

Affective State Based Anomaly Detection in Crowd

Glorija Baliniskite¹, Egons Lavendelis^{2*}, Mara Pudane³
¹⁻³Riga Technical University, Riga, Latvia

Abstract – To distinguish individuals with dangerous abnormal behaviours from the crowd, human characteristics (e.g., speed and direction of motion, interaction with other people), crowd characteristics (such as flow and density), space available to individuals, etc. must be considered. The paper proposes an approach that considers individual and crowd metrics to determine anomaly. An individual's abnormal behaviour alone cannot indicate behaviour, which can be threatening toward other individuals, as this behaviour can also be triggered by positive emotions or events. To avoid individuals whose abnormal behaviour is potentially unrelated to aggression and is not environmentally dangerous, it is suggested to use emotional state of individuals. The aim of the proposed approach is to automate video surveillance systems by enabling them to automatically detect potentially dangerous situations.

Keywords – Anomaly detection in crowd, dangerous anomaly detection, emotional state, person extraction from crowd, surveillance system automation.

I. INTRODUCTION

The importance of security as a human value is broadly reflected in development of the latest smart home and surveillance technologies. Examples include automated population monitoring in China [1] and Great Britain [2], and surveillance cameras and security guards in almost every public institution.

In most used video surveillance systems nowadays monitoring and reaction are performed by trained employee or employees. In such a configuration, human perception and reasoning play a crucial role in the effectiveness of the whole system, because it depends on the collection and processing of information from cameras. Perception of a person is characterised by apperception – dependence of personality's perception and emotional or rational interest at a moment in time can be influenced by health status, age, interests, current needs, value systems, mood, etc. [3]. For this reason, a dangerous situation may be overlooked, especially when the crowd is dense [4]. Even if the system is assumed to work in an ideal manner minimising human factor, for example, by providing redundancy and ability to merge observations, such a system is expensive to maintain.

If the goal of automation is to create a system that enables the identification of an individual whose behaviour is potentially dangerous and then to pass this information on to a user for further decision making, then the anomaly detection can be used as a foundation of the system. An anomaly in this case is the difference in behaviour of the individual from that

of other individuals around him or generally in crowd. Besides, in choosing parameters and context for samples in the system, attention should be paid to concepts described in sociology and psychology – individual's behaviour and emotions. In this case, context is the description of the situation in which the dataset was obtained.

Since behaviour is a reaction to internal or external irritation and emotions are a person's internal reaction to events, individuals and objects [5], information about a person's emotional state can serve as a context for validating anomalies.

The remainder of the paper is organised as follows. Section I outlines the related work. The proposed approach is described within Section II. Section III presents the results obtained and Section IV concludes the paper and provides the possible areas of future work.

II. RELATED WORK

Studies dealing with the problem of defining an individual's abnormal behaviour are rooted in psychology and sociology [6]. For example, in sociology, interaction with the world, including the behaviour of an individual, triggers a particular emotional state as an internal response, which in turn leads to appropriate behaviour [5].

Security monitoring systems are the most widely used in crowd monitoring because the information obtained by a camera is visually comprehensible to a person. Nevertheless, automation-based research has focused on the use of acoustic sensors to detect anomalies [7] or events [8], fusing thermal cameras and conventional camera data to predict crowd behaviour [2], [9] and offers describing dangerous behaviour assuming it is behaviour that will be different from the majority or will be atypical of the environment or will be anomalous. Following those assumptions, it becomes possible to use several standard algorithms for anomaly detection.

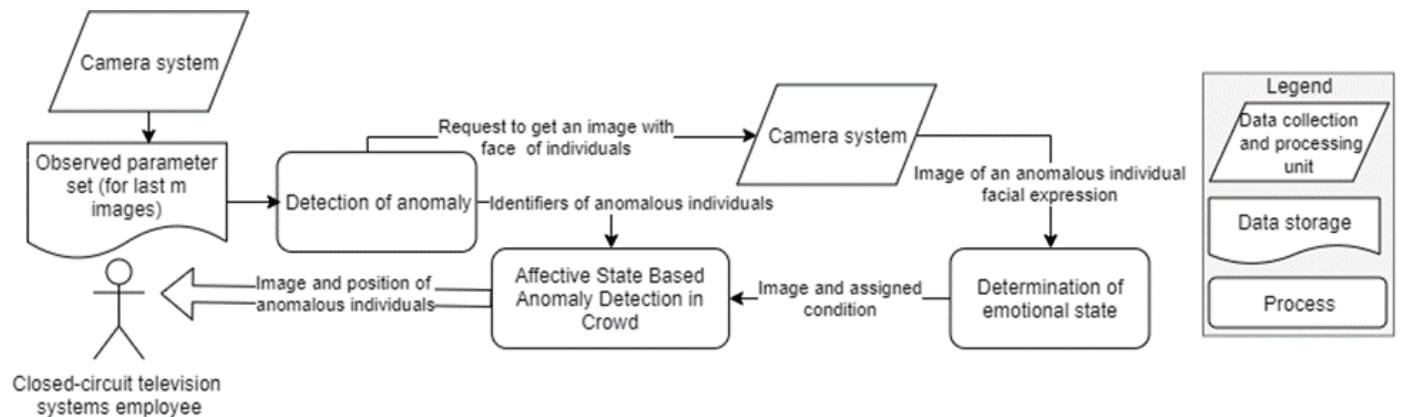
Based on the labelled data, the used methods of detecting anomalies are divided into three groups – supervised, semi-supervised and unsupervised anomaly detection. Supervised anomaly detection algorithms include Bayesian networks [10] and k -nearest neighbours (k -NN) [11]. The goal of these algorithms is to create normal and anomalous data models. These methods have the advantage of being highly accurate when the available labelled data are general enough. Disadvantage is the need for marked data because it is time-

