

Detection and Localization of Anomalies from Videos based on Optical flow Magnitude and Direction

S. Bansod and A. Nandedkar

Deptment of Electronics and Telecommunication, SGGSIE&T, Vishnupuri, Nanded.
{bansodsuprit@sggs.ac.in;avnandedkar@sggs.ac.in}

Abstract. Anomalies in video scenes means unexpected or unusual activity which is usually not frequently observed. Such activities hence are rare and require sudden attention so that it can be detected as early as possible. There is a need to automatically identify and locate where such anomaly is present. Optical flow magnitude and direction based method is an automated system built on motion, position and statistical features of moving objects present in video. Moving objects are identified by means of optical flow and are represented using bounding box. The normal behaviors is learned beforehand for different objects. A generalization of normal behavior is captured by clustering different directional motions in the scene. Anomalous behavior of objects are detected and localized using motion and positions differing from normal behavior. The performance of proposed method is compared with existing methods by using standard benchmark datasets available online such as UCSD and UMN.

Keywords: *Anomalies, optical flow, directional motion.*

1 Introduction

After the invention of video camera, due to easy access and ease of use video data is available in chunks and generated day by day. Even an automatic camera like closed circuit television cameras is widely established at malls, square of cities, railway station etc. The video from such cameras can be analyzed for various purposes like object detection, person re-identification, pedestrian or vehicle detection and counting, activity and behavior recognition etc. From these wide range of applications anomalous behavior detection from the crowd can be a challenging task. Video comprises many objects and more specifically detecting or identifying the behavior of the crowd require more efforts. The crowd is a collection of a group of people moving randomly or sometimes mannerly in a scene. Major hurdles in such situations are occlusion, irregular shapes, varying illumination conditions depending on the environment and camera position etc. [1]. Abnormality can be any irregular behavior, such as fighting, yelling, theft in public, and in case of traffic not observing traffic rules, wrong side, taking wrong U – turn etc. As abnormal situations are irregular, there is a need to understand properly the normal occurrences of scenes from input video. The approaches [5],[8] recently has developed a trend to first learn the normal situations based on context, since as the scene changes the same object or the same activity may become anomalous which is again a challenge for computer vision researchers while developing automated system for crowd behavior detection. The benchmark dataset [14, 15] contain training videos with only normal behaviors and testing videos with abnormalities, which help to clearly distinguish between normal and abnormal situations so that an automated system can take a proper decision. For learning normal behaviors, recent approaches are built on motion and appearance [2, 3], spatiotemporal [6-8] and trajectories [9, 10] as prime features. The input frame is subdivided into small blocks called cell and the features are computed on this small part so that detail analysis can be done. Also pre-processing may involve background removal so that foreground is employed for feature calculation. It is found in the literature that various ways have been proposed by researchers to solve the problem of abnormal activity detection from videos. As mentioned previously, there are approaches based on features calculated to distinguish between normal and abnormal activities. Adam et al. [2] developed a real time robust unusual event detection by capturing the scenes from fixed location cameras. Optical flow based speed and direction and their variations are the key features of the work. Kim and Grauman [3], detect the abnormality in videos using Baye's Rule, input to Baye's classifier is mixture of probabilities computed at atomic level of the frame and at each atomic level optical flow features are calculated. Xu et al [5], designed a technique to detect anomalies at local and global level using histogram of optical flow calculated at each cell level. A unified energy function is designed at cell level using both local and global approaches to detect anomalies.

Mehran et al. [4], used a social force model which is based on pedestrian movement dynamics, where the crowd is collection of interacting particles. A force flow matrix is constructed from the forces of interacting particles.

