Unsupervised Abnormal Crowd Activity Detection in Surveillance Systems

Łukasz Kamiński, Paweł Gardziński, Krzysztof Kowalak, Sławomir Maćkowiak

Poznań University of Technology, Polanka 3, 60-965 Poznań Poland
{lkaminski, pgardzinski, kkowalak, smack}@multimedia.edu.pl

Abstract – **We propose an unsupervised method for abnormal crowd activity detection in surveillance systems. Proposed solution is using MPEG-7 Motion Activity descriptors and Particle Filter algorithm for classification. The experiments were performed on UMN dataset sequences. The detection results are comparable to results obtained by supervised methods.**

Keywords – **particle filter, unsupervised anomaly detection, UMN**

I. INTRODUCTION

The number of surveillance cameras has increased significantly during recent years. Therefore, the amount of video material to analyze is still increasing, which is a crucial problem is crowd observation, where fast abnormal activity detection is very critical. Abnormal crowd activity detection is very important and useful in surveillance systems and security control.

The common approach in abnormal activity detection methods is usage of a known normal activity pattern. Firstly, the pattern is designed during learning process from normal activity dataset. Next, in classification step, actual data is compared with the pattern. Finally, a decision whether the activity is normal or not is made based on this comparison.

The requirement of predefined normal activity pattern complicates abnormal activity detection in real surveillance systems. Furthermore, predefined normal activity patterns may change to abnormal in some situations. An example of this situation is people taking an escalator, where the abnormal activity is movement of people against escalators movement direction. In this case, the change of escalators movement direction causes the need of a new normal activity pattern. Otherwise, this system would generate false alarms.

The abnormal crowd activity detection is a very popular topic. Anomaly detection methods can be divided into 3 categories.

First category treats anomaly detection as a binary problem. These methods are fully supervised. During classifier training process both normal and abnormal activity samples are required. For example, algorithms presented in [3,6,10] use SVM classifier for abnormal crowd activity detection. However, classifier training is a complex task because of a large number of possible abnormal activity patterns, which makes algorithms in this category rarely used.

Second category of detection algorithms does not require abnormal activity patterns. However, normal activity patterns are still required. Algorithms which belong to this group were described in [1,17,21].

The last category consists of unsupervised algorithms. These kind of algorithms do not require normal nor abnormal patterns. Several unsupervised algorithms have already been presented, but their usage is limited to uncrowded scenes [14,15,16,24]. Solution presented in this paper belongs to this last category and can be used in both crowded and uncrowded scenes, where normal activity patterns may change over time.

The most work in the topic of anomaly detection is focused on designing a crowd characteristic descriptor, e.g. social force model [5,23] or mixture of dynamic textures [9]. Instead of developing a new descriptor, the MPEG-7 Motion Activity descriptor [8,18] has been used in this paper. This descriptor has already been used in [7], but in that paper the presented method was supervised. As a classifier we use Particle Filter (PF) [12,20], which is a well-known technique used in moving object tracking [2,19,22]. A combination of Particle Filter algorithm and MPEG-7 Motion Descriptors is a novel approach for anomaly detection.

The rest of the paper is organized as follows. Section 2 presents the method of Motion Activity descriptor extraction. Section 3 provides a detailed information about Particle Filter prediction and update equations. Section 4 contains experimental results and conclusions. Section 5 contains a summary of this paper and presents possible future work.

II. MOTION ACTIVITY DESCRIPTOR

In the presented solution, MPEG-7 [18] descriptors were used. Those descriptors belong to Motion Activity descriptors group [8]. The calculation of those descriptors is based on the simple operations on motion vector components [11]. Therefore, the descriptors calculation can be done on compressed images, which significantly accelerates feature extraction process.

The first descriptor is a motion intensity. It is used to describe crowd dynamic. High value of this descriptor corresponds to high dynamic of the observed scene. In order to calculate this descriptor all motion activity energy in the image must be summed. Assuming, that i-th motion vector energy is equal to

$$E_i = x_i^2 + y_i^2, \tag{1}$$

where x is a horizontal component of motion vector and y is a vertical component of motion vector, the motion intensity is equal to

$$E_c = \sum_{i=0}^{N} E_i. \quad (2)$$

The second descriptor is an 8-bin histogram of movement directions. This descriptor describes object movement directions in the scene. In order to calculate this descriptor an angle of each motion vector is calculated. In the next step, the histogram with bins separated from each other by 45 degrees is created.