* Corresponding author's e-mail: egons.lavendelis@rtu.lv

consuming, complex, and not necessarily class-specific. The normal-to-anomalous data ratio is also problematic because anomalies in data occur at a lower frequency than normal data [9]. These algorithms are used when anomalous events such as

an unattended bag, a fight, a visible weapon, etc., must be detected [13].

Semi-supervised anomaly detection algorithms include k -NN [11], a single class Support Vector Machine (SVM) [13] and



isolation forest [14]. These algorithms require labelled nor

Fig. 1. The proposed approach diagram.

data. One-class marking is a much simpler task, as it requires the developer to understand the nature of the studied phenomenon. However, these algorithms have to be modified to accommodate the changing crowd environment. Labelled normal data are used in research [15], in which the speed of movement is used as the individual's behaviour.

There are no such disadvantages to unsupervised methods such as k -mean [16], Local Outlier Factor (LOF) [17], Gaussian model clusterisation [18] and histogram analysis [10], but these methods are less accurate than the supervised and semi-supervised algorithms [12]. These methods have been used to automate surveillance systems in [19], [20], [21], where the individual's behaviour is determined by the speed of movement.

The above-mentioned methods of automating the closed-circuit television (CCTV) systems consider individual's speed of movement as a characteristic for detecting anomalous behaviour. However, much more metrics and context can be obtained from camera data, which can potentially improve the anomaly detection algorithm, for example, when anomalous behaviour is towards human movement in the crowd, interactions with other people and so on. Based on the fact that human behaviour can depend on emotional state, the author believes that emotions can be used as a context to determine if an anomaly is dangerous by reducing the amount of information passed on to the observer.

III. THE PROPOSED APPROACH

The camera system captures images and processes the incoming image to produce the individual's behaviour. The behaviour of the individual is described by a pre-selected set of parameters, such as the speed of movement, acceleration, trajectory, etc. The most recent observations of the last m images are stored in the data store, where m expresses a time

interval (m is the time multiplied by the number of camera frames per second), in which the context or the behaviour observed in the crowd should not change drastically. The anomaly detection process looks at the data in the data warehouse and determines which individuals behave abnormally using the isolation forest algorithm. A request is made to assess the emotional state of individuals with abnormal behaviour. Based on the emotional state and the anomaly observed by the anomaly detector, affective state-based anomaly detection process provides the CCTV employee with an image of tagged individuals, which can be potentially dangerous (an image from which the emotion was determined may also be issued). The process diagram can be seen in Fig. 1.

A. Detection of Anomaly

It is assumed that each individual is characterised by the following characteristics: speed of movement, acceleration, trajectory and its change, individual size and distance to other agents. When individuals are at a distance where their direct physical interaction is possible, the following additional data are collected: the difference in velocity of movement between agents, the dynamics of the distance between them, the trajectory similarity. To identify individuals whose behaviour is significantly different from the behaviour of the majority of the current crowd, an isolation forest algorithm supplied with a dynamic component is used. Such an approach is chosen to detect the anomaly in the way so that the developed solution is flexible to context change. An example where context-dependent behavioural patterns change – at a shopping mall opening celebration, the density of people would be very high, such an observation would be normal, while the dispersion of people in some part of the room would be considered anomalous, while the next day the mall has a daily attendance,

so the density of people in the room is low and now it is the normal behaviour.

Isolation forest anomaly detection algorithm belongs to a group of semi-supervised algorithms, which can also be used in an unsupervised manner. This algorithm is applied within this paper because it can correctly handle cases where multiple normal behaviour patterns are observed in one place, there is no need for the training set, no need to normalise all parameters. The disadvantage of this algorithm is the need to define a parameter – the average cutting length of a branch, which defines a threshold beyond which observation is considered anomalous.