From the force flow matrix likelihood probabilities are computed, and by fixing a threshold the image is categorized as normal or abnormal. Kratz and Nishino [6], has proposed method to detect anomalous behavior from the extremely crowded scenes in videos. Dense local motion patterns are analyzed to describe the common behavior within a scene. 3D Gaussian distributions of spatiotemporal gradients are computed in a set of non-uniform local motion patterns. Temporal relations are captured by Hidden Markov Model (HMM) and spatial by coupled HMM. Li et al [7], proposed a combine detection and localization of anomalous behaviors in the crowded scenes for temporal and spatial anomalies based on Mixture of Dynamic Textures (MDT). Authors have introduced two normalcy models that jointly account for the appearance and dynamics of complex crowd scenes, temporal normalcy and spatial normalcy. Li et al [8], developed spatiotemporal video volume (STVV) based approach which computes spatial and temporal gradients in a fixed cell from a frame as raw features. These are clustered and represented within a large video cubes using BoW method. At last a dictionary is prepared to learn information about STVVs using the sparse representation approach and detection of anomalies is done using sparse reconstruction cost.

Wu et al. [9] proposed a method for anomaly detection in crowded scenes based on Lagrangian particle dynamics and chaotic invariant. By using particle advection method positions of moving particles are estimated, and representative trajectories of moving objects are generated. Trajectories are clustered using K-means to obtain position information. Anomaly detection is based on probability map of trajectories. Zhang et al. [10] proposed a trajectory based approach to identify anomalies. Author used motion, size and texture descriptor features to learn the behavior of normal and abnormal scenes at region of interest. With the help of these descriptors from region to consecutive frames small pathlets are drawn and are merged to get trajectory. Pathlet merging and path prediction is done by SVM classifier.

The novelty of our method lies in the fact that it uses motion as well as position of objects as a key features. We utilize the information in optical flow to separate the dominant moving objects so no background removal step is needed making the system time efficient. The feature computations are done on available foreground objects only. Extracted features along with their location are clustered using K-means to represent prominent motion and position of normal objects. Abnormality is decided based on motion and position of objects deviating from normal objects.

The remaining paper is organized as follows, Section 2, describes about proposed method; Section 3 discusses about experimental results and performance comparison on benchmark datasets and in Section 4 we conclude.

2 Optical Flow Magnitude and Direction based Anomaly Detection and Localization System

Fig. 1 gives generalized system block diagram for detection and localization of anomalies from the video [11]. The input video frames can be first preprocessed by applying a low pass filter to neglect the effects of noise present in the frames while capturing video. For all such preprocessed frames, features are calculated to identify motion. The motion features are further clustered to represent the dominant motion, these dominant motion helps to understand the behaviors. At last the anomaly is detected and localized from these behaviors.

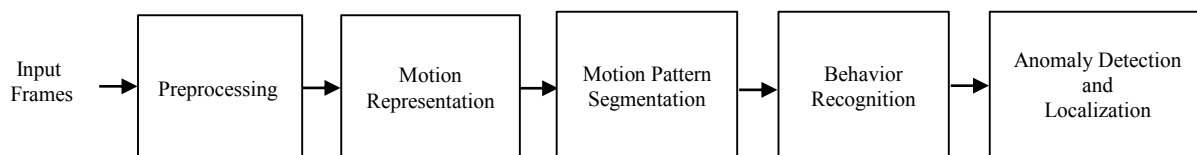


Fig. 1. Generalized System block diagram for Anomaly Detection and Localization

The flow chart for optical flow magnitude and direction based anomaly detection and localization is shown in Fig. 2. The proposed approach is divided into three steps, i) Pre-processing and Feature Calculation, ii) Clustering and iii) Anomaly Detection and Localization.