The third descriptor describes the number and size of active objects in the scene. This descriptor consist of three spatial distribution parameters: NSR, NMR, NLR (Number of Short/Medium/Long Runs). These values are extracted from thresholded motion vector energy matrix. This matrix is calculated for each image in the sequence. Assuming, that E_{avg} is an average motion vector energy in the image and motion vector coordinates are (i, j), the elements of the matrix are equal to

$$E_{i,j}^{thrsh} = \begin{Bmatrix} E_{i,j} & if & E_{i,j} > E_{avg} \\ 0 & if & E_{i,j} \leq E_{avg} \end{Bmatrix}. \quad (3)$$

The calculation of NSR, NMR and NLR parameters is based on a raster scan over E^{thrsh} matrix. During this scan, the length of continuous non zero motion vector energy sequence is calculated. Assuming, that W is a width of E^{thrsh} and N is a length of actual non-zero sequence, the NSR, NMR and NLR are presented as follows

$$\begin{array}{lll} NSR = NSR + 1 & if & N > \frac{2W}{3} \\ NMR = NMR + 1 & if & \frac{W}{3} < N \leq \frac{2W}{3}. \\ NLR = NLR + 1 & if & N \leq \frac{W}{3} \end{array} \quad (4)$$

The fourth descriptor describes the ratio of motion intensity in the largest active region to the total motion intensity in a scene. This descriptor is useful in case of gathering and splitting of the crowd and is equal to

$$E_{ratio} = \frac{E_{area}}{E_c}, \quad (5)$$

where E_{area} is the largest active object energy.

Finally, all of the presented descriptors are merged together in order to form a single descriptor $descr$. This descriptor is more practical to use and is given by

$$descr = [d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8, NSR, NMR, NLR, E_c, E_{ratio}] \quad (6)$$

where $d_1 - d_8$ are movement direction histogram bins.

III. Anomaly Detection

In order to detect which images in the image sequence show abnormal crowd activity, PF algorithm was used [12]. PF algorithm is an implementation of recursive Bayes filter by Monte Carlo simulations. The goal of this algorithm is to estimate an unknown probability density function. Assuming, that the crowd realizes a sequence of states, we can assume, that actual state x_k depends only on the previous state

$$x_k = f_k(x_{k-1}, v_{k-1}), \quad (7)$$

where $f(.)$ is a nonlinear function and v is a process noise. In this case, it is possible to estimate state x_k based on the measure z_k

$$z_k = h_k(x_k, n_k), \quad (8)$$

where $h(.)$ is a nonlinear function and n is a measurement noise. Detailed information about PF algorithm can be found in [20].

PF algorithm is using particles, which in this case describe a state of the crowd. The number of particles in this algorithm is constant and is equal to $N = 100$. In this case, particle is described by 13 elements vector, which corresponds to descriptor size.

Crowd state estimation is divided into two steps: prediction and update. In prediction step, a transition model is used to predict the future state of the crowd. Taking into account the random nature of crowd motion, proposed solution uses a random walk model, which is based on an addition of properly crafted noise to particle components

$$x_k = \boldsymbol{A}x_{k-1} + \boldsymbol{B}v_{k-1}, \quad (9)$$

where $\boldsymbol{A}, \boldsymbol{B}$ are identity matrices and v is a process noise.

In the update step, the z_k measure is available so particles weights update is possible. In this paper, z_k measure is equal to descriptor computed for current image. An update of w_k weights is performed by calculation of exponential function values of negative Mahalanobis distance between particle and z_k measure, which is given by

$$w_k = \exp(-MD(descr, x_k)). \quad (10)$$

Particle weight corresponds to probability value. This means that high w_k weight value for particle x_k corresponds to high probability that state of the crowd is normal. Anomaly is detected when all particles weights are smaller than predefined threshold value $thrsh$. In other words, abnormal activity is detected when computed descriptor does not match to predicted state of crowd.

In order to prevent degeneration [20] of PF algorithm, in the last step, particles with low weight are replaced by a new ones with higher weight. This approach allow to continuously track the changes of crowds state.

IV. Experimental Results

Experiments were performed on UMN dataset [25]. This dataset contains 11 testing sequences. Each sequence begins with a normal activity presenting walking people followed by the crowd splitting which is an abnormal activity. Sample images from UMN sequences dataset are shown in Fig. 1. This dataset was used to evaluate the effectiveness of the algorithm for global activity recognition, because the entire crowd participates in the abnormal activity.