Using the average cutting length of a branch of each individual (derived from forest generation and ranked in ascending order), the set of individuals with anomalous behaviour is calculated using formula (1):

$$\begin{aligned}
 A &= \{id_1, id_2, \dots, id_m\}; \\
 d_{\text{koef_avd}} &= \frac{1}{n} \sum_{j=1}^n d_{\text{koef_}j}, \\
 d_{\text{koef_error}} &= \frac{1}{n} \sum_{i=1}^n |d_{\text{koef_}i} - d_{\text{koef_avd}}|, \\
 \begin{cases} id_i \in A \text{ if } d_{\text{koef_}i} > d_{\text{koef_avd}} - d_{\text{koef_error}}, \\ id_i \notin A \text{ in other cases.} \end{cases}
 \end{aligned} \tag{1}$$

In formula (1), A – a set containing indices of abnormal behaviour, id_i – indices of abnormal behaviours, n – the number of observed behaviours – the number of elements in total A , i – the individual number of decomposition coefficients, $d_{\text{koef_}i}$ – branch average cut-off coefficient of i -th, $d_{\text{koef_avd}}$ – an average value of the pruning coefficient, $d_{\text{koef_error}}$ – the absolute mean error of the pruning coefficients (used here to determine the cut-off threshold at which the behaviour is considered anomalous).

B. Potentially Dangerous Emotional States

There are six basic emotions describing emotional states – sadness, anger, surprise, fear, disgust and joy. Emotions do not have strict boundaries and there can be a fluent transition between them. This transition can be represented in a two-dimensional space (see Fig. 2). The placement of emotions within the emotion space depends on whether the emotion induces a passive or active reaction and whether it is positive or negative. For example, joy is a positive, active emotion driven by a person’s behaviour that has no negative effect on themselves or others. Joy can be slightly motivating because it is on the active side of the emotion space. Fear, on the other hand, is a negative and active emotion and person driven by this emotion may endanger others. For example, somebody can accuse an innocent person just because he himself is afraid of being punished. Negative and active emotions not always can lead to dangerous behaviour; the expression of reaction is also influenced by a person’s emotional education, temperament, social status, etc. [22]. Negative emotions spread in the crowd a lot faster than positive; thus, fast

negative emotion detection is vital for threat prevention [23], [24].

Using computer vision, person’s emotions can be seen via their facial expressions or body postures. When assigning the task of defining the emotions of a crowd of individuals, it should be borne in mind that most of the existing techniques of emotion detection algorithms are not directly applicable to crowd analysis.

Facial emotion detection is the most widely researched and used methods in a variety of applications – commerce (service quality assessment), security systems (augmented reality tools for police units), human rehabilitation and community integration (integration of people with autism spectrum disorder through training to understand emotions), etc. [4], [25]–[28]. These approaches have reached 90 % accuracy

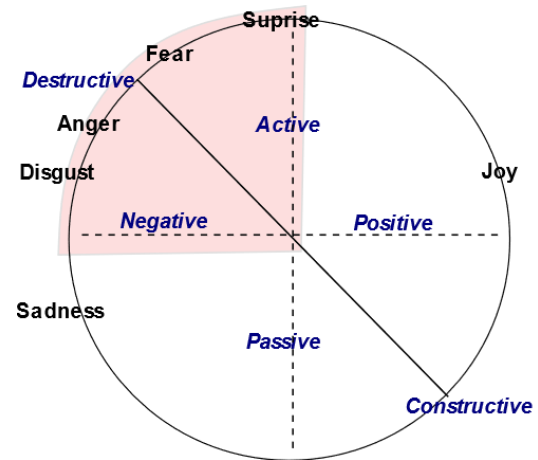


Fig. 2. Emotional state distribution in 2D space. The area containing potentially dangerous emotions is marked in red.

[25], which is higher than the human accuracy estimated in the experiment conducted in 2013 (75 non-professional participants participated in the experiment). However, the obvious and significant limitation of the approach is the need for the face to be seen in the image, as well as it is a quadrilateral resolution to be sufficient for the algorithm to work correctly.

The authors suggest that any emotion that is neutral or negative and active is considered a dangerous emotional state because it is more rapid in the crowd, more persistent, and mainly motivates the individual to cause aggressive, dangerous situations for the individual himself or others. The part of the 2D emotional space is marked red in Fig. 2.

The rules for detecting anomalies based on the emotional state are as follows:

- if during the processing of information obtained from the static camera, a behavioural anomaly is detected for any individual and its emotional state is located in the negatively active part of the emotion space, then the emotional state-based anomaly for a particular individual is confirmed;
- if there is a behavioural anomaly in any individual but the emotional state cannot be determined, then the emotional state-based anomaly for a particular individual is assumed to be confirmed. The reason for such an

approach is the fact that the danger of the situation, if the anomaly is not detected, is greater than when the emotional state-based anomaly is false;

- if behavioural anomalies are observed but the identified emotional state is not within the space of the potentially dangerous emotions, then the emotional state-based anomaly is not confirmed.