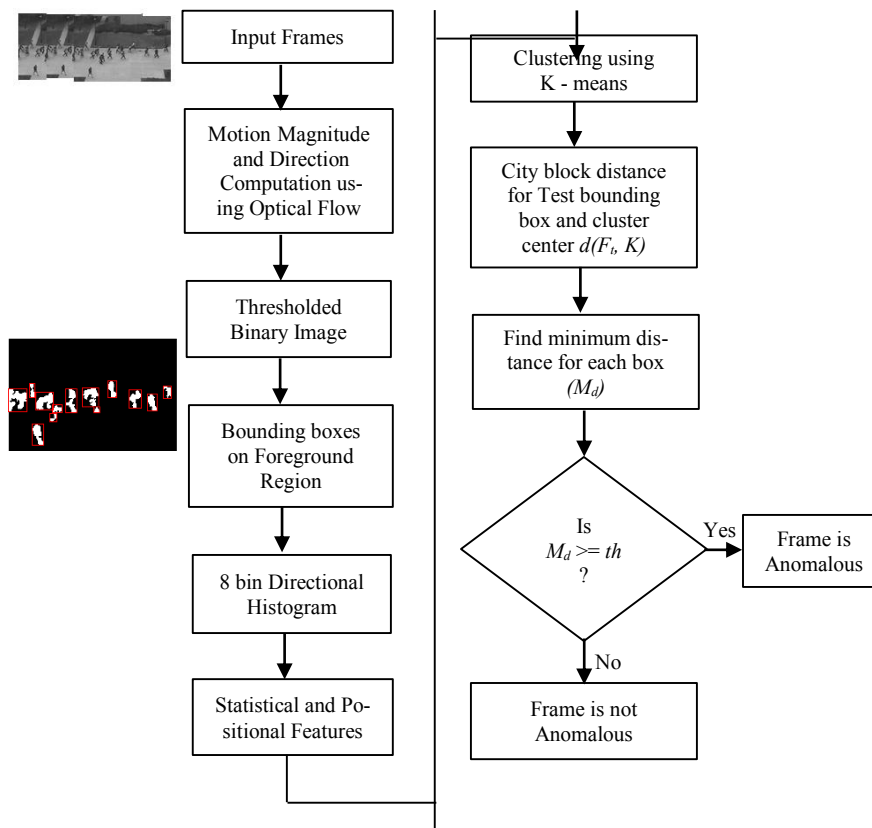


Fig. 2. Flowchart for Proposed System

2.1 Pre-processing and Feature Calculation

In this step, motion, position and statistical features are calculated for normal as well as abnormal scenes. At first, it is important to understand the behavior of normal scenes, so the input normal video frames are filtered using a Gaussian filter as a pre-processing step. We compute optical flow for consecutive two frames using Liu et al [12] method. The output of optical flow is the velocities of X- and Y-direction, the magnitude and direction of each pixel in the frame are computed by Eq. (1) and Eq. (2).

$$mag = \sqrt{u_x^2 + v_y^2} \quad (1)$$

$$dir = \tan^{-1} \left(\frac{v_y}{u_x} \right) \quad (2)$$

where u_x – velocity in X-direction, v_y – velocity in Y-direction.

We apply a fixed threshold for these magnitude values and obtain a binary image which helps to remove the background and the remaining foreground area indicates the motion present in the frame. To neglect small to very small motion regions morphological operation is applied to remove them and the output binary image represents the most dominant motion objects for the frame. These moving objects are highlighted by bounding boxes, as well as object location and its centroid is calculated for each bounding box in each frame. Dimension of each bounding box differs according to the object size. Now for each object three sets of features are calculated as follows-

2.1.1 Histogram of Magnitudes

We divide magnitudes within each bounding box into 8 directions from -180° to 180° to obtain 8 directional histogram of magnitudes, as shown in Fig. 3. Normalization of each bin is done by dividing the maximum magnitude for that bin. Maximum value is found out from the whole normal video scenes for each bin.

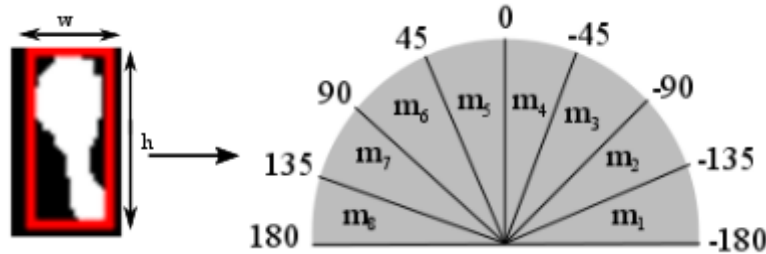


Fig. 3. Eight Directional Histogram representation for Object

2.1.2 Statistical Features

These are derived features from the motion features. It includes –

1. The difference of maximum magnitude and minimum magnitude from the 8 bin for each object, it is the dominant motion for the object.
2. Standard deviation for the bin, representing the variation of magnitudes for the object, given by Eq. (3).

$$\sigma_m = \sqrt{\frac{1}{N} \sum_{i=1}^N (m_i - \mu_m^2)} \quad (3)$$

$$\mu_m = \frac{1}{N} \sum_{i=1}^N m_i \quad (4)$$

where N – number of bins, $N=8$, m_i – magnitude of i^{th} bin, μ_m – mean of magnitude vector, σ_m – standard deviation for magnitude vector.

