Extraction of Semantic Information from Events

Panagiota Antonakaki *

Department of Informatics and Telecomunications
National and Kapodistrian University of Athens
ganton@iit.demokritos.gr

Abstract. Behavior understanding from events can be considered as a typical classification problem using predefined classes. In our research presented in this thesis, the main idea is to deal with this classification problem using motion analysis for behavior recognition. We describe our research work from its first steps to the final accomplishments. Novel methods are introduced including a methodology whereby frame information is coded in graphs (called Optical Flow Proximity Graphs - OFPGs), using only optical flow to form the feature vector. A symbolic method including two levels of graph representation is also proposed dealing with the same recognition problem. OFPGs are also proposed for video indexing. Furthermore, a bottom-up approach for anomaly detection using a multi camera system is proposed. In the framework of this system, we present the use of one class continuous Hidden Markov Models (cHMMs) for the task of human behavior recognition. An approximation algorithm, called Observation Log Probability Approximation (OLPA), is proposed to overcome numerical stability problems in the calculation of probability of emission for very long observations.

Keywords: video processing, behavior recognition, event detection, anomaly detection

1 Introduction

One of the most important goals of visual surveillance systems is to track objects and further analyze their activities in order to recognize behaviors and even detect anomalies. To this end we try to detect objects in a scene, track them, recognize their action and interactions in order to understand and describe their behaviors. As a behavior we determine a performed motion by a human subject (object of interest). This motion is captured by low-cost cameras (web-cameras) placed in the room where the object of interest is moving, from a distance in order to observe the whole room.

This thesis is concerned with two main subjects. The first subject is based on methods used during our research in order to classify observed motion to predefined behaviors. The part of the thesis focusing on this subject includes all

* Dissertation Advisor: Sergios Theodoridis, Professor - Dr Stavros Perantonis, Researcher

our research steps for behavior recognition and the corresponding experimental results per approach. Before the completion of this part, a novel model applied on raw data of the video for behavior recognition using graph representation is proposed. The behaviors recognized are also used for video indexing. The second subject deals with the classification of behaviors as normal or abnormal. In the part considering this subject we also propose a novel approach of anomaly detection based on short term behavior and trajectory classification using one class Support Vector Machines and an alternation of continuous Hidden Markov Models also used as one class classifiers.

2 Related Work

Human motion analysis is receiving increasing attention from computer vision researchers [8]. For behavior recognition in video streams, many features based on motion, trajectory or shape of the object are used in different methods. For example, Ribeiro et al. in [9], deal with the problem of feature selection and they present features based on the motion and the trajectory of the object, like velocity, speed, optical flow, etc.

Most approaches for 2D interpretation of human body structure focus on motion estimation of the joints of body segments between consecutive frames. Leo et al. in [10] attempt to classify actions at an archaeological site. They present a system that uses binary patches and an unsupervised clustering algorithm to detect human body postures. A discrete HMM is used to classify the sequences of poses into a set of four different actions. In [11], the used feature is called *Star Skeletonization* and is a distance of the extremities of the silhouette from the blob's center of gravity. In [12] the vertical and horizontal projections of the blob are used to recognize people posture. A simple nearest neighbour classifier, that compares the current projections with manually tuned models, is adopted. Roh et al. in [13] base their action recognition task on curvature scale space templates of a player's silhouette.

Recently, several researchers have dealt with the problem of anomaly detection, which is the process of behavior classification as normal or abnormal. A variety of methods, ranging from fully supervised [14, 15], to semi-supervised [16] and unsupervised systems [17–20], have been proposed in the existing literature. It should be noted, however, that most of the existing approaches do not use multi-camera information, except for Zhou et al. in [21], where multiple video streams are combined via a coupled Hidden Markov Model.

Tarassenko et al. in [22] are proponents of the idea that learning normality alone is all that is required for the detection of abnormality.

3 Human Behavior Recognition and Video Indexing

In this thesis we present all steps made during our research work for one person's behavior recognition, including two similar approaches.