IV. EXPERIMENT

An experiment was run to test the proposed approach. The experiment aimed at applying the proposed approach to real-life records and using computer vision methods to detect, distinguish individuals and collect the needed data for anomaly detection.

An isolation forest algorithm is used, which generates one hundred trees and determines the average cutting length of the tree branch for each observation. There are two sets of parameters that serve as input for algorithms – displacement along the x axis (dx), y axis (dy), and both axis (d) (hereinafter group A) and velocity (v), acceleration (a) (hereinafter group B), see (2), where x_i, y_i – the position of the individual in the i^{th} frame, the coordinate system of the image, dx_i, dy_i – the components of the displacement of the individual between $(i-1)^{\text{th}}$ and the i^{th} frame. The number of parameters in a group corresponds to the number of dimensions, in which the observations are located and in which trees are created.

$$\begin{aligned} (dx_i, dy_i) &= (x_i - x_{i-1}, y_i - y_{i-1}), \\ d_i &= \sqrt{dx_i^2 + dy_i^2}, \\ v_i &= d_i, a_i = v_i - v_{i-1}. \end{aligned} \quad (2)$$

Due to the fact that extracting emotional state from the video is a time-consuming task, within the first prototype the emotional states of individuals were simulated. In particular, the described emotional state was generated as a number from 0–100; an emotional state above 50 would be equivalent to observing a negative, active emotion.

A. Data Set

To visualise the operation of the proposed approach, three

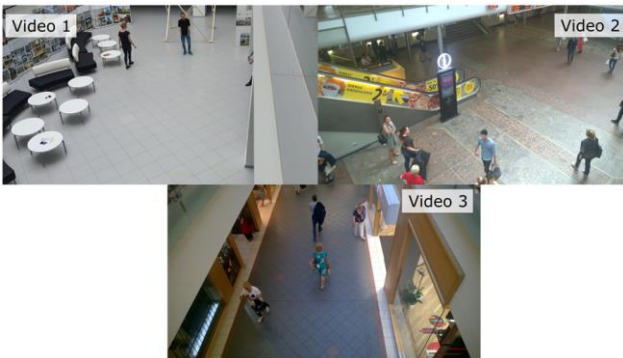


Fig. 3. Snapshots of a captured video set. Video 1 simulates anomalous behaviour in a crowd, while Video 2 and Video 3 are taken under natural conditions in crowded space.

videos were used for emotional state recognition. Sample images of the set can be seen in Fig. 3. Initially, the data were obtained in mp4 format, which was later split into images. The OpenCV library was used to perform the task [29].

B. Splitting the Crowd Into Individuals

To divide the crowd into individuals, the authors use the You Only Look Once (YOLO) open-access project [30]. This project classifies and detects objects, i.e., determines the object's belonging to the class and its location in the image.

The YOLO project uses the open-access convolutional neural network framework Darknet [31] and the training weights of the Common Objects in Context (COCO) dataset. It is a dataset containing 91 classes of marked objects in natural context images [32].

However, it must be considered that this approach is chosen due to the limited resources during the development of the first prototype to prove the concept. To develop a software product for video surveillance applications, the neural network should be trained on the images obtained from the surveillance cameras, as it would adapt to YOLO specific application, providing higher detection accuracy. During the processing of the video captured in the work, samples were selected that could be successfully processed using data trained with the COCO dataset.

C. Retrieving Characteristics Needed to Detect Anomaly

As the set of parameters is different to each person, individual identification must be performed. Within the present research, the Simple Online and Realtime Tracking project (SORT) [33]–[36] is used to solve this problem.

The centre of lower edge of the resulting outline will be used as a location mark of the individual. It will enable retrieving further parameters since detection of movement requires interaction of snapshots in time (in the standard case). Using an outline also enables normalisation of these snapshots by transforming perspective.

For perspective transformation, four points are known in the image, the relative positions of which are known in space. Within the framework of the research, datasets were taken in places with tiled floors, so that in the post-processing step there would be guidelines for positioning the points and indicating the ratio of the distance between the points indicated.

V. RESULTS

The section considers the results of the experiment. During the experiment, the captured video was processed and, using the suggested approach, it was determined whether the individual was behaving abnormally. The results are listed in order of data processing to view the performance of each component of the approach.

In visualisations, colour of the elements indicates difficulty of distinguishing one element from others. The colour gradient used can be seen in Fig. 4. In contrast, the histogram of the average branch splitting length is shown in Fig. 5, where all histograms have an exponential distribution. Comparison of

the histograms of parameter Group A and Group B shows that for Video 1 and Video 3, the data of Group B are more diffused.

The graphs of the data obtained in Video 1 (see Figs. 6 and Fig. 7) show that the data are arranged as a single cluster with a centre at the beginning of the coordinates, which is explained by the fact that some people are at rest, standing or moving very slowly. Majority of the observations indicate that the context of Video 1 is slow motion or resting (based on the



Fig. 4. Colour gradient (from red to green), which is then used for visualisation to convey the average length of branch splitting. The colour lines represent the n^{th} measurement: a) the easiest element to detach from the set, b) the most difficult element to detach from the set, c) anomalous observation, d) normal observation.

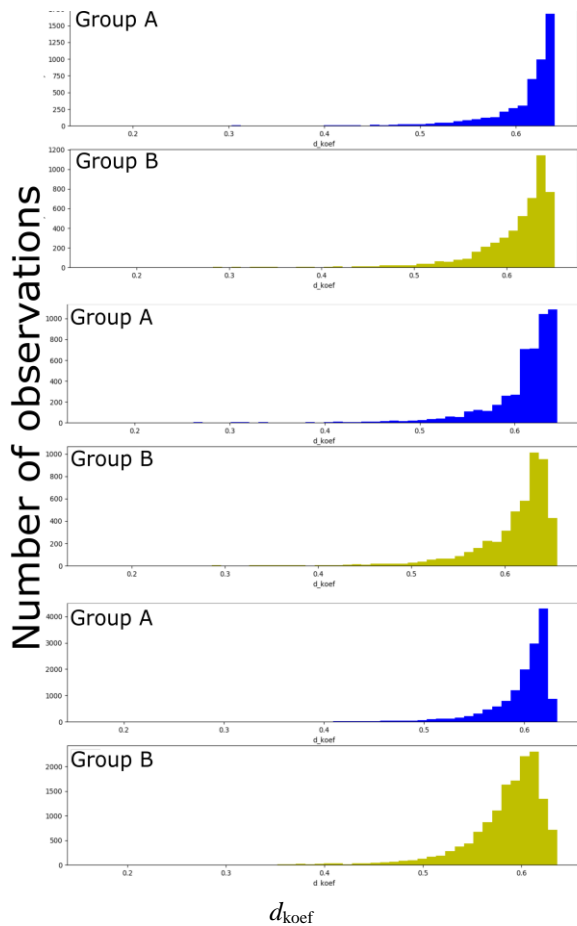


Fig. 5. Histograms of middle branch splitting length. Blue for Group A histogram, yellow for Group B.

acceleration graph in Fig. 7); abnormal behaviour occurs at a steady speed n ; most often acceleration occurs.

As Video 1 is designed with the idea that the resulting data will mimic anomalous behaviour, it can be observed that the

data are significantly different from other, so the system decodes them as anomalies.

Observations of Video 1 with anomalous behaviour can be seen in Fig. 8. Compared to other sets of graphs (see Figs. 8 and 9), it can be observed that the anomaly detection method used is capable of separating points in dense cloud compaction from those distant from other points; consequently, they are respectively located in diffused point clouds.

The algorithm output of whether an observation belongs to a normal or anomalous set has no boundary and can be seen in the gradient transition of group B point, where the anomalous

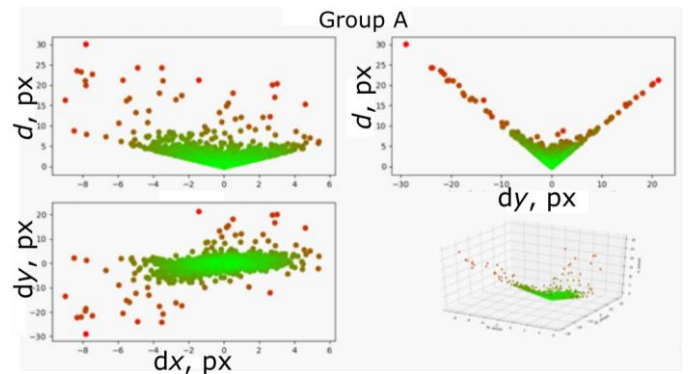


Fig. 6. Video 1, group A observations.

observations are in the same location as the normal observations. This may indicate a too high cut-off ratio for the branching coefficient method developed. To investigate whether this is the case, a data set with available ground truth is required. This phenomenon is not necessarily a bad one, since the normal set as abnormal within this use is less dangerous than the other way around.

Comparing the graphs obtained with all video data, it can be concluded that there is no significant difference in the results obtained (most of the people in crowd have normal behaviours) by the algorithm, which indicates that the results are obtained depending on the context.

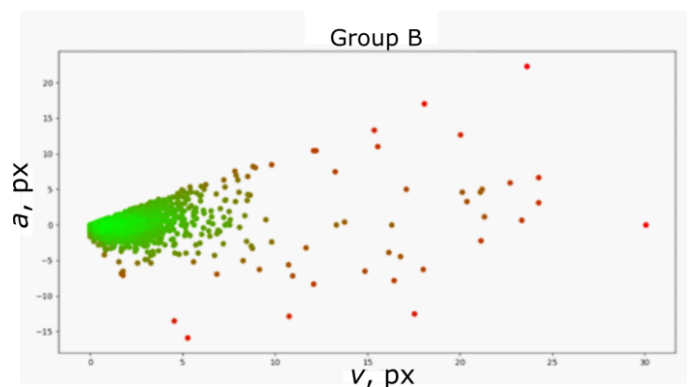


Fig. 7. Video 1, group B observations.

Note that in Video 3 (see Fig. 9) the score on the graph is higher than in the other videos. Fig. 9 shows that Video 3 scatter at a peak value is wider than that of Video 1 and Video 2 (Fig. 8), leading to the conclusion that along with the

increase in the number of pixels used to detect anomaly, scattering changes its shape.

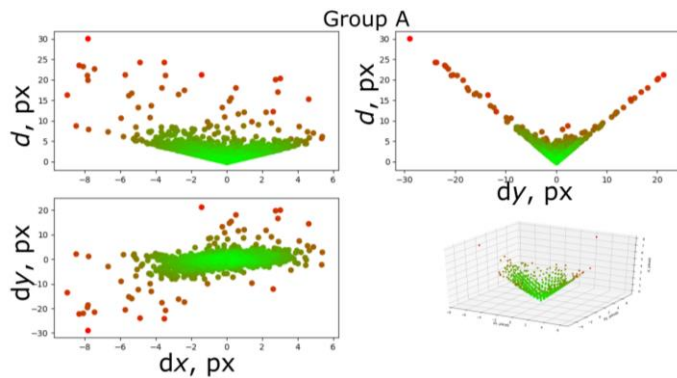


Fig. 8. Video 2, group A observations.

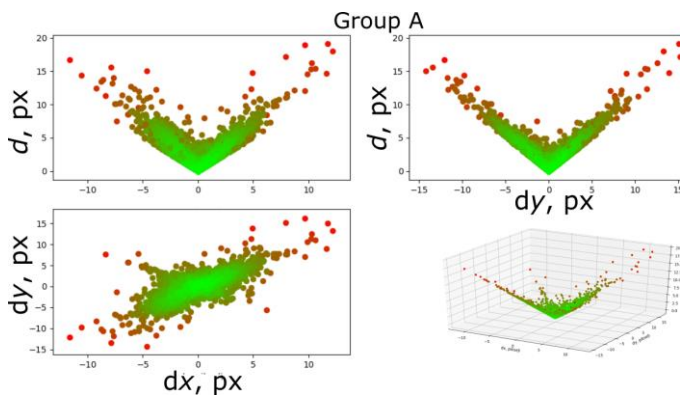


Fig. 9. Video 3, group A observations.

VI. CONCLUSION AND FURTHER RESEARCH

The developed crowd anomaly detection algorithms extract parameters from an image; anomaly detection is performed using already existing anomaly detection algorithms. The results show that unsupervised machine learning allows for context-dependent anomalies, whereas emotional states provide additional information and reduce the number of anomalies, thus facilitating the work of the crowd observer.

To further improve the crowd behaviour detection methods, it is necessary to create a data set for testing algorithms. As of the authors' knowledge, no well-organised and directly usable data set is available, but it could significantly improve the ability to validate and compare algorithms.

To capture the emotional state of humans from the images, it is suggested to use a camera system consisting of a static basic surveillance camera and dynamic (can rotate without changing the focal point location of the camera, without translating) cameras equipped with an optical lens system that automatically zooms in. The goal of the camera is to provide a high-quality image of the human face that can be used to recognise emotions from behaviours and facial expressions by applying existing emotion retrieval methods. Such a system allows for the creation of a new CCTV system on the existing surveillance system. The camera is designed to record the facial expression of individuals in the crowd when the

obtained basic camera images show anomalous individual behaviour; the algorithm proposes a dynamic camera to capture an image of the anomalous individual's facial expression.

Future research is related to improving and adapting each stage of the proposed implementation to the requirements of the specific field, replacing or customising the open access projects YOLO and SORT, as well as the general application training set COCO with solutions tailored specifically to human and crowd dynamics. Emotion detecting package should be added. This, in itself, could be a challenge because while the face is in front, the accuracy of emotion recognition is sufficient, even if the crowd is dense and it is not possible to get a face even with a rotating camera, the accuracy of emotion recognition could be reduced.

Further research is needed regarding the human environment factors that can influence human behaviour in order to select characteristics of human behaviour and to determine cause-effect relationships that are normal or abnormal.

REFERENCES

- [1] IHS Markit video surveillance, "Top Video Surveillance Trends for 2017," *Mark. Week*, pp. 14–18, 2016.
- [2] M. Andersson, J. Rydell, and J. Ahlberg, "Estimation of Crowd Behavior Using Sensor Networks and Sensor Fusion," *12th Int. Conf. Inf. Fusion*, Aug. 2009.
- [3] A. Kondrova, *Kognitīvo procesu sistēma*. Rīga, 2010.
- [4] H. U. Keval, "Effective, Design, Configuration, and Use of Digital CCTV," Doctoral thesis, University College London, 2009.
- [5] M. W. Baig, E. I. Barakova, L. Marcenaro, M. Rauterberg, and C. S. Regazzoni, "Crowd Emotion Detection Using Dynamic Probabilistic Models," in *From Animals to Animals 13*, A. P. del Pobil, E. Martinez-Martin, J. Hallam, E. Cervera, A. Morales, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014, pp. 328–337. https://doi.org/10.1007/978-3-319-08864-8_32
- [6] W. Little, *Introduction to Sociology – 1st Canadian Edition Edition*. pp. 141–168, 2014.
- [7] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, "Unsupervised Detection of Anomalous Sound based on Deep Learning and the Neyman-Pearson Lemma," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212–224, Jan. 2019. <https://doi.org/10.1109/taslp.2018.2877258>
- [8] D. Chakrabarty and M. Elhilali, "Abnormal Sound Event Detection Using Temporal Trajectories Mixtures," *2016 IEEE Int. Conf. Acoust. Speech Signal Process.*, pp. 216–220, 2016. <https://doi.org/10.1109/ICASSP.2016.7471668>
- [9] A. G. Abuarafeh, M. O. Khozium, and E. Abdrabou, "Real-Time Crowd Monitoring Using Infrared Thermal Video Sequences," *J. Am. Sci.*, vol. 8, no. 3, 2012.
- [10] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, vol. 41, no. 3, pp. 1–58, Jul. 2009. <https://doi.org/10.1145/1541880.1541882>
- [11] H. Parvin, H. Alizadeh, and B. Minati, "A Modification on k -Nearest Neighbor Classifier," *Global Journal of Computer Science and Technology*, vol. 10, no. 14, pp. 37–41, 2010.
- [12] R. Chalapathy and S. Chawla, "Deep Learning for Anomaly Detection: A Survey," pp. 1–50, 2019.
- [13] Y. Chen, J. Qian, and V. Saligrama, "A New One-Class SVM for Anomaly Detection," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3567–3571, 2013. <https://doi.org/10.1109/ICASSP.2013.6638322>
- [14] F. T. Liu and K. M. Ting, "Isolation Forest," *2008 Eighth IEEE Int. Conf. Data Min.*, pp. 413–422, Dec. 2008. <https://doi.org/10.1109/ICDM.2008.17>
- [15] C. Chen, Y. Shao, and X. Bi, "Detection of Anomalous Crowd Behavior Based on the Acceleration Feature," *IEEE Sens. J.*, vol. 15, no. 12, pp. 7252–7261, 2015. <https://doi.org/10.1109/JSEN.2015.2472960>

- [16] J. Z. C. Lai and T. J. Huang, "Fast Global k -Means Clustering Using Cluster Membership and Inequality," *Pattern Recognit.*, vol. 43, no. 5, pp. 1954–1963, 2010. <https://doi.org/10.1016/j.patcog.2009.11.021>
- [17] M. M. Breunig, H. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying Density-Based Local Outliers," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data (SIGMOD '00)*, pp. 1–12, 2000. <https://doi.org/10.1145/342009.335388>
- [18] M. Markou and S. Singh, "Novelty Detection: A Review – Part 1: Statistical Approaches," *Signal Processing*, vol. 83, no. 12, pp. 2481–2497, Dec. 2003. <https://doi.org/10.1016/j.sigpro.2003.07.018>
- [19] W.-L. Hsu, Y.-C. Wang, and C.-L. Lin, "Abnormal Crowd Event Detection Based on Outlier in Time," in *2014 Int. Conf. Mach. Learn. Cybern.*, vol. 1, pp. 359–363, 2014. <https://doi.org/10.1109/icmlc.2014.7009142>
- [20] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent Flows: Real-Time Detection of Violent Crowd Behavior *," *2012 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.*, pp. 1–6, 2012. <https://doi.org/10.1109/CVPRW.2012.6239348>
- [21] C.-L. L. Wei-Lieh Hsu, Yu-Cheng Wang, "Spatio-Temporal Anomaly Detection in Crowd Movement Using SIFT," in *Proc. 2014 Int. Conf. Mach. Learn. Cybern.*, 2014.
- [22] Y.-T. Matsuda, T. Fujimura, K. Katahira, M. Okada, K. Ueno, K. Cheng, and K. Okanoya, "The Implicit Processing of Categorical and Dimensional Strategies: An fMRI Study of Facial Emotion Perception," *Front. Hum. Neurosci.*, vol. 7, no. September, 2013. <https://doi.org/10.3389/fnhum.2013.00551>
- [23] R. Fan, K. Xu, and J. Zhao, "Higher Contagion and Weaker Ties Mean Anger Spreads Faster Than Joy in Social Media," pp. 1–23, 2016.
- [24] L. Coviello, J. H. Fowler, and M. Franceschetti, "Words on the Web: Noninvasive Detection of Emotional Contagion in Online Social Networks," in *Proc. IEEE*, vol. 102, no. 12, pp. 1911–1921, 2014. <https://doi.org/10.1109/JPROC.2014.2366052>
- [25] D. Mehta, M. F. H. Siddiqui, and A. Y. Javaid, "Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality," *Sensors*, vol. 18, no. 2, pp. 416, 2018. <https://doi.org/10.3390/s18020416>
- [26] S. Petrovica and H. K. Ekene, "Emotion Recognition for Intelligent Tutoring," *CEUR Workshop Proc.*, vol. 1684, 2016.
- [27] Z. S. Hippe, J. L. Kulikowski, T. Mroczek, and J. Wtorek, "Emotion Recognition and Its Applications," *Adv. Intell. Syst. Comput.*, vol. 300, no. July, 2014.
- [28] K. Glanz, "Social and Behavioral Theories 1. Learning Objectives," *Off. Behav. Soc. Sci. Res.*, 2005.
- [29] G. Bradski, "The OpenCV Library," *Dr. Dobb's J. Softw. Tools*, vol. 120, pp. 122–125, 2000.
- [30] J. Redmon and A. Farhadi, *YOLO v.3*, 2018.
- [31] J. Redmon, "Darknet: Open Source Neural Networks in C," *Pjreddie*. [Online]. Available: <http://pjreddie.com/darknet/>
- [32] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft COCO: Common Objects in Context," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 8693 LNCS, no. PART 5, pp. 740–755, 2014. https://doi.org/10.1007/978-3-319-10602-1_48
- [33] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in *2016 IEEE International Conference on Image Processing (ICIP)*, 2016, pp. 3464–3468. <https://doi.org/10.1109/ICIP.2016.7533003>
- [34] T. Simon, H. Joo, I. Matthews, and Y. Sheikh, "Hand Keypoint Detection in Single Images Using Multiview Bootstrapping," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4645–4653, 2017. <https://doi.org/10.1109/CVPR.2017.494>
- [35] Z. Cao, T. Simon, S.-E. Wei, and Y. Sheikh, "Realtime Multi-Person 2D Pose Estimation Using Part Affinity Fields," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1302–1310, 2017. <https://doi.org/10.1109/CVPR.2017.143>
- [36] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional Pose Machines," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, pp. 4724–4732, 2016. <https://doi.org/10.1109/CVPR.2016.511>

Glorija Baliniskite is a Doctoral student of the study programme "Computer Systems" at Riga Technical University. She obtained the Master's degree in computer systems in 2019 at Riga Technical University, Latvia. The topic of her Doctoral Thesis is related to surveillance system automation. Ms. Baliniskite's research interests include anomaly detection in crowd, emotional state and surveillance system automation.
E-mail: glorija.baliniskite@edu.rtu.lv

Egons Lavendelis is an Associate Professor and Researcher in the area of artificial intelligence. Egons defended his Doctoral Thesis in 2009 at Riga Technical University (RTU) in the area of agent-oriented software engineering proposing the MASITS methodology. After the Doctoral studies, he has been working as a Senior Researcher and Assistant Professor at the Department of Artificial Intelligence and Systems Engineering. Egons' research interests are multi-agent systems, agent-oriented software engineering, communication among autonomous entities, including semantics, different applications of agent paradigm and various artificial intelligence techniques; lately his attention has been devoted to the use of multi-agent systems for integration of autonomous robots into collaborative multi-robot systems. Currently, Egons is delivering courses related to artificial intelligence and databases at RTU. He is the author of more than 30 publications and book chapters at various international conference proceedings and scientific journals.
E-mail: egons.lavendelis@rtu.lv
ORCID iD: <https://orcid.org/0000-0001-9912-035X>

Mara Pudane is a Researcher and Doctoral student of the study programme "Computer Systems" at Riga Technical University. She obtained the Master's degree in computer systems in 2013 at Riga Technical University, Latvia. During her Master studies, she started to work as a Research Assistant at the Department of Artificial Intelligence and System Engineering (Riga Technical University). The topic of her Doctoral Thesis is related to human group behaviour imitation with focus on affective factors. Ms. Pudane's research interests include human modelling, multi-agent systems and affective computing.
E-mail: mara.pudane@rtu.lv
ORCID iD: <https://orcid.org/0000-0002-9188-5478>