3. Entropy for the bounding box is the measure of randomness, more the entropy more is the randomness. It is normalized with respect to the maximum entropy of the scene.

$$E = \sum_{i=1}^M (p * \log_2 p) \quad (5)$$

2.1.3 Positional Features

It becomes necessary to know the positions of the abnormalities which help to locate them and deviate from the positions of the normal objects. Position of normal object can be fixed or it can be predicted, but anomalies may occur suddenly, so it is necessary to know positions of normal objects. Also, with motion features, position helps to detect slower motion anomaly. Positional feature mainly includes-

1. The height and width of the bounding box, covering the object as shown in Fig. 3. The height and width of each bounding box are normalized with respect to the height (h) and width (w) of a frame.
2. Area of the object, product of height and width of bounding box.

Combining all these features, we finally obtain a 16 dimensional feature vector.

2.2 Clustering

For every frame we get a certain number of bounding boxes, and after calculation of above features for the normal scene we have huge data. To learn the normal scene behavior, an unsupervised clustering is done using K-means technique [16]. The cluster centers indicate the normal motion and positions of normal scenes. We choose the cluster number, $K = 20$.

2.3 Anomaly Detection and Localization

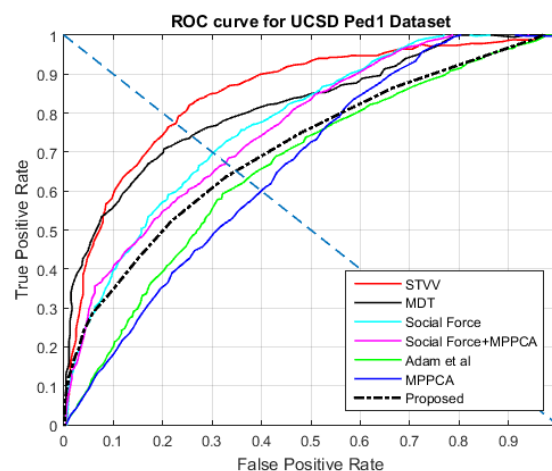
For each object from the test set, we find out the city block distance $d(F_i, K)$ between a test set feature F_i and cluster centers of K – means (n_1, n_2, \dots, n_K), so a $b \times k$ dimensional vector is obtained. A minimum value of distance M_d for the object is find out from this vector, and the process is repeated for each object from all the frames. The minimum distance value for each object is compared with a threshold, for a frame if the distance for a single object is greater than the threshold then that frame is treated as anomalous.

For such a frame containing an anomaly, we arrange distances of objects in descending order and the object with the highest distance is marked in red bounding box with anomaly and remaining in green with normal objects in the frame, hence localization is done. The bounding boxes are marked with the help of bounding box locations calculated at the time of feature computation.

3 Experimental Results

The aim of experiment to detect and locate anomalies from videos and to compare the performance of the proposed approach by using Area Under Curve (AUC), Equal Error Rate (EER) and time requirement analysis to detect and locate anomalies in each frame. EER is $1 - \text{Efficiency}$. For experimentation we used two benchmark datasets of crowd anomaly available online, UCSD [14] and UMN [15]. UCSD has two scenes, UCSD Ped1 and UCSD Ped2. Ped1 contains 34 image sequences for training and 36 sequences for testing. Both train and test sequence have 200 frames, of dimension 158×238 . Ped2 contains 16 sequences for training and 12 sequence for testing with frame number varies from 120 to 180. It has frame dimension of 360×240 . Anomalies included in Ped1 and Ped2 are cycle, skater, truck, three wheeler etc. UMN contains three scenes namely, ground, museum and court. People walking is a normal activity and sudden running is considered as abnormal. The experiment on both datasets is performed on MATLAB, the optical flow code is mixed, with system configuration of Intel i7 processor 3.40 GHz and 8 GB RAM.

As mentioned in [7], the detection can be of two types, frame level and pixel level. The proposed method at frame level has achieved for Ped1 dataset AUC of 67.78%, EER of 34.85% and computation time is 0.9 fps. For Ped2 it is 52.67% and EER of 17.42% with computation time of 2.12 fps. UMN dataset has AUC of 80.41%, EER of 25.30% and computation time of 2.22 fps. Comparison of AUC and EER are shown in Table 1 with state of art methods, indicating proposed method has least EER for Ped2 dataset.



(a)

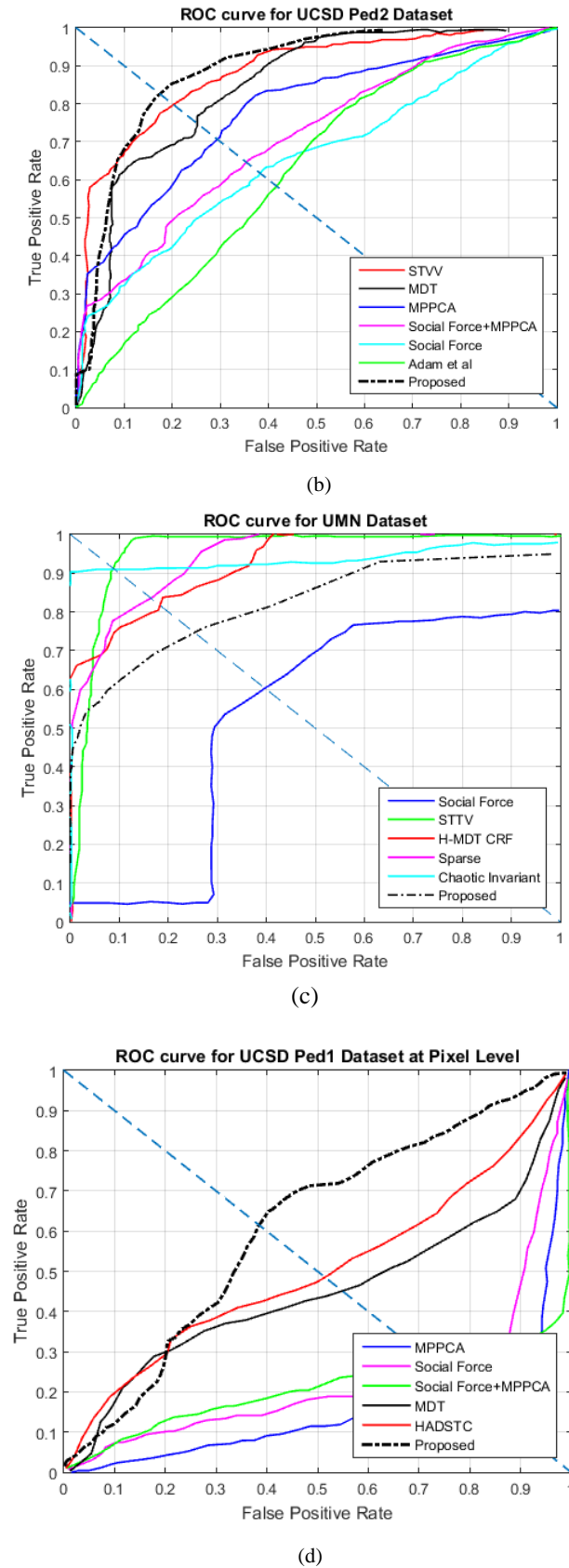


Fig. 4. Frame level and pixel analysis (a) frame level ROC curve for UCSD Ped1 dataset, (b) frame level ROC curve for UCSD Ped2 dataset, (c) frame level ROC curve for UMN dataset and (d) pixel level ROC curve for UCSD Ped1 dataset.

ROC for frame level detection are presented in Fig. 4 (a-c). The pixel level detection is considered when the 40% of anomalous region in a frame are detected. Fig. 4 (d) shows the pixel level detection ROC with AUC of 60.58% and Rate Detection (RD) is $1 - EER$, which is 61.50% for the proposed method. Fig. 5 and Fig. 6 shows the anomaly localization results from UCSD Ped1 and Ped2 datasets.

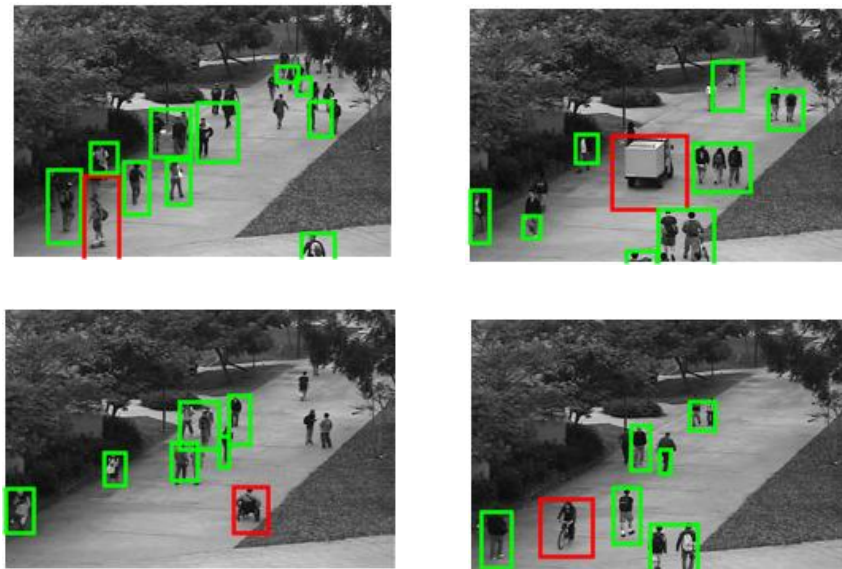


Fig. 5. Anomaly Localization from UCSD Ped1 dataset



Fig. 6. Anomaly Localization from UCSD Ped2 dataset

Table 2 gives the time requirement analysis for UCSD dataset in frames per second (fps). The proposed method requires very less time to detect and locate anomaly from UCSD Ped1 dataset and from UCSD Ped2 it is comparable to state of the art methods.

Table 1. Performance of Anomaly Detection, Comparison of AUC/EER (percent)

Method	Dataset		
	UCSD Ped1	UCSD Ped2	UMN
Adam [2]	62.4 / 38	60.8 / 42	–
MPPCA [3]	63 / 40	72.4 / 30	–
Social Force [4]	68.3 / 31	61.3 / 42	94.9 / 12.6
Social Force + MPPCA	69.3 / 32	67.9 / 36	–
Sparse [13]	46.1 / 19	–	99.6 / 2.8
MDT [7]	80.7 / 25	83.4 / 25	99.5 / 3.7
HADSTC [5]	85.4 / 22	88.2 / 21	–
STTV [8]	87.2 / 21	89.1 / 20	93 / 10
Proposed	67.78 / 34.85	52.67 / 17.42	80.41 / 25.30

Table 2. Time requirement analysis for UCSD dataset in (frames per second, fps)

Method	Ped1	Ped2
Sparse [13]	3.8	3.8
MDT [7]	1.11	1.38
HADSTC [5]	5	5
STTV [8]	3.4	4.8
Proposed	0.9	2.12

4 Conclusion

In this paper, we proposed optical flow magnitude and direction based method to identify and localize anomalies from videos. The important features used to verify the frame abnormality are motion, and positional features. We separate the dominant moving objects by using motion information, for those objects feature computation is done. The features are clustered to represent prominent motion and positions. Anomalies are detected for the objects deviating from normal motion and positions. Localization is done with the help of positional features. We conducted experiments on standard datasets UCSD and UMN widely used for performance evaluation. The proposed method outperforms with very less EER for Ped2 subset of UCSD dataset and high RD of 61.50% compared with other techniques available. Also it is time efficient as it takes very less time to process a frame, 0.9 fps for Ped1, 2.12 fps for Ped2 and 2.22 fps for UMN.

