Spatio-temporal Classification of Aggression in Video Surveillance using Optical Flow History

Kevin Hardeman

Bachelor Thesis
Artificial Intelligence
7.5 ECTS

April 9, 2018

Supervisors

dr. S.M. Stuit (UU)	dr. G.J. Burghouts (TNO)
dr. F.P.M. Dignum (UU)	dr. R.P.J. Nieuwenhuizen (TNO)

Utrecht University

TNO innovation for life

Abstract

The role of automatic surveillance in modern society is rapidly increasing. While most of these systems operate by making a judgment of aggression based on two frames, in this paper we explore the effect of adding more frames to the quality of detecting aggression in video surveillance. The premise of doing this is that the system must be able to run in real-time, preferably using as little computing power as possible. We evaluate an algorithm by TNO that was developed with this goal in mind. The results show a significant increase in detecting aggressive instances when more frames are added. The tests were run on a dataset containing instances of street aggression supplied by the Dutch police. A history of 1.5 seconds worth of frames was found to deliver the best results on the dataset.

1 Introduction

Occurrences of aggression and violence between individuals have been present since before people as we know them were even around [14]. Thousands of years later, in modern society, the problem still persists. The most aggressive incidents today happen in night-life settings. Researchers at the Trimbos Institute in the Netherlands found out that the amount of violence is significantly larger when individuals are in a mental state where they are less inhabitant of direct impulsive reactions to the environment, and in situations where a greater level of anonymity is felt [17]. In night-time entertainment areas, the presence of these factors is highest, and often combined due to above average alcohol intake and lower-light settings. Thereby, these conditions serve as a breeding ground for violent incidents. Municipalities struggle to minimize the quantity of such incidents. Police and law enforcement teams try their hardest but evidently can't be present at all places that are at risk of aggressive incidents. If and when the authorities are alerted, they often appear late on the scene. The violence may have already subsided and the offenders may be long gone.

With the aim of detecting more violent incidents as they happen, increasingly more use is being made of security cameras. The footage of these cameras is usually shown in a control room where all different camera streams are watched by security staff members, who can alert their colleagues that are out on the street when instances of violent behaviour are noticed. These control rooms are a worthwhile addition in reducing public violence, though not a perfect solution to this problem. In 2004, G.J.D. Smith [15] examined some of these control room situations and found that in the study, control room operators were expected to make shifts with a duration of 6 to 8 hours. During these shifts, an operator is expected to monitor up to 15 video streams on his own. Because of the scarcity of incidents happening, losing focus and getting bored or distracted is very likely for personnel. This phenomenon is also known under the term of ”operator fatigue” [8]. This led to the fact that even though in theory all camera screens would be watched 24/7, some break-ins, assaults and car thefts had been going on due to lack of focus, or because the other screens were being watched.

Beside these shortcomings, the control room operators reported more suspicious behaviour in people of specific ethnic groups, genders and clothing styles. In recent years, many of these control room situations are aimed to be strengthened with the help of computer algorithms both to improve the likelihood of detecting the signals and to make the job of the control-room personnel easier and more efficient, or even redundant in the long run.

This research will focus on an algorithm for violence detection developed by TNO on behalf of a municipality in the Netherlands in cooperation with the Dutch police, to be used as an extra aid for the control room personnel. It is designed to be directly applicable to the security cameras already in place and to work in real-time, and to work in such a way that it will be able to guide the control room operator in its task.

Due to the difficulty of developing general surveillance algorithms, practical systems usually consist of a combination of algorithms, which are selected on a case by case basis. Different researches use different sorts of tactics in tackling these issues [5, 9]. The most relevant and influential of them will be laid out in the background section of this paper.

Furthermore, a custom-made surveillance system will be explained in this paper. This system is specifically made for the real-time detection of aggression, applied in settings where continuous detection and tracking of individuals is not possible due to crowdedness and low resolution of existing security cameras on which this algorithm will be applied. Because of this, the algorithm in question uses different features to try and detect aggression, namely information obtained from the movement of pixels from one video frame to the next. These pixel-movements have a magnitude and a direction, and are therefore called optical flow vectors. This will be laid out more in-depth in the methods-section of this research.