3.1 First Approaches for Short Term Behaviors ([3])

Our first approaches propose a hierarchical method separated into steps, each step dealing with a distinct sub-problem. The main idea is to recognize short term behavior and detect events in order to use this information for long term behavior recognition. As short term behaviors we define actions observed under specific time and space constraints and as events we define the change between two different successive short term behaviors. Long term behaviors are described as scenarios which involves different short term behaviors and events.

Because the low performance of the first approaches, the fact that the number of features used in the feature vector increases the computational cost, and based on the assumption that optical flow by itself can be descriptive enough, we also propose a different approach for both one person and multiple persons behavior recognition, using features calculated directly from raw data without tracking being necessary ([23], [26], [**?**]).

3.2 Final Approach for Behavior Recognition and Video Indexing

For the behavior recognition problem we propose two methods based on graph representation. The first method includes graphs (OFPGs) extracted by each frame containing motion neighbourhood information from the whole frame (WFGA). The second method is a symbolic approach including two levels of graph representation (SGA).

In Whole Frame Motion Representation (WFGA), we use the optical flow vectors calculated from the whole frame using a pixel-by-pixel analysis. In the training step we calculate the optical flow vectors from the whole frame, for each frame. Then, the corresponding OFPG per frame is extracted. Thus, all observed behaviors in a frame are represented by the same graph. The graphs extracted from each frame are merged using the U operator into the graphs which represent the frame's assigned classes.

After that we describe a frame as a feature vector. The vector contains the similarities of our frame OFPG to all the class graphs.

Then, an one-class Support Vector Machine (SVM) SVM_c per class c is trained (see Figure 1).

In the testing step, a similar procedure is used. Each testing instance (frame) is represented by one graph. We create the similarity-based feature vector. Each trained SVM_c model decides whether that feature vector (and the corresponding frame) belongs to class c (see Figure 2)..

In the Symbolic Craph-Based Approach (SGA) the frame is segmented into equal areas, each of which is represented by its OFPG (see Figure 3). The resulting graphs of this level are used as symbols forming an index of graphs to hold the different symbols observed in the frame. At the second level, we use this index in order to map each segment of the frame with a specific symbol.

In the second level of graphs, we use the index of graphs to create a symbolic representation of the original frame in the video sequence (see Figure 4 for an overview of the process). We create and update an index, mapping symbols to

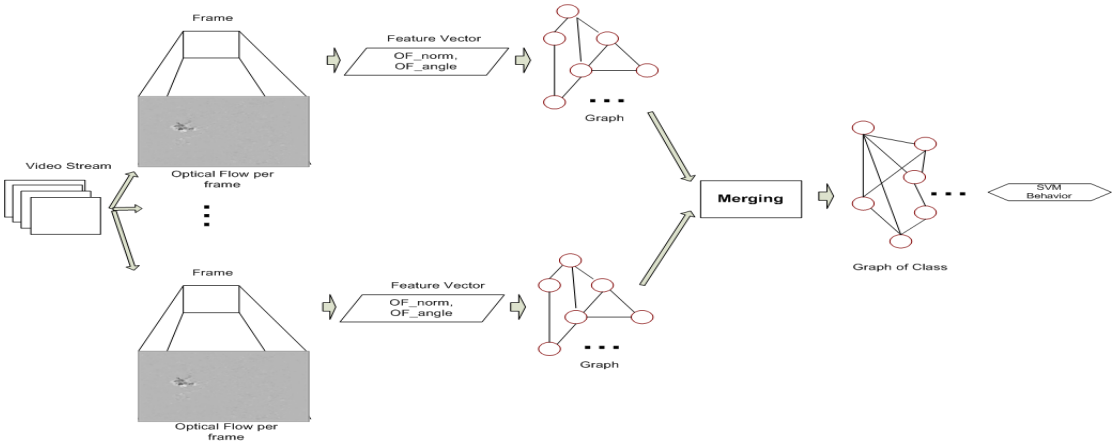

Fig. 1. The training step of WFGA.

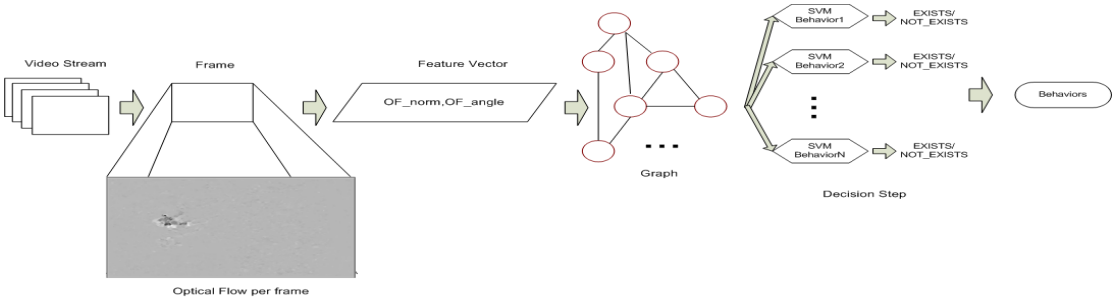

Fig. 2. The testing step of the WFGA framework.

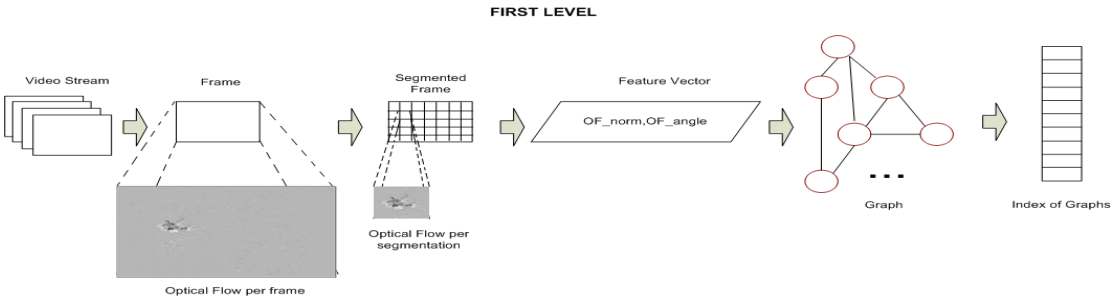

Fig. 3. Extraction of OFPGs from each segment.

subgraphs. These subgraphs are OFPGs of image segments. Each subgraph is assigned by the index a symbol.

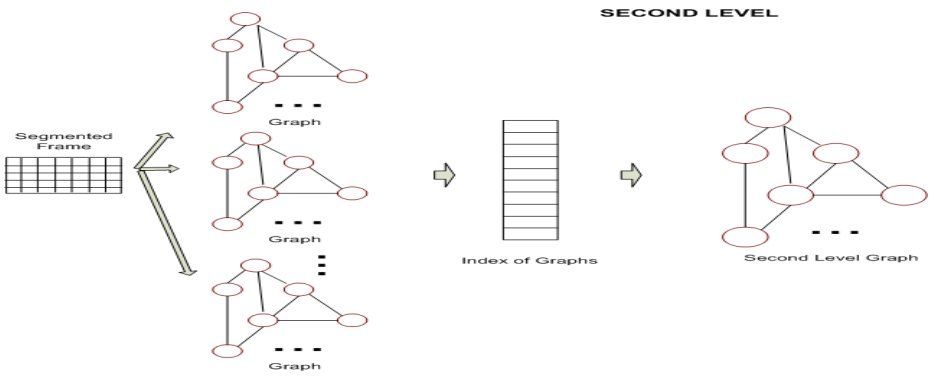

Fig. 4. Extraction of second level graphs.

Once the second level graph of each known class is generated in the training step, the classification step follows. For the classification, each testing instance is represented by a second level graph. The similarity feature vectors are extracted and the SVM classifiers give a binary answer on whether a particular behavior is observed in the frame or not. See Figure 5 for the overview of the testing step in the SGA case. Similarly to WFGA, for the video indexing procedure all behaviors observed in the video are used to tag this video.

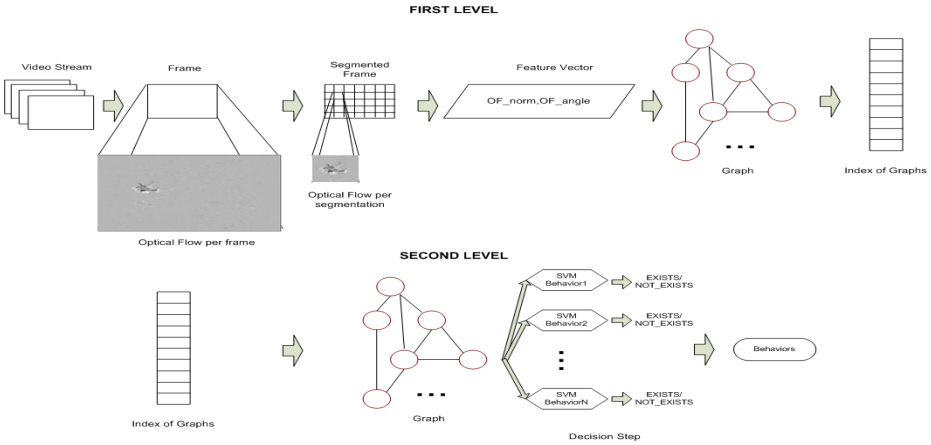

Fig. 5. The testing step of SGA.

4 Anomaly Detection

One of the goals of Smart Surveillance is anomaly detection. Anomaly detection refers to detecting patterns in a given data set that do not conform to an established normal behavior. The patterns are called anomalies and often translate to critical and actionable information in video surveillance calling for human attention if necessary. The second subject of this thesis deals with this problem proposing a novel method for anomaly detection.

The proposed methodology is based on the fusion of data that we collect from several cameras with overlapping fields of view ([**?**]). We perform classification using two different one-class classifiers, a Support Vector Machine (for motion classification) and a continuous Hidden Markov Model (for trajectory classification). A trajectory is the path that a moving object of interest follows through space as a function of time. In our experiments the trajectory of a person includes points (x,y) on the common views of multiple cameras. The final decision on the behavior is made by taking into account outputs from both classifiers.

In motion representation and analysis, our methodology uses information obtained by preprocessing, namely the object's bounding box, the object's blob and sequential positions.

The short term activity is represented by a 7-dimensional feature vector, as follows:

$$f = (v(t), \widehat{v_T}(t), R_T(t), F(t), \Delta F(t), \max(\Delta H(t)), \max(\Delta SD(t))) \quad (1)$$

The corresponding features calculated are: speed, algebraic mean speed, algebraic mean bounding box difference, mean optical flow, mean optical flow difference, max entropy difference and maz standard deviation difference.

The decision whether a short-term behavior is normal or not can be taken by employing a one-class SVM as proposed by Scholkopf [6].

Our second information source for evaluating behavior is the trajectory. The problem of discriminating between normal/abnormal trajectories concerns the definition of a measure that would give sufficiently different values for the two classes. The variable length of the trajectories poses additional difficulties. Long, normal trajectories would have cHMM generation probability values comparable to small values of short, abnormal trajectories, so the observation's length factor needs to be removed.

We proved that for a normal observation sequence (O_{normal}) and for an abnormal one ($O_{abnormal}$) the following condition must hold:

$$\frac{\log P(O_{abnormal}|\lambda)}{length(O_{abnormal})} \ll \frac{\log P(O_{normal}|\lambda)}{length(O_{normal})} \quad (2)$$

then we will be able to use it as a classification measure.

Long normal sequences give small values of cHMM probabilities, due to successive multiplications, making the logarithm of those probabilities to be too high to let the 1 to be damaging. Assuming that the approximation $\frac{\log P(O|\lambda)}{length(O)}$ with $\frac{\lfloor \log P(O|\lambda) \rfloor}{length(O)}$ is acceptable, it can be inserted to Forward Backward algorithm.

According to the above approximations, we can express the algorithm as follows.

1. Initialization:
   $\tilde{\alpha}_1(i) \simeq \lfloor \log \pi_i \rfloor + \lfloor \log b_i(O_1) \rfloor$
2. Induction:
   $\tilde{\alpha}_t(i) \simeq max_j(\tilde{\alpha}_{t-1}(j) + \lfloor \log a_{ij} \rfloor) + \lfloor \log b_j(O_t) \rfloor)$
3. Termination:
   $\tilde{P}(O|\lambda) \simeq \max_i \tilde{\alpha}_t(i)$

This Observation Log Probability Approximation (OLPA) algorithm helps us overcome the problem of consecutive multiplications, by making it possible to use a sum of integers.

5 Experiments

5.1 Behavior Recognition and Video Indexing

For our experiments in behavior recognition and video indexing, we used three different sets of data. In the first data set, simple behaviors were captured in our lab[1](for Semveillance project). The second data set is the commonly used data set from the PETS04 workshop [4]. These latter video sequences have already been used by the CAVIAR project. The third data set is the dataset collected in *The Weizmann Institute of Science*, used by the authors in [5], for comparison reasons.

Our evaluation of the final approach started with the PETS04 data set to determine whether WFGA is better than SGA. On our corpus, the WFGA failed to run in some cases, because of memory complexity.

WFGA

Behavior	Precision	Recall	F-measure
browser	0.2093	0.3459	0.3273
walker	0.9423	0.9491	0.9456
fighters	0.1263	0.9461	0.2223
meeters	0.2934	0.9810	0.4294

SGA

	Precision	Recall	F-measure
browser	0.8065	0.7366	0.7377
walker	0.9918	0.8480	0.9129
fighters	0.5608	0.8766	0.6437
meeters	0.6685	0.8537	0.7448

a

SGA

Behavior	Precision	Recall	F-measure
walk	0.7165	0.9437	0.8138
run	0.3367	0.9376	0.4952
abrupt	0.4003	0.9457	0.5616
active	0.7520	0.9479	0.8381
inactive	0.9853	0.9478	0.9662

b

Behavior	Precision	Recall	F-measure
walk	0.9687	0.9464	0.9573
run	0.7343	0.9473	0.8271
abrupt	0.7782	0.9468	0.8533
active	0.9899	0.9468	0.9678
inactive	0.9999	0.9442	0.9712

c

Fig. 6. (a). Experimental results on PETS04 dataset with noise removal. (b) Precision, Recall and F-measure of SGA on our dataset with noise removal and (c) without noise removal.

[1] The dataset can be made available on demand to the author.

The results of our experiments using the PETS04 data for both proposed approaches are shown in Figure 6. The results of the SGA are higher than those of the WFGA. This indicates that SGA has both lower computational cost and better experimental results.

The application of the SGA method on our dataset yielded promising results as well, especially if one considers that we only use optical flow to detect behavior. The results from experiments using noise removal are shown in Figure 6.

The results of our experiments using the Weizmann data set follow the leave-one-out method, i.e. for every video sequence we remove the entire sequence from the database while other actions of the same person remain. The results are shown in the Figure 7. The results presented in those tables include again low precision values but higher recall values. This is, again, due to the fact that the classes are unbalanced. In Figure 7 we included the results after noise removal and sampling our data set. The latter results are in most cases lower than the results without sampling, due to the fact that video sequences in this data set are small enough (few frames per video) and important information was excluded due to sampling. The same is valid for noise removal.

The experimental results for video indexing using the PETS04 data are shown in Figure 7. These results imply that, when a behavior observed in the video is not included in the training set — and thus not modeled — the system can make the decision that the specific behavior does not belong to any observed and modeled behavior. The specificity values are high enough (~ 0.83), allowing us to conclude that in video indexing the success of the approach follows the success of the behavior recognition case.

Behavior	Precision	Recall	F-measure
walk	0.3188	0.9460	0.4768
run	0.1780	0.9413	0.2994
skip	0.2149	0.9428	0.3499
jack	0.7452	0.9467	0.8339
jump	0.2040	0.9437	0.3354
pjump	0.2041	0.9449	0.3357
side	0.4636	0.9447	0.6219
wave1	0.2556	0.9462	0.4024
wave2	0.2190	0.9460	0.3557
bend	0.5634	0.9456	0.7058

a

Behavior	WFGA Specificity	SGA Specificity
browser	0.3521	0.8444
walker	0.1829	0.8613
fighters	0.5257	0.8157
meeters	0.2605	0.8077

b

Fig. 7. (a). Precision, Recall and F-measure of SGA on Weizmann dataset with noise removal. (b) Experimental results for video indexing.

5.2 Anomaly Detection

As a scene for our experiments we have used our lab, where we installed three cameras, and there we tried to simulate some common scenarios[2]. The experiments measure the performance of two variations of our process, namely the offline and the real-time process.

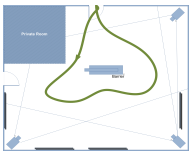

Fig. 8. Example of normal trajectory in the scene.

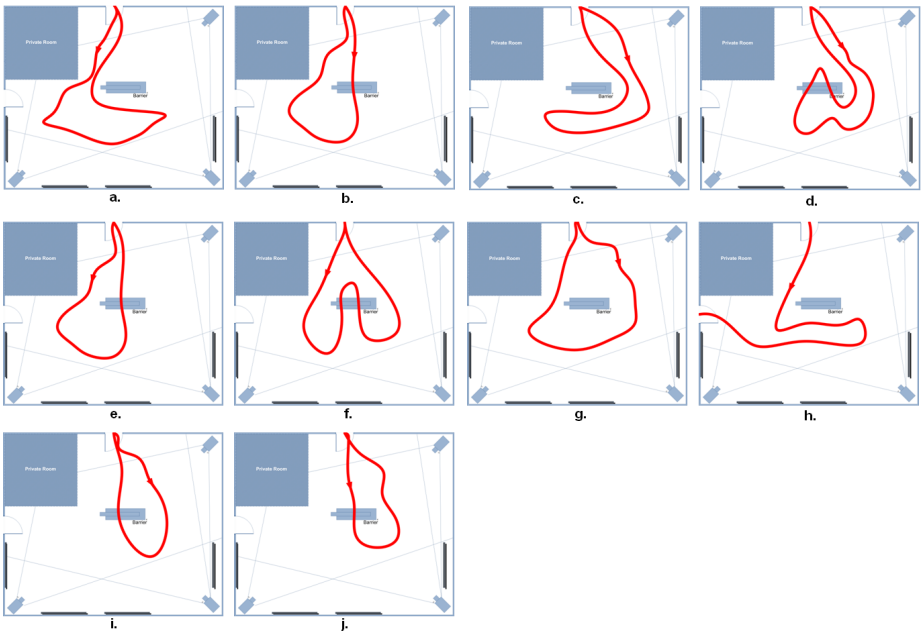

Fig. 9. Examples of abnormal behaviors in the scene.

We have performed a 10-fold cross validation method to test the effectiveness of our system using the offline approach. The videos with normal behaviors

[2] The custom corpus used within our experiments can be made available to any interested party, via e-mail correspondence.

illustrate a person entering the room, buying a ticket, browsing and looking around for several minutes and exiting the room using a preset path (Figure 8). The abnormal behaviors consist of running, abrupt motion or unexpected trajectory (Figure 9).

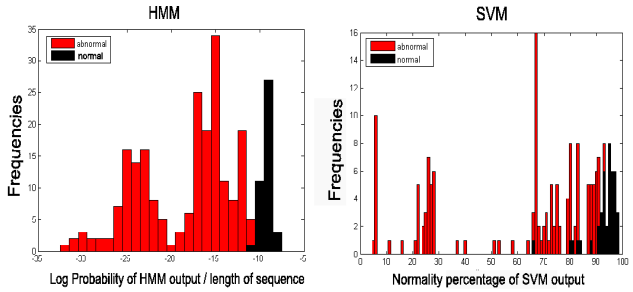

Fig. 10. a. Percentage normality in normal and abnormal behaviors for Support Vector Machine, b. output of continuous Hidden Markov Model for normal and abnormal behaviors. Black color is for normal behaviors and red for abnormal behaviors

For SVM-based classification we set the threshold to be the following function of the mean and the standard deviation of the distribution of the number of allowed abnormal frames within a normal sequence:

$$threshold_{SVM} = mean(Hsvm_{normal}) - 2.5 \cdot std(Hsvm_{normal}) \quad (3)$$

For HMM outputs the minimum value of the distribution of normalized log-probabilities of the normal instances was considered to be the threshold value that separates normal trajectories from the abnormal ones:

$$threshold_{HMM} = \min(Hhmm_{normal}), \quad (4)$$

where Hsvm is the histogram of SVM's outputs and Hhmm is the histogram with HMM's outputs.

Precision and recall have been calculated for the offline and the real-time experiments. For each approach we give the performance for both the SVM and HMM classifier models separately, as well as for the whole system in Figure 11.

6 Conclusions

This thesis was concentrated on high level processes of behavior recognition problem extracting semantic information from events taking place in a video. The first part of this thesis includes preprocessing steps such as background subtraction and tracking in order for the proposed high level approach that we propose to accomblish behavior recognition. In the second part of this thesis a

3-cameras

Offline

	SVM		HMM		Overall	
	Precision	Recall	Precision	Recall	Precision	Recall
Normal	0.9048	0.9286	1	0.9762	1	0.9286
Abnormal	0.7674	0.7071	0.95	1	0.88	1

Real-time

	SVM		HMM		Overall	
	Precision	Recall	Precision	Recall	Precision	Recall
Normal	0.9875	0.9228	0.9960	0.9770	0.9960	0.9105
Abnormal	0.2419	0.6788	0.8478	0.9704	0.8478	0.9375

one camera

Offline

	SVM		HMM		Overall	
	Precision	Recall	Precision	Recall	Precision	Recall
Normal	0.9788	0.8375	1	0.95	1	0.8
Abnormal	0.6708	0.9464	0.913	1	0.7366	1

Real-time

	SVM		HMM		Overall	
	Precision	Recall	Precision	Recall	Precision	Recall
Normal	0.9945	0.9148	0.9953	0.9597	0.9975	0.8525
Abnormal	0.2696	0.8569	0.7544	0.9637	0.5042	0.9861

Fig. 11. Precision and Recall for 3-camera and one camera System on our dataset. The column "Overall" indicates the performance of the combined decision.

bottom-up approach for human recognition understanding is presented, using a multi camera system for anomaly detection.

For behavior recognition adn video indexing we have presented a unified system, which needs no preprocessing steps and a priori knowledge about the surveillance room or how many people are being observed. We have proposed two innovative approaches for classification — and consequently semantic annotation — and indexing of videos, using graph-based representations and analysis methods, with promising experimental results.

Our experimental results demonstrated the good performance of the proposed approaches in the task of recognizing human behaviors in a somewhat noisy environment, with different scenarios of action and participation of different actors. The experiments were implemented using two different datasets with good performance, implying the robustness of the method.

For anomaly detection we have presented a set of theoretical and practical tools for the domain of behaviour recognition, which have been integrated within a unified, automatic, bottom-up system based on the use of multiple cameras performing human behaviour recognition in an indoor environment, without a uniform background.

References

1. Hu, W. and Tan, T. and Wang, L. and Maybank, S.: A Survey on Visual Surveillance of Object Motion and Behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews. 34, 3, 334–352 (2004)
2. Zhou, H. and Kimber, D. and com Inc, A. and Seattle, WA: Unusual Event Detection via Multi-Camera Video Mining. 18th International Conference on Pattern Recognition. 3 (2006)
3. Efros, A.A. and Berg, A.C. and Mori, G. and Malik, J.: Recognizing Action at a Distance. IEEE International Conference on Computer Vision. 2, 726–733 (2003)
4. Fisher, R.B.: The PETS04 Surveillance Ground-Truth Data Sets. Proc. 6th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance. 1–5 (2004)

5. Lena Gorelick and Moshe Blank and Eli Shechtman and Michal Irani and Ronen Basri: Actions as Space-Time Shapes. Transactions on Pattern Analysis and Machine Intelligence (2007)
6. Schölkopf, B. and Platt, J.C. and Shawe-Taylor, J. and Smola, A.J. and Williamson, R.C.: Estimating the Support of a High-Dimensional Distribution. Neural Computation. 13, 7, 1443–1471 (2001)
7. Manevitz, L.M. and Yousef, M.: One-Class SVMs for Document Classification. The Journal of Machine Learning Research. 2, 154 (2002)
8. Moeslund, T.B. and Hilton, A. and Kruger, V.: A Survey of Advances in Vision-Based Human Motion Capture and Analysis. Computer Vision and Image Understanding. 104, 90–126 (2006)
9. Ribeiro, P.C. and Santos-Victor, J. and Lisboa, P.: Human Activity Recognition from Video: Modeling, Feature Selection and Classification Architecture. Proceedings of International Workshop on Human Activity Recognition and Modelling. 61–78 (2005)
10. Leo, M. and D'Orazio, T. and Gnoni, I. and Spagnolo, P. and Distante, A.: Complex Human Activity Recognition for Monitoring Wide Outdoor Environments. Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. 4 (2004)
11. Fujiyoshi, H. and Lipton, A.J. and Kanade, T.: Real-Time Human Motion Analysis by Image Skeletonization. IEICE Transactions on Information and Systems E Series D. 87, 1, 113–120 (2004)
12. Haritaoglu, I. and Harwood, D. and Davis, L.S.: Ghost: A Human Body Part Labeling System using Silhouettes. International Conference on Pattern Recognition. 14, 77–82 (1998)
13. Roh, M. and Christmas, B. and Kittler, J. and Lee, S.: Robust Player Gesture Spotting and Recognition in Low-Resolution Sports Video. LECTURE NOTES IN COMPUTER SCIENCE. 3954, 347 (2006)
14. Dee, H. and Hogg, D.: Detecting Inexplicable Behaviour. British Machine Vision Conference. 447, 486 (2004)
15. Duong, T. and Bui, H. and Phung, D. and Venkatesh, S.: Activity Recognition and Abnormality Detection with the Switching Hidden Semi-Markov Model. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1, 838 (2005)
16. Zhang, D. and Gatica-Perez, D. and Bengio, S. and McCowan, I.: Semi-Supervised Adapted HMMs for Unusual Event Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 1, 661 (2005)
17. Jiang, F. and Wu, Y. and Katsaggelos, A.K.: Abnormal Event Detection from Surveillance Video by Dynamic Hierarchical Clustering. Proceedings IEEE International Conference on Image Processing. 5, 145–148 (2007)
18. Lee, C.K. and Ho, M.F. and Wen, W.S. and Huang, C.L. and Hsin-Chu, T.: Abnormal Event Detection in Video using N-cut Clustering. International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP). (2006)
19. Mahajan, D. and Kwatra, N. and Jain, S. and Kalra, P. and Banerjee, S.: A Framework for Activity Recognition and Detection of Unusual Activities. Indian Conference on Computer Vision, Graphics and Image Processing. 15–21 (2004)
20. Xiang, T. and Gong, S.: Incremental and Adaptive Abnormal Behaviour Detection. Computer Vision and Image Understanding. (2008)

21. Zhou, H. and Kimber, D. and com Inc, A. and Seattle, WA: Unusual Event Detection via Multi-Camera Video Mining. 18th International Conference on Pattern Recognition. 3 (2006)
22. Tarassenko, L. and Nairac, A. and Townsend, N. and Buxton, I. and Cowley, P.: Novelty Detection for the Identification of Abnormalities. International Journal of Systems Science. 31, 11, 1427–1439 (2000)
23. Dimitrios I. Kosmopoulos, Panagiota Antonakaki, Konstandinos Valasoulis, Dimitrios Katsoulas: Monitoring Human Behavior in an Assistive Environment using Multiple Views. Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments (PETRA 2008). 32, (2008)
24. Dimitrios I. Kosmopoulos, Panagiota Antonakaki, Konstandinos Valasoulis, Anastasios L. Kesidis, Stavros J. Perantonis: Human Behavior Classification Using Multiple Views. Proceedings of the 5th Hellenic conference on Artificial Intelligence: Theories, Models and Applications (SETN 2008). 123–134, (2008)
25. Dimitrios I. Kosmopoulos, Anastasios Kesidis, Panagiota Antonakaki, Konstandinos Valasoulis, Stavros J. Perantonis: SemVeillance System: Tracking and Behavior Recognition Under Occlusions. European Conference on Artificial Intelligence (ECAI 2008). (2008)
26. Panagiota Antonakaki, Dimitrios I. Kosmopoulos, Stavros J. Perantonis: Detecting Abnormal Human Behaviour using Multiple Cameras. Signal Processing. 89(9), 1723–1738 (2009)