References

- [1] Chandola, V., Banarjee, A., Kumar, V.: Anomaly Detection: A Survey, Technical Reprot, ACM Digital Library, 41, 3, 1-15, (2009). doi: 10.1145/1541880.1541882
- [2] Adam, A., Rivlin, E., Shimshoni, I., and Reinitz, D.: Robust real-time unusual event detection using multiple fixed location monitors, IEEE Tran. on Pattern Analysis and Machine Intell., 30, 3, 555–560, (2008). doi: 10.1109/TPAMI.2007.70825
- [3] Kim, J., Grauman, K.: Observe locally, infer globally: A space-time MRF for detecting abnormal activities with incremental updates, IEEE Conf. on Computer Vision Pattern Recognition, 2921–2928, (2009). doi: 10.1109/CVPR.2009.5206569
- [4] Mehran, R., Oyama, A., and Shah, M.: Abnormal crowd behavior detection using social force model, IEEE Conf. on Computer Vision Pattern Recognition, 935–942, (2009). doi: 10.1109/CVPR.2009.5206641
- [5] Xu, D., Song, R., Wu, X., Li, N., Feng, W., Qian, H.: Video Anomaly Detection based on a Hierarchical Activity Discovery within Spatio temporal contexts, Int. Journal of Neurocomputing, 143, 2, 144-152, (2014).

- [6] Kratz, L., Nishino, K.: Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models, *IEEE Conf. on Computer Vision Pattern Recognition*, 1446–1453, (2009). doi: 10.1109/CVPR.2009.5206771
- [7] Li, W., Mahadevan, V., Vasconcelos, N.: Anomaly Detection and Localization in Crowded Scenes, *IEEE Tran. on Pattern Analysis and Machine Intell.*, 36, 1, 18 – 32, (2014). doi: 10.1109/TPAMI.2013.111
- [8] Li, N., Wu, X., Xu, D., Guo, H., Feng, W.: Spatiotemporal Context Analysis within Video Volumes for Anomalous Event Detection and Localization, *Int. Journal of Neurocomputing*, 155, 1, 309-319, (2015).
- [9] Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, *IEEE Conf. on Computer Vision Pattern Recognition*, 1-6, (2007). doi: 10.1109/CVPR.2007.382977
- [10] Zhang, T., Wiliem, A. and Lovell, B.: Region-based Anomaly Localization in Crowded Scenes via Trajectory Analysis and Path Prediction, *IEEE Int. Conf. on Digital Image Computing: Tech. Appl.*, 3-9, (2013). doi: 10.1109/DICTA.2013.6691519
- [11] Li, T., Chang, H., Wang, M., Ni, B., Hong, R. and Yan, S.: Crowded Scene Analysis: A Survey, *IEEE Tran. on Circuits and Systems for Video Tech.*, 25, 3, 367-386, (2015). doi: 10.1109/TCSVT.2014.2358029
- [12] Liu, C.: Beyond pixels: exploring new representations and applications for motion analysis (Ph.D. thesis), Massachusetts Institute of Technology, (2009).
- [13] Cong, Y., Yuan, J, Liu, J.: Sparse Reconstruction Cost for Abnormal Event Detection, *IEEE Conf. on Computer Vision Pattern Recognition*, 3449-3456, (2011). doi: 10.1109/CVPR.2011.5995434
- [14] UCSD Anomaly Detection Dataset by University of California and San Diego, <http://www.svcl.ucsd.edu/projects/anomaly/dataset.html>
- [15] Unusual Crowd Activity Dataset by University of Minnesota, <http://mha.cs.umn.edu./movies/crowdactivity-all.avi/>
- [16] Duda, R. O., Hart, P. E., Stork, D. G.: *Pattern Classification*, John Wiley & Sons, (2012).