimson W.E.L. Adaptive Background Mixture Models for Real-time Tracking. *Proc. Computer Vision and Pattern Recognition, Conf. 1999,* Vol. 2, url: http://www.ai.mit.edu/projects/vsam/Publications/stauffer_cvpr98_track.pdf.

3. Han F., Shan Y., Cekander R., Sawhney H.S., and Kumar R. A two-stage Approach to People and Vehicle Detection with HOG-based SVM. The 2006 Performance Metrics for Intelligent Systems Workshop, (2006), pp.133 – 140, url:http://pdf.aminer.org/001/102/676/vehicle_detection_using_normalized_color_and_edge_map.pdf.

4. Chang I-Ch., Liu Ch.-Yu, and Huang Ch.-L. Human Activity Linkage Recording for Multiple Cameras with Disjoint Views, (2010), *Proceedings of the International Conference on* IIH-MSP '10. Washington, DC, 15–17 Oct. 2010, pp. 635–638, url: http://sci-hub.org/pdfcache/ 6daef99defd3c460476ad466e88f43c4.pdf.

5. Cooper G., and Herskovits E. A Bayesian Method for the Induction of Probabilistic Networks from data, (1992), *Machine Learning,* October 1992, Vol. 9, Iss. 4, pp 309 – 347 url: http://www.inf.ufrgs.br/~alvares/CMP259. DCBD/Bayes.pdf.

6. Murphy K.P., and Hon B.A. Dynamic Bayesian Networks: Representation, Inference and Learning. PhD thesis, (1992), *University of California at Berkley,* url : http://www.cs.ubc.ca/~murphyk/Thesis/thesis.

7. Rickert M., and Eibl M. Evaluation of Media Analysis and Information Retrieval Solutions for Audio–visual Content Through their Integration in Realistic Workflows of the Broadcast Industry, (2013), *Proceedings of the 2013 Research in Adaptive and Convergent Systems, RACS 2013,* Montreal, QC; Canada; 1 October 2013 through 4 October 2013; pp. 118 – 121. ISBN: 978-145032348-2Source Type: Conference Proceeding Original language: English, doi: 10.1145/2513228.2513285 Document Type: Conference Paper, url: http://sci-hub.org/pdfcache/daa1082beabe4 0a85611c8ddd7d2f908.pdf.

8. Cozzolino A.b, Flammini, F.a, Galli V.a, Lamberti M.a, Poggi G.c, and Pragliola C.a Evaluating the Effects of MJPEG Compression on Motion Tracking in Metro Railway Surveillance *(Conference Paper).* ISSN: 03029743 ISBN: 978-364233139-8Source Type: Book series Original language, English, doi: 10.1007/978-3-642-33140-4_13Document Type: Conference Paper, url: http://95.25.47.1/scimag4/10.1007/978-3-642-33140-4_13.pdf.

9. Janvier B., Bruno E., Marchand-Maillet S. and Pun T. Performance Evaluation of a Contextual News Story Segmentation Algorithm *(Conference Paper),* Vol. 6073, 2006. ISSN: 0277786X ISBN: 081946113X;978-081946113-1 CODEN: PSISDSource Type: Conference Proceeding Original language: English.

10. Piriou G.a , Bouthemy P.a, and Yao J.-F.b. Recognition of Dynamic Video Contents with Global Probabilistic Models of Visual Motion IEEE Transactions on Image Processing, vol. 15, Iss. 11, November 2006, pp. 3417– 3430. ISSN: 10577149 CODEN: IIPRESource Type: Journal Original language, English, doi: 10.1109/TIP.2006.881963 PubMed ID: 17076401Document Type: Article, url: http://sci-hub.org/pdfcache/8d4d82f60e 9fc96dd8c9b6d42798db76.pdf.

11. Parameswaran V.a , Shet V.c , and Ramesh, V.bd. Design and Validation of a System for People Queue Statistics Estimation *Studies in Computational Intelligence.* Vol. 409, 2012, pp. 355-373 © 2012 Springer Berlin Heidelberg. ISSN: 1860949X. ISBN: 978-364228597-4Source Type: Book series Original language, English, doi: 10.1007/978-3-642-28598-1_11Document Type: Article , url1: http://95.25.47.1/scimag 5/10.1007/978-3-642-28598-1.pdf, url2: http://sci-hub.org/pdfcache/db9501d20b 3bcc48a00899d6303c35d3.pdf.

Received 05.10.2013

Antoshchuk Svetlana,
Dr., Professor, Department of Information System. Odessa National Polytechnic University,
tel. (048) 705-85-25;
e-mail:
asgonpu@gmail.com

Brovkov Volodimir, Dr., Professor, Department of Information System. Odessa National Polytechnic University, tel. (048) 705-83-56; e-mail: Vladimir. Brovkov@HTW-Berlin.de

Kovalenko Nikita ,
Assistant, Department of Information System. Odessa National Polytechnic University, tel. (048) 705-83-56;e-mail: kov.nikit@gmail.com