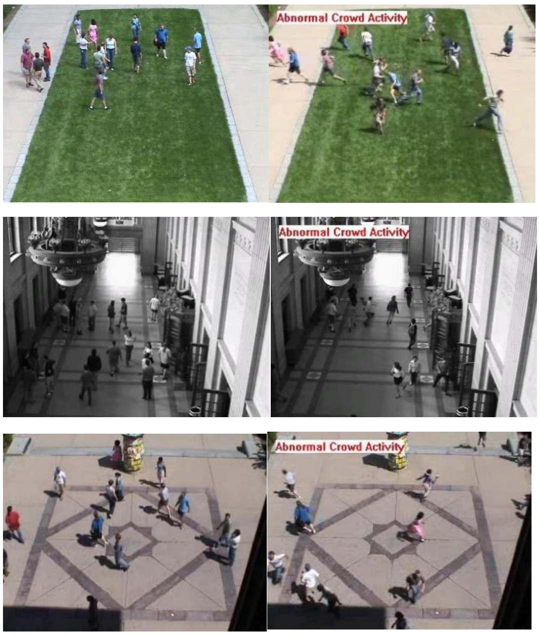

Figure 1. Sample images from UMN sequences dataset. Left column shows normal activity and right column shows abnormal activity.

The UMN dataset provides manually marked images which contain abnormal crowd activity. To measure the effectiveness of presented approach *precision* and *recall* metrics were used, which are defined as follows

$$precision = \frac{TP}{TP+FP}, \quad (11)$$

$$recall = \frac{TP}{TP+FN}, \quad (12)$$

where TP is true positive, FP is false positive and FN is false negative. Calculation of *precision* and *recall* metrics for different values of *thrsh* parameter allows to plot Receiver Operating Characteristic curve (ROC curve), which is shown in Fig. 2.

To compare presented method with other algorithms the Area Under the Curve (AUC) was used. High value of this parameter corresponds to high efficiency of the algorithm. The results are shown in Tab. I.

Effectiveness of presented method is comparable with other algorithms. It is worth mentioning that other algorithms require samples of normal activity to train the classifier. On contrary the proposed method is free from this requirement and is fully unsupervised. Normal activity model in proposed solution is built on the fly without prior initialization of the algorithm. Obtained results confirm that PF algorithm can be effectively used to detect abnormal crowd activities.

TABLE I. COMPARISON OF OBTAINED RESULT WITH OTHER METHODS

Method	*AUC*
Yang 2012 [3]	0.975
Mehran 2009 [5]	0.960
Proposed algorithm	**0.945**
Zhao 2011 [13]	0.940
Wang 2012 [4]	0.900

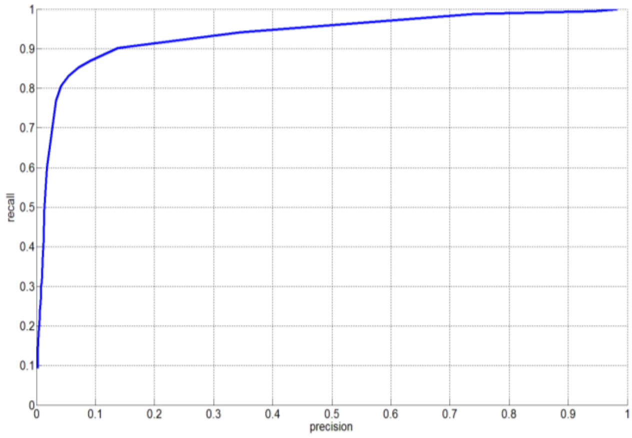

Figure 2. ROC curve of proposed algorithm for different value of *thrsh* parameter.

V. CONCLUSION

In this paper an unsupervised abnormal crowd activity detection method was presented. Proposed method uses MPEG-7 motion activity descriptors to characterize crowd motion in observed scene and Particle Filter algorithm to detect abnormal crowd activities. Unlike other methods, presented algorithm does not require predefined samples of normal nor abnormal activities. Proposed method is fully unsupervised, which simplifies practical implementations of such system, because the detection of different types of anomalies is possible. The experiments were performed on UMN sequences dataset. Obtained results present high efficiency compared to other methods which require training.

Abnormal crowd activity detection is a complex problem. Presented method can recognize the presence of anomaly in the scene, but cannot point the localization of this anomaly. It is caused by descriptor which is calculated for entire image. In future work, we plan to modify the algorithm so that it can localize anomalies in the scene.

ACKNOWLEDGMENT

The presented work has been funded by the Polish Ministry of Science and Higher Education for the status activity consisting of research and development and associated tasks supporting development of young scientists and doctoral students..

REFERENCES

[1] S. Wu, H.-S. Wong, Z. Yu, “A Bayesian model for crowd escape behavior detection”, IEEE Transactions on Circuits and Systems for Video Technology, vol. 24(1), pp. 85-98, 2014.

[2] K. Okuma, A. Taleghani, N.D. Freitas, J.J. Little, D.G. Lowe, “A boosted particle filter: multitarget detection and tracking”, Proceedings of Eighth European Conference on Computer Vision, pp. 28-39, 2004.