In some existing algorithms [3, 13, 18], a better performance is achieved when more than two consecutive frames were taken into account, thus using more information to be able to achieve better detections. Inspired by these findings, this paper will research whether taking multiple frames prior to the current one into account also has an effect on correctly classifying aggression. The goal of this research will be to see if this is also the case when analyzing aggressive behaviour.

Combined with the need for a real-time algorithm based on optical flow, the ultimate goal of this particular research is to find the optimal parameters for training a model for aggression detection, with the most focus on the amount of frames that is taken into account. This history-parameter will be tweaked to find the amount of time and frames that has the optimal effect on correctly classifying aggression. Hereby it is hypothesized that taking history into account in the first place will have a positive effect on the model, as opposed to only examining the optical flow vectors of the current and past frame.

In the rest of this paper, a brief history and context of computer vision algorithms focused on behaviour and anomaly detection will be outlined in section 2 (Background). In section 3 (Methods), the basis of the algorithm will be laid out. Specific details about the settings of parameters like the amount of optical flow history used will be discussed in section 4 (Parameters). In section 5 (Results) the performance of this algorithm using the most successful parameters are shown and will be compared with the performances of existing algorithms. Finally, section 6 (Discussion) will focus on the practical implications, and improvements and variations on this research will be considered and discussed.

2 Background

Many different approaches have been tried and researched for detecting human behaviours from visual footage. A few of the most popular and commonly used methods will be laid out here. In the rest of this paper, information extraction from a static single-viewpoint camera-position is assumed. Behaviour detection in moving or multiple cameras take different approaches which go beyond the scope of this paper.

2.1 Detecting people

The first step of every behaviour detection algorithm is to determine the areas of focus in an image or video-frame. This helps to reduce the computational costs and time needed for in-depth analysis. With the specific goal of human behaviour detection in mind, this step aims to ideally extract the precise location of humans in video footage, sometimes going as far as extracting the position of an individual or even body parts, focussing on the movement of limbs or heads (e.g. in the case of emotion detection). A popular low-level method for achieving this goal is background subtraction. It compares the current frame to a reference image of the background. The parts of a frame with the most difference compared to the reference image are then selected, with the underlying assumption that these areas showing the biggest change are the main interest points [16]. A more high-level method of extracting the relevant parts of an image is to use Space-time interest points [6]. This method detects where the largest movements between frames take place. The power of this method is that multiple frames can also be taken into account, potentially resulting in a better estimation of correct interest points, later to be used as the main areas for feature extraction.

2.2 Feature extraction

After a selection of relevant regions in the frames has been made, the next step is to analyze the extracted regions in more depth, aiming to obtain features that can be fed into a learning algorithm. Within the research area of behaviour detection, there are different approaches to this problem.

At the end of the 21st century when the field of behaviour detection was still emerging, it was assumed that successful recognition of human actions would only be possible using the 3D posture of an individual could be determined [12]. The tracking of people was also made possible using this approach. Obtained tracks, together with the posture and movement of an individual's body parts could now be analyzed. This method had its limitations, especially in low-resolution footage or when occlusions occurred, making it difficult to successfully track people in a robust way. In 1996, a breakthrough paper by Bobick and Davis [2] was published that claimed behaviour recognition was possible in other ways than by analyzing human posture. The inspiration came from the insight that humans can identify behaviour by recognizing low-level movement features without knowing the exact positions of the limbs. This point was proven by training an algorithm on blurred video data of which no human postures could be detected by humans nor a computer. Action recognition on the basis of this data did however prove to be possible [2].

This finding, combined with progressions in the field of data analysis, gave rise to a new philosophy of 2D view-based approaches for selecting other sorts of features. This approach laid more focus on the learning algorithm instead of on the pre-processing of data, which made it possible to essentially give a learning algorithm less precise information to achieve better results in the end.

For a long time, the differences in visual information between two frames were considered. In 2001, Bobick and Davis [3] introduced the idea of comparing and analyzing more than two frames to obtain more information, resulting in what they called a Motion History Image (MHI). With this method, motion information from multiple frames is merged into one image, which was fed into a learning algorithm. This sparked inspiration for more methods that used history information, a successful and promising result in fall detection using motion history was achieved by Rougier et al. in 2007 [13].

2.3 Visual aggression detection

Whereas much of the methods discussed above have already been used on relatively simple human actions like walking, running, falling or shaking hands, less research had existed on the detection of more complicated human behaviour like aggression until the early 2000s. In relatively many researches on aggression detection, the focus laid on the combination of auditive and visual information, which is not viable in the set-up in this paper due to the lack of auditive information in the security cameras already in use [1]. Furthermore, most of the datasets used are violent scenes from movies, sometimes going as far as using explosions and blood as classifiers to detect aggression [7].

The most recent research that tries to classify non-scripted aggression was done by a Dutch research team and focused on classifying aggression inside train compartments [18]. In this research use is made only of a single viewpoint, using only visual information in a real-world setting (namely train compartments) which makes it similar to the current paper. The Dutch researchers used a bagof-words approach to classify aggressive behaviour. In this approach, a video sequence is represented as a collection of spatial-temporal ”words” and analyses are made in the same way as in the linguistic domain, where predictions are made on the basis of word orders and occurences in a corpus used for training [10]. However, classifications like this usually don't take feature history in mind.

3 Methods

This section consists of a description of the behaviour detection algorithm used in this research, complete with the default parameters that were chosen as the starting point and the reasoning behind those default parameter values.

3.1 Specifications

The algorithm was made and run on the academic version of Matlab 2017a on a 64 bit-version of Windows 10. The hardware used for this was an x64-based i7-5500U 2.40 GHz CPU equipped with 8 GB of RAM.

3.2 Data and annotations

The data consists of 106 security camera videos in AVI-format. The colored data obtained from the security cameras is transformed to grayscale images to reduce computational costs. From these videos, 208 intervals were selected, with a length ranging from 2 to 57 seconds. 54 of these intervals contain aggressive behaviour.

For all different camera-angles present in the videos, a region of interest (ROI) has to be specified. This ROI consists of the area where the presence of people is expected (pavements, stairs, etc). The area of the frames to be analyzed will later be confined to this region. This way, moving parts of buildings like doors or sunscreens are prevented from causing false positives prior to any analysis. Furthermore, in all camera-perspectives, the size of a person in the foreground as well as in the background were defined. With this information, the estimated size of people is taken along in the analysis. This serves as an extra way of preventing false positives, because this decreases the probability of plastic bags or flying birds being classified as aggression. On the basis of all former information, a grid of bounding boxes is laid over the ROI (figure 1).

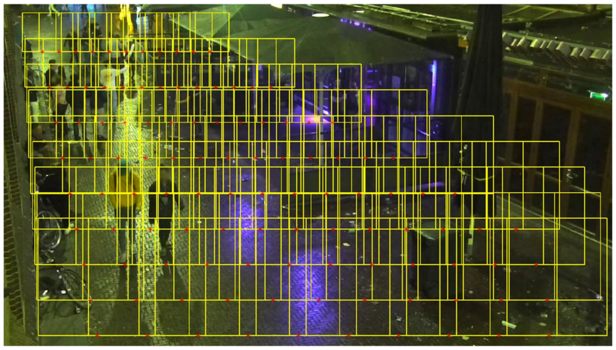

Fig. 1 - Example of a bounding box-grid (stride-parameter 0.5)

Annotations in the form of bounding boxes are indicated around the area containing aggression in every relevant frame. For this algorithm, the choice was made to annotate the full area of violence. For example, if two people were observed fighting, a bounding box was placed around the two people (in contrast to for instance only annotating striking arms or kicking feet). This way, the information about the full area of aggression is taken into account. The duration of kicks and hits is also very short, so an algorithm taking optical flow history into account using these annotations would be difficult or impossible to achieve.

3.3 Data-selection

When all data would be taken into account, a lot of computing time and memory space would be needed to analyse an abundance of motionless areas in the data. This could pose a threat to the goal of running the algorithm in real-time. To this goal, smart selections are made to extract the important data. To achieve this, two selection-methods were applied to pre-process the data by extracting the most promising bounding boxes.

3.3.1 Non maximum-suppression

The first step in removing uninformative parts of the data is done by applying a low-level analyzing technique called normalized cross-correlation (NCC).

This method limits the number of overlapping bounding boxes that are covering the same action or object. By subtracting the current frame from the previous frame we achieve a spatial mapping of the difference measured between these two frames. Some parts of the frame will differ more than other parts. The bounding boxes with areas that have changed are the points that have to be analyzed in more depth. Using a threshold, a selection of bounding boxes is made out of the original grids. The stricter the threshold is selected, the less bounding boxes are left over. The bounding boxes containing the least movement are the first ones to be excluded. This is the way NCC makes a pre-selection of bounding boxes to be used in the next steps, and ignores the non-important or static areas.

3.3.2 Detector score

The next step is to also involve the aggression-annotations. This is achieved by another NCC-threshold for making an even stricter selection of boxes. This time however, extra information from the aggression-annotations are involved. This information consists of how many leftover boxes from the former step would overlap with the aggression-annotations. From this information, a ratio of potential correct and false classifications is extracted. In this step, a new threshold can be set.

Selecting the right threshold revolves around the question of whether it is more important to achieve as many true positives as possible at the risk of allowing more false positives, or if it is more important to limit the amount of false positives as much as possible. Where the focus lies depends on the practical application of the algorithm. With the applied use of this particular research in mind, it was decided that there is equal importance in maximizing true positives as there is in minimizing false positives. When a false detection is made by the algorithm, it will be verified by a surveillance camera operator. On the other hand, while keeping this balance, the more subtle occurrences of aggression could be missed by the algorithm. This should not be a problem because the main focus lies on the quick detection of the most severe occurrences of aggression.

3.4 Features

3.4.1 Optical flow

The contents of the selected bounding boxes can then be qualitatively analyzed. Specifically, an analysis of the movement that occurs in these boxes is made with the use of the Lucas-Kanade differential method for optical flow estimation. This method estimates in what direction certain pixels have moved from one frame to the next. This is done by taking the 3x3-pixel neighbourhood of a pixel in the first frame into account, and looking for the same set of pixels in the new frame that is most similar to that region in the original frame. This way, it can be established which way a pixel has moved from one frame to the next. An optical flow vector is established, consisting of both the length that the pixel has moved as the (x,y)-directions. This method is applied to a number of pixels in every bounding box, the number is customizable by a parameter. In this experiment, from 64 pixels per bounding box, the optical flow was analysed. These points were laid out in a grid evenly divided over every box. The optical flow vectors are the data points that are central to this algorithm.

3.5 Learning methods

The obtained optical flow features are turned into a histogram using a random forest. This data is then processed by a support vector machine with the aim to separate the aggression-instances from the other data. This method will be explained in more detail in the rest of this section.

3.5.1 Random forest

The information that is obtained so far are bounding boxes with their corresponding optical flow-vectors, of which some are annotated as containing aggression. With this information, future predictions about whether new data is aggressive or not can be made. To achieve this goal, the optical flow vectors are put into a random forest (RF) with the implementation of Breiman and Cutler [4]. The forest consists of 4 random trees with 16 leaves on each tree(see section 4.2 for more information). After some experiments, this proved to be the optimal cut-off point between accuracy of predictions and time efficiency. On every node in the RF, a split is made on the basis of a random threshold, by nature of a RF. This threshold is based on the movements inside a bounding box, meaning that this division can be made on the basis of the direction or size of the optical flow vectors, the differences in directions of movements within a bounding box, or a combination thereof. This process is visualized in figure 2a.

Once the information in a bounding box is put through the random forest, the output data is then reshaped into a histogram containing the information from the RF (figure 2b). This is done to have a more compact representation of the data, and most importantly, so that the support vector machine (SVM) in the next step can process this data.

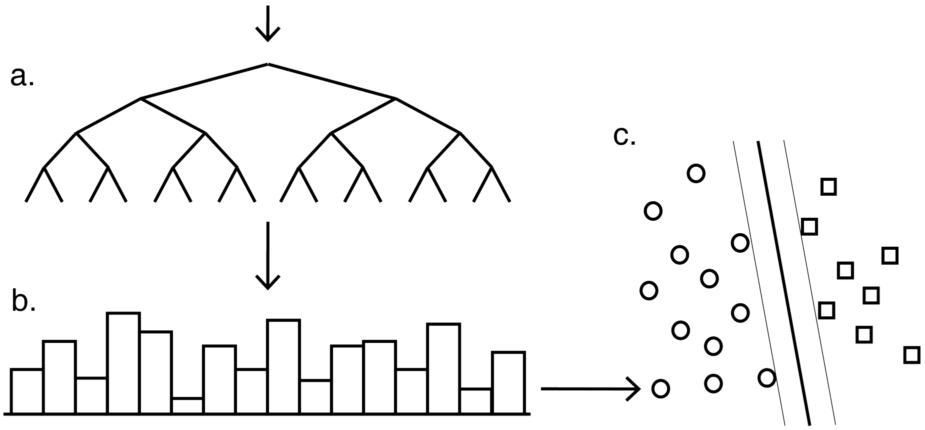

Figure 2. a: Random tree, of which four were used as a random forest. The optical flow-vectors from the bounding boxes are used as inputs, the output is the data separated by the random forest. b: The output from the random forest is transformed and stored in the form of a histogram. c: Support vector machine splitting the training-data, and used to make predictions on the basis of similarity to training data.

3.5.2 Support vector machine

Each bounding box obtains its own histogram from the random forest. From this information, a distinction can be made between the histograms that are aggressive and that are non-aggressive. To achieve this task, an SVM with a χ^2-kernel was used, which aims to split the two categories in the best way possible.

This way, when a new histogram is given as a new input, a classification can be made on the basis of similarity to the existing annotated histograms. If the new histogram is very similar to other aggression-histograms, it will be classified as aggression and vice versa for non-aggressive histograms.

3.5.3 Performance measure

For every leaning problem, a performance measure has to be chosen. This choice depends on whether the focus lies more on preventing false positives (FPs) or detecting the most true positives (TPs).

When minimizing the number of FPs is the main concern, the most common performance measure used is precision. Precision is the number of correct classifications out of all returned positives. This works by the following formula:

$$precision = \frac{TPs}{TPs + FPs}$$

When the goal is to maximize the number of TPs that is correctly classified, the most common performance measure is recall. This corresponds to the fraction of aggression occurrences that are successfully classified. In this calculation, all occurrences of aggression in the data are taken into account, including the false negatives (FNs):

$$recall = \frac{TPs}{TPs + FNs}$$

The practical requirements of the municipality are in the middle ground: to be able to detect as many cases of aggression (TPs), while also preventing FPs. Because of this practical middle ground, both are considered equally important. This is why a performance measure was chosen that is a weighted average between precision and recall: The F1-score:

$$F1 = 2 * \frac{precision * recall}{precision + recall}$$

4 Parameters

Beside the optical flow history-parameter, other customizable parameters are also used in this algorithm. This section gives a short overview regarding choices for other parameters that were made, and the effect of these parameters have on the results.

4.1 Bounding box-size

The first variable settings are the size and ratio of bounding boxes. The parameter-values that were chosen were bounding boxes with the height of a person used as both the height and width, resulting in a square bounding box. The height of a person was estimated to be about 1.8 metres on average on the Dutch dataset used. The size of the bounding box thus resulted to be a square of 1.8 by 1.8 metres. The width of a box is larger than the width of one person. This ratio was chosen because aggression involves more than one person. Aggressive behaviour usually takes place in a horizontal direction. By making the width much larger than the width of one person, the desired effect was achieved.

4.2 Random forest-parameters

For the random forest, the total number of trees and the number of leaves per tree are variable parameters. Because optical flow is the only parameter that is used as a feature, relatively little variation is possible. In other random forest-algorithms that don't make use of an SVM to train a model and consider a lot of different features, sometimes a lot of trees are needed to obtain an effective size for the random forest [11]. Because this algorithm only covers optical flow vectors as features, using an SVM is supposedly more effective for this classification problem than using many trees [14]. For this reason, relatively little trees will be used in this case. With this reasoning and after some testing, it was decided that the optimal cut-off point between accuracy and time-efficiency lies with 4 random trees and 16 leaves per tree in the forest.

4.3 Optical flow history

Optical flow history is the main parameter in this research. Hypothesized is that the more optical flow history in a bounding box is taken into account, the more history there is to be trained on, so in the end a more educated model can be made, and classifications of aggression could be more accurate in the end. A note to take with this approach is that not all occurrences of aggression are equal. For instance, one instance can consist of one person swiftly giving a strike at the head of another person, while another can consist of a whole group of people running around and making multiple aggressive movements. Both examples are instances of violence that ideally are equally likely to be detected by the algorithm. It is likely however that when taking into account a long history of five seconds, in the swift strike-example the first few seconds do not yet contain any aggressive behaviour. The base value of this parameter is to take into account only the current and last frame. Then, more frames will be added to the optical flow history and eventually an optimal value will be found that will be reported and interpreted in the next section.

5 Results

This section discusses the results that were obtained in the research. All of the analyses were done in real-time at minimally 15 times the required frame-rate of 15 frames per second using the hardware earlier specified. This implies that the algorithm can run on much less powerful hardware, ensuring low costs for implementing the algorithm in practice.

5.1 Data-selection

For the pre-processing step of non maximum-suppression, a relatively high cut-off point was proven to be most successful, conserving around 85% of all bounding boxes that were originally present. The detector score was chosen more strictly, only 20% of the bounding boxes that had passed the former selection were kept. These parameters were established by way of trial and error on multiple bounding box-sizes and proved to yield the best results.

5.2 Parameters

On a subset consisting of 15 intervals of which 5 aggressive, bounding box-dimensions with a width and height of 1.6 metres showed the best results, as can be seen in more detail in table 1.

Bbox-dim (m)	1.0	1.2	1.4	1.6	1.8
Precision	0,889	0,991	0,995	0,995	0,983
Recall	0,870	0,995	0,905	0,995	0,947
F1	0,880	0,993	0,948	0,995	0,964

Table 1 - Results of varying bounding box-dimensions

The optical flow history-parameter was finally tested using the full dataset and with the bounding-box dimensions of 1.6×1.6 metres. The results can be seen below in table 2.

History (s)	0.0	0.5	1.0	1.5	2.0
Precision	0,994	0,946	0,938	0,953	0,980
Recall	0,835	0,876	0,908	0,922	0,873
F1	0,907	0,910	0,923	0,938	0,923

Table 2 - Results of varying optical flow history

The optical flow history of 1.5 seconds has the highest F1-score and yields the best tested value of the history parameter on the used dataset.

6 Discussion

The goal of this research was to find out if there exists a relationship between the amount of optical flow history and the accuracy of predictions. As can be seen in the results, there is an undeniable improvement in performance when history was used for analysis in comparison to only comparing the last two frames. This result is in line with what was expected prior to the research, on the basis of results that were found earlier in fall detection using optical flow history [13]. This finding could prove to be a relatively easy way for related video-detection algorithms to achieve a better detection accuracy.

The data that was used was limited to an hour's footage, where years and years of video material exists from all over the world. Of course, when all footage that existed would be fed into the algorithm, a better model would be expected at the cost of a considerably larger training time.

Another possible addition to this algorithm that would contribute to better detections would be to add colour information to the optical flow-algorithm. With the addition of this information, the accuracy of establishing where a certain pixel has moved would be higher [8]. While this could increase the accuracy of this algorithm, it would take more computational time to compute this extra colour-data, which is the reason it was avoided in this research.

Furthermore, aggression is difficult behaviour that occurs in many forms. There is a striking difference in movements between the case of a person giving someone a strike to the head in comparison to a chaotic fight involving a large amount of people. Both cases were present in the used dataset and were marked as equally important behaviour with the same "aggression"-label. It might be interesting to see whether different parameters work better with specific sorts of aggressive behaviour. As was reasoned earlier, smaller bounding boxes and less history could be more effective for detecting kicks and hits, whereas a longer history could work better for classifying group aggression. If this would be the case, then possibly both models could be merged into one detector, giving an alarm if a high enough certainty is obtained by one of the models.

Note should be taken that some parameters have a direct influence on other parameters. For example, the dimensions of the bounding boxes were altered after deciding which bounding box-dimensions to take. If this sequence would have been reversed, it is possible that it would have resulted different combination of history and bounding box-dimensions. However, the order in which parameter-values were selected are good enough to be used by the Dutch municipality and police forces.

References

[1] E. Bermejo, O. Deniz, G. Bueno, R. Sukthankar, Violence Detection in Video using Computer Vision Techniques, 2011.

[2] A.F. Bobick, J.W. Davis, An Appearance-based Representation of Action, 1996.

[3] A.F. Bobick, J.W. Davis, The Recognition of Human Movement using Temporal Templates, 2001.

[4] D.R. Cutler, T.C. Edwards, Jr., K.H. Beard, A. Cutler, K.T. Hess, J. Gibson, J.J. Lawyer, Random Forests for Classification in Ecology, 2007.

[5] T. Ko, A Survey on Behaviour Analysis in Video Surveillance Applications, 2011.

[6] I. Laptev, T. Lindeberg, On Space-Time Interest Points, october 2003.

[7] J. Lin, W. Wang, Weakly-Supervised Violence Detection in Movies with Audio and Video Based Co-training, 2009.

[8] Matthews, G., & Hancock, P. A. (2017). The handbook of operator fatigue. CRC Press.

[9] Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. Computer vision and image understanding, 81(3), 231-268.

[10] Niebles, J. C., & Fei-Fei, L. (2007, June). A hierarchical model of shape and appearance for human action classification. In Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on (pp. 1-8). IEEE.

[11] Oshiro, T. M., Perez, P. S., & Baranauskas, J. A. (2012, July). How many trees in a random forest?. In International Workshop on Machine Learning and Data Mining in Pattern Recognition (pp. 154-168). Springer, Berlin, Heidelberg.

[12] J. Park, H. Oh, D. Chang, E Lee, Human Posture Recognition using Curve Segments for Image Retrieval, 1999.

[13] C. Rougier, J. Meunier, A. St-Arnaud, J. Rousseau, Fall Detection from Human Shape and Motion History using Video Surveillance, 2007.

[14] Sala, N., Arsuaga, J. L., Pantoja-Prez, A., Pablos, A., Martnez, I., Quam, R. M., ... & Carbonell, E. (2015). Lethal interpersonal violence in the Middle Pleistocene. PloS one, 10(5), e0126589.

[15] G.J.D. Smith, Behind the Screens: Examining Constructions of Deviance and Informal Practices among CCTV Control Room Operators in the UK, 2004.

[16] P. Spagnolo, T.D. Orazio, M. Leo, A. Distante, Moving object segmentation by background subtraction and temporal analysis, 2006.

[17] Trimbos Institute, Netherlands Institute of Mental Health and Addiction. 2014. *A*lcohol en uitgaansgeweld, De stand van zaken. `https://assets.trimbos.nl/docs/f29a86b3-1b4d-4049-a0f0-28dc3d6cca12.pdf` Trimbos Instituut, Utrecht

[18] Z. Yang, L.J.M. Rothkrantz, Automatic Aggression Detection inside Trains, 2010.