[3] H. Yang, Y. Cao, S. Wu, W. Lin, S. Zheng, Z. You, “Abnormal crowd behavior detection based on local pressure model”, Signal & Information Processing Association Annual Summit and Conference, pp. 1-4, 2012.

[4] B. Wang, M. Ye, X. Li, F. Zhao, J. Ding, “Abnormal crowd behavior detection using high-frequency and spatio-temporal features”, Machine Vision and Applications, vol. 23(3), pp. 501-511, 2012.

[5] R. Mehran, A. Oyama, M. Shah, “Abnormal crowd behavior detection using social force model”, IEEE Conference on Computer Vision and Pattern Recognition, pp. 935-942, 2009.

[6] Y. Miao, J. Song, “Abnormal Event Detection Based on SVM in Video Surveillance”, IEEE Workshop on Advance Research and Technology in Industry Applications, pp. 1379-1383, 2014.

[7] H. Liao, J. Xiang, W. Sun, Q. Feng, J. Dai, “An Abnormal Event Recognition in Crowd Scene”, Sixth International Conference on Image and Graphics, pp. 731-736, 2011.

[8] A. Divakaran, “An Overview of MPEG-7 Motion Descriptor and Their Applications”, Computer Analysis of Image and Patterns, Lecture Notes in Computer Science, vol. 2124, pp. 29-40, 2001.

[9] W. Li, V. Mahadevan, N. Vasconcelos, “Anomaly Detection and Localization in Crowded Scenes”, IEEE Transactions on Pattern Analysis and Machine Intalligence, vol. 36(1), pp. 18-32, 2014.

[10] H. Lin, J.D. Deng, B.J. Woodford, “Anomaly detection in crowd scene via online adaptive one-class support vector machines”, IEEE International Conference on Image Processing, pp. 2434-2438, 2015.

[11] ISO/IEC 14496-10, “Coding of Autio-Visual objects – Part 10: Advanced Video Coding”, 2010.

[12] M. Isard, A. Blake, “CONDENSATION – Conditional Density Propagation for Visual Tracking”, International Journal of Computer Vision, vol. 29(1), pp. 5-28, 1998.

[13] J. Zhao, Y. Xu, X. Yang, Q. Yau, “Crowd instability analysis using velocity-field based social force model”, IEEE Visual Communications and Image Processing, pp. 1-4, 2011.

[14] Y. Ito, K.M. Kitani, J.A. Bagnell, M. Hebert, “Detecting Interseting Events Using Unsupervised Density Ratio Estimation”, Computer Vision – ECCV 2012. Workshops and Demonstrations, Lecture Notes in Computer Science, vol. 7585, pp. 151-161, 2012.

[15] O. Boiman, M. Irani, “Detecting irregularities in images and in video”, Tenth IEEE International Conference on Computer Vision, vol. 1, pp. 462-469, 2005.

[16] H. Zhong, J. Shi, M. Visontai, “Detecting unusual activity in video”, Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 819-826, 2004.

[17] M.D. Dreitenstein, H. Grabner, L. Van Gool, “Hunting Nessie – Real-time abnormality detection from webcams”, IEEE 12^{th} International Conference on Computer Vision Workshops, pp. 1243-1250, 2009.

[18] ISO/IEC 15938-3:2002/Amd 3:2009. Information Technology – Multimedia content description interface – Part 3: Visual, Amendment 3: Image signature tools.

[19] Z. Khan, T. Balch, F. Dellaert, “MCMC-based particle filtering for tracking a variable number of interacting targets”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27(11), pp. 1805-1819, 2005.

[20] A. Doucet, S. Godsill, C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering”, Statistics and Computing, vol. 10(3), pp. 197-208, 2000.

[21] B. Zhao, L. Fei-Fei, E.P. Xing, “Online detection of unusual events in videos via dynamic sparse coding”, 2011 IEEE Conference on Computer Vision and Pattern Recognition, pp. 3313-3320, 2011.

[22] Z. Qi, R. Ting, F. Husheng, Z. Jinlin, “Particle Filter Object Tracking Based on Harris-SIFT Feature Matching”, Procedia Engineering, vol. 29, pp. 924-929, 2012.

[23] D. Helbing, P. Molnar, “Social Force Model for Pedestrian Dynamics”, Physical review E, vol. 51(5), pp. 4282-4286, 1995.

[24] X. Wang, X. Ma, E. Grimson, “Unsupervised Activity Perception by Hierarchical Bayesian Models”, 2007 IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

[25] UMN, Unusual crowd activity dataset, http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi