Towards Unsupervised Sudden Group Movement Discovery for Video Surveillance

Sofia Zaidenberg[1], Piotr Bilinski[1] and François Brémond[1]
[1]*Inria, STARS team, 2004 Route des Lucioles - BP 93, 06902 Sophia Antipolis (France)*
{*Sofia.Zaidenberg, Piotr.Bilinski, Francois.Bremond*}*@inria.fr*

Keywords: Event detection, Motion estimation, Anomaly estimation, Situation awareness, Scene Understanding, Group Activity Recognition, Stream Selection

Abstract: This paper presents a novel and unsupervised approach for discovering "sudden" movements in video surveillance videos. The proposed approach automatically detects quick motions in a video, corresponding to any action. A set of possible actions is not required and the proposed method successfully detects potentially alarm-raising actions without training or camera calibration. Moreover, the system uses a group detection and event recognition framework to relate detected sudden movements and groups of people, and provide a semantical interpretation of the scene. We have tested our approach on a dataset of nearly 8 hours of videos recorded from two cameras in the Parisian subway for a European Project. For evaluation, we annotated 1 hour of sequences containing 50 sudden movements.

1 INTRODUCTION

In video surveillance, one goal is to detect potentially dangerous situations while they are happening. We propose a detector sensitive to rapid motion or *sudden movements*, specifically for the context of underground railway stations security. We aim at detecting any action performed at high speed. Many dangerous actions that raise an alert for security are quick movements such as fighting, kicking, falling or sometimes running. Those actions are especially alarming when occuring within a group or people. Human attention is sensitive to rapid motion. When something moves at a higher speed than usual in our field of view, our attention is involuntarily drawn to that motion because that motion could be a sign of danger. Rash, unexpected movements in a subway station may cause a feeling of insecurity, even if they are actually not violent. Detecting such movements could allow the security operators to send a team to the affected area to reassure surrounding users.

Detecting sudden movements adds a cue for *stream selection* in the video surveillance system of a subway network. One problem that subway security comes up against is that there are too many cameras for too few human operators to watch. Stream selection is the process of intelligently selecting the most relevant cameras to display at any given moment. That way, the chances of missing an important event are reduced. False positives are not a big issue in such systems. Indeed, most of the time there are actually no true positives to show and the system still needs to select streams to display. In that case, we want to display the most relevant streams even if no event is detected. A camera in which a sudden movement is detected is more interesting to display to the security operator than a camera where the motion is slow. This system can help to avoid missing a situation requiring attention.

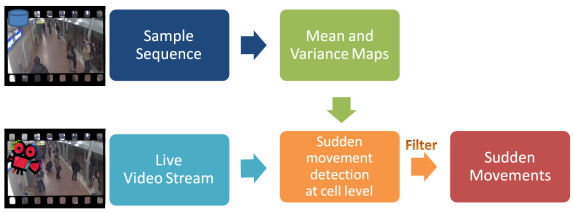

Figure 1: System overview.

Most approaches in the state of the art focus on action recognition – the labeling of a video sequence with an action label. Our problem is different in that we do not aim at recognizing which action is performed, but at detecting when an unusually rapid movement occurs. In fact, we do not (and can not)

have a database of actions that we want to recognize, on the contrary, we aim at discovering on the fly in a live stream any action that can be categorized as "sudden". Two "sudden" actions do not have to look similar, they can be visually as different as kicking and falling. On the other hand, we process live streams of data from video surveillance cameras in an underground railway station. We process videos with an arbitrary number of people anywhere in the camera view, in different views with different viewpoints and perspectives. This causes huge variations of motion and visual appearance of both people and actions in subway videos, and makes action recognition unadapted to the problem at hand. Therefore, we define our problem as sudden movement discovery. We tested and evaluated our approach on a dataset recorded in the Paris subway for a European project.

The strength of our approach, and the contribution of this paper, lies first in that it is *unsupervised*. It does not require training or calibration. This aspect is very important for application in large video surveillance systems composed of a large number of cameras. For instance cities such as London or Mexico City count up to 1 million of video surveillance cameras in their streets. It is not conceivable to do supervised training or calibration on such numbers of cameras. Even in a subway network such as in Paris, the number of cameras is several tens of thousands. Our approach allows for an automatic application on any number of cameras, without human intervention. Moreover, we do not impose any constraint on the camera placement. Video surveillance cameras are usually placed height up and thus see people with a strong perspective. The visual aspect of people and their motion is very different whether they are close under the camera or far from it. Our approach detects sudden movements in any typical video surveillance stream without constraints and without human supervision. Our approach, detailed in section 3, is summarized by figure 1. The sample sequence used to acquire mean and variance maps is not chosen by a human operator. It is a random sequence that can be automatically acquired at system set-up. The only constraint is to acquire it during presence hours and not when the scene is empty. Second, our approach can be combined with a more semantic interpretation of the scene, such as the system proposed by (Zaidenberg et al., 2012). This system detects groups of people in the scene and recognizes pre-defined scenarios of interest. The sudden movement detector is an addition to the group activity recognition system. When a sudden movement is detected inside a group's bounding box, a special event is triggered and can be used as it is, as an alert, or it can be combined into a higher-level scenario to provide a more accurate semantical interpretation of the scnene.

2 Related Work

There are many optical flow algorithms. Among the most popular are Large Displacement Optical Flow (LDOF) (Brox and Malik, 2011), Anisotropic Huber-L1 Optical Flow (Werlberger et al., 2009) and Färneback's algorithm (Farnebäck, 2003). Despite the first two implementations are more accurate than the third, however, operate very slowly on a single CPU (more than 100 seconds for 2 consecutive video frames of spatial resolution 640×480 pixels). Therefore, for extracting dense optical flow we use Färneback's algorithm (more precisely its implementation from the OpenCV library[1]) as a good balance between precision and speed ($1-4$ fps depending upon displacement between frames).

Optical flow algorithms are widely used for tracking. Recently, dense tracking methods have drawn a lot of attention and have shown to obtain high performance for many computer vision problems. Wang *et al.* (Wang et al., 2011) have proposed to compute HOG, HOF and MBH descriptors along the extracted dense short trajectories for the purpose of action recognition. Wu *et al.* (Wu et al., 2011) have proposed to use Lagrangian particle trajectories which are dense trajectories obtained by advecting optical flow over time. Raptis *et al.* (Raptis and Soatto, 2010) have proposed to extract salient spatio-temporal structures by forming clusters of dense optical flow trajectories and then to assembly of these clusters into an action class using a graphical model.

Action Recognition is currently an active field of research. Efros *et al.* (Efros et al., 2003) aims at recognizing human action at a distance, using noisy optical flow. Other efficient similar techniques for action recognition in realistic videos can be cited (Gaidon et al., 2011; Castrodad and Sapiro, 2012). Kellokumpu *et al.* (Kellokumpu et al., 2008) calculate local binary patterns along the temporal dimension and store a histogram of non-background responses in a spatial grid. Blank *et al.* (Blank et al., 2005) uses silhouettes to construct a space time volume and uses the properties of the solution to the Poisson equation for activity recognition.

Another related topic is *abnormality detection*. In papers such as (Jouneau and Carincotte, 2011) and (Emonet et al., 2011), authors automatically discover recurrent activities or learn a model of what is

[1] `http://opencv.willowgarage.com/wiki/`

normal. Thus they can detect as *abnormal* everything that does not fit to the leaned model. Additionally, (Mahadevan et al., 2010) propose a method to detect anomalies in crowded scenes using mixtures of dynamic textures, but they do not focus on sudden movement anomalies, which is our topic of interest. (Daniyal and Cavallaro, 2011) propose a supervised approach to detect abnormal motion based on labeled samples. This approach classifies the abnormal motion and uses contextual information such as the expected size of a person. (Xiang and Gong, 2005) and (Xiang and Gong, 2008) propose an unsupervised approach to detect anomalies but they also learn distinct behavior patterns, whereas we tend to detect one precise category of anomalies: sudden movements.

Apart from above cited fields, our problem relates to the fall detection problem. Indeed, falling is a fast motion and is one of the events we aim at detecting. Belshaw *et al.* (Belshaw et al., 2011) present a classifier *fall/no fall* but they use supervised machine learning techniques that require a training set of annotated videos.

However, as stated in the introduction (section 1), our problem has different goals and constraints. In the field of video surveillance, the camera is fixed, at a high viewpoint, providing strong perspective. Moreover, we can not make assumptions on the number of people, their location and the type of actions performed. Hence, we need an unsupervised approach, which is able to deal with strong camera perspective and does not require training. Our approach should not just classify videos as *sudden/non sudden*, but also localize sudden movements in the scene. Additionally, an unsupervised approach will allow for easy deployment on a large set of cameras.

3 Overview

As we have explained above we are interested in an unsupervised technique which does not require any ground truth for training nor calibration of a video camera. However, in that case we have to deal with problems related to camera view point. In particular, tracklets of a person who is close to a camera and is moving slowly can be seen in the image much bigger than tracklets of a person who is far away from the camera but it is moving quickly (see Figure 5). To solve this problem, we propose the simple but effective following algorithm.

3.1 Dense Tracklets

Our goal is to detect sudden movements in a video sequence. Therefore, we firstly have to register the motion of people both near and far away from a video camera. To do this, we use optical flow. As optical flow from two consecutive frames might contain data with noise, we apply short dense tracking based on optical flow to reduce the amount of noise. Firstly, for each video frame, we define a grid with a cell size of C pixels. Experiments show that a grid with a step size of $C = 5$ pixels gives good results and is a good balance between the amount of points and speed of an algorithm. We sample feature points on such defined grid and track each of these sampled points for L consecutive frames. The tracking is done using median filtering in a dense optical flow field extracted using Färneback's algorithm (Farnebäck, 2003). We limit the length of tracklets to L frames, allowing us to register short movements of objects, eliminate noise, and also avoid drifting problems. As we are also interested in the detection of short sudden movements, we select to track flow points for $L = 5$ consecutive frames. Finally, we remove tracklets that are static or contain sudden large displacements. The use of dense tracking is somehow similar to (Wang et al., 2011). However, it differs in computing the feature point set at every frame and computing much shorter tracklets for the purpose of noise elimination.

As a result of the above dense tracking algorithm, for each video sequence we obtain a set of tracklets $S = \{T_i\}$, where each tracklet T_i is represented as a list of points in the $2D$ space $T_i = ((x_1, y_1), (x_2, y_2), ..., (x_L, y_L))$. Finally, as the extracted tracklets are very short (several frames) we represent each tracklet T_i as 3 numbers (x_L, y_L, Δ), where (x_L, y_L) is the final position of the tracklet T_i and Δ is the total length of the tracklet's displacement vector, *i.e.*:

$$\Delta = \sum_{i=2}^{L} \sqrt{(x_i - x_{i-1})^2 + (y_i - y_{i-1})^2} \qquad (1)$$

3.2 Sudden Movement Discovery

To deal with problems related to camera view point (described above), we propose to divide images into small regions. In particular, similarly as for dense tracklets, we propose to compute a dense grid, that is however more sparse than the one used for tracking. We have empirically chosen a step size of $C' = 10$ pixels for this grid given the good results obtained and for it is a good balance between the amount of regions in the image and the speed of the algorithm. Then, the

process of sudden movements discovery is carried out in two phases.

First Phase

In the first phase of the algorithm, we apply dense tracking on a randomly chosen sample video (without any manual labelling). Then, we quantize all the extracted tracklets from all the videos to just defined cells of the grid. The quantization process is based on the final positions of the tracklets. As a result, for each cell of the grid we obtain a set of tracklets. Finally, we compute a simple statistic information for each i-th cell of the grid, *i.e.*mean μ_i and standard deviation σ_i. These statistics are calculated from tracklets' total length of the displacement vectors Δ.

Figures 2 and 3 present the obtained grid with means and standard deviations and show that they are in fact representative of the various motion patterns in the videos. In this example, the view is one of a platform with trains visible on the right side of the image. We can clearly see that in average, the motion is faster in the region of the train, slower on the left (where there is a wall) and top (above the height of a person) parts of the image. One can also notice that there is more motion closer to the camera than further from it. This is visible in the mean values (figure 2), but even more in the standard deviation values (figure 3). As stated above, this is due to the perspective of the view.

Second Phase

In the second phase of the algorithm, we use already extracted and quantized dense tracklets. Then, we calculate how far is each of these tracklets from the calculated means, in respect to the standard deviations. We assume that a cell of the grid contains a fast motion if at least one tracklet is farther from the mean μ_i by $\alpha \times \sigma_i$. We call such cell *activated*. The parameter α defines the sensitivity of the approach and is evaluated in Section 4.

3.3 Sudden Movement Localization and Noise Removal

We filter out detections due to noise. We consider 3 consecutive frames and merge all of their detections in one grid. We create a graph from these detections by creating a vertex for each activated cell in the grid and an edge between two activated vertices when the corresponding spatio-temporal cells are neighbors (*e.g.* if a is the blue activated cell in figure 4, we create the edge $a \longrightarrow b$ for every activated cell b in the orange region around a). Then we find all the connected components of this graph by looping through its vertices, starting a new depth first search whenever the loop reaches a vertex that has not already been included in a previously found connected component. Finally, we eliminate components with only 1 vertex because they are due to noise. The remaining components contain the localized sudden movements, along with information about the intensity of the movement encoded within the number of activated cells in each component. Then, this information can be used to rank cameras by the stream selection system.

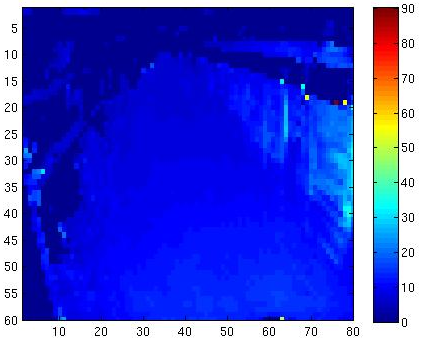

Figure 2: Illustration of computed mean value for each cell of the grid.

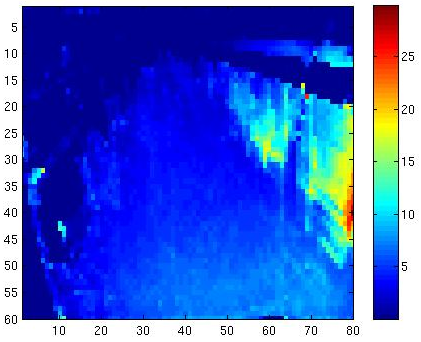

Figure 3: Illustration of computed standard deviation value for each cell of the grid.

4 Experiments

4.1 Dataset

The novelty of this topic and the absence of benchmarking datasets led us to use our own dataset recorded for a European project. This dataset contains data from several days recorded from two cameras in

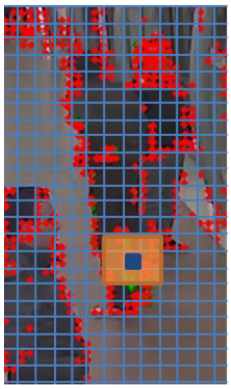

Figure 4: Filtering out noise.

the "Bibliothèque François Mitterrand" subway station in Paris. For this experiment we focused on the camera showing the platform (illustrated on the figure 5) because in this view people are the most stagnant, waiting for the train, and thus the more prone to perform sudden actions such as the ones we detect (see section 4.3). We also tested our approach the other camera, shown on the figure 6.

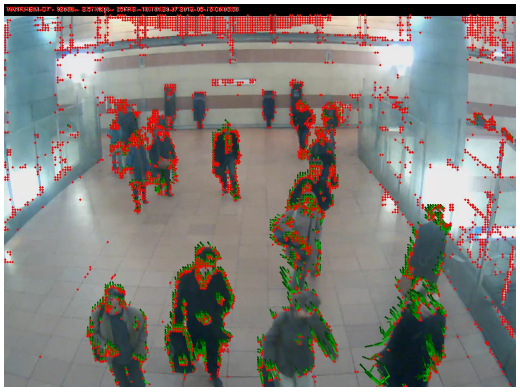

Figure 5: Left: Example of dense tracking.

We have used 8 hours of video recorded at 30 frames-per-second for the sudden movement discovery experiment. These videos contain various kinds of data, from low or no occupation of the platform to high density crowd, with slow or no motion to sudden movements such as the ones our system is built to detect.

4.2 Evaluation method

As we stated before, out method does not require any human annotations. However, for the purpose of evaluation, we provide them. We created Ground Truth data from selecting the most salient and relevant sud-

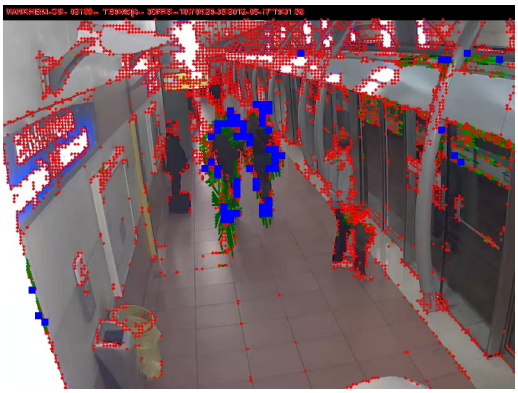

Figure 6: Example of a false positive which we do not considered as an error. The blue squares are cells in the grid where a sudden movement was detected.

den movements in our data. In fact, most of the time, people are quiet in the subway and no sudden motions are observed. The 50 annotated sudden movements correspond to 30 sequences, adding up to 1 hour of video (nearly 73000 frames). The annotation consists in defining a bounding box around the person performing the sudden movement for the duration of the movement. A such extract is called a Ground Truth *object* and our goal is to detect all objects. An object is detected if in a given minimum number (parameter δ) of its frames a sudden movement is detected. Indeed, the annotation is subjective and it is difficult for the annotator to decide when exactly a "sudden movement" begins and ends. An example of an annotated Ground Truth object is shown in figure 7, where the movement is annotated for 17 frames (which is less than 1 second). One can notice on figure 7 that at the beginning and the end of the action, the movement is actually slow, only the middle frames will be detected as containing a "sudden movement".

Figure 7: Example of Ground Truth of a sudden movement.

$\delta = 1$	GT	TP	FN	success rate	FP
$\alpha = 0.75$	50	50	0	100%	47.4%
$\alpha = 1.0$	50	50	0	100%	43%
$\alpha = 1.25$	50	50	0	100%	37.8%
$\alpha = 1.35$	50	48	2	96%	36.3%
$\alpha = 1.5$	50	48	2	96%	33.3%
$\alpha = 1.75$	50	47	3	94%	28.6%
$\alpha = 2.0$	50	45	5	90%	23.7%
$\alpha = 2.5$	50	44	6	88%	15.8%
$\alpha = 3.0$	50	44	6	88%	10.4%

Table 1: Detection of sudden movements in annotated video sequences with $\delta = 1$. GT: number of Ground Truth objects, TP: number of correctly detected Ground Truth objects, FN: number of missed Ground Truth objects, success rate: percentage of TP among GT, FP: percentage of frames where a sudden movement was detected but not annotated.

$\delta = 6$	GT	TP	FN	success rate	FP
$\alpha = 0.75$	50	50	0	100%	21.2%
$\alpha = 1.0$	50	49	1	98%	19.8%
$\alpha = 1.25$	50	47	3	94%	17.9%
$\alpha = 1.35$	50	47	3	94%	17.4%
$\alpha = 1.5$	50	47	3	94%	16.2%
$\alpha = 1.75$	50	44	6	88%	14.3%
$\alpha = 2.0$	50	44	6	88%	12.2%
$\alpha = 2.5$	50	43	7	86%	8.7%
$\alpha = 3.0$	50	39	11	78%	6%

Table 2: Detection of sudden movements in annotated video sequences with $\delta = 6$. Acronyms: see caption of table 1.

4.3 Results

Tables 1, 2 and 3 sum up the results obtained with the proposed algorithm with different values of parameters α and δ on the annotated sequences described above (section 4.2).

Tables 1, 2 and 3 (summarized in figure 8) show that the proposed algorithm successfully detects sudden movements (the success rates varies between 76% and 100% depending on the parameters). Varying the α parameter allows to adjust the sensitivity of the detection. With a low value of α (*e.g.* 0.75 or 1.0), we detect slightly slower sudden movements. An example of a movement that is detected with $\alpha = 1.0$ and not with $\alpha = 1.5$ is shown figure 9. In this example, a person was tying his laces and is now standing up. This action being a little unusual, it has drawn the attention of the annotator, but is not a very sudden movement. This proves that the very definition of what we want to detect is difficult to provide with precision, hence it is difficult to evaluate quantitatively this algorithm.

$\delta = 10$	GT	TP	FN	success rate	FP
$\alpha = 0.75$	50	48	2	96%	17.3%
$\alpha = 1.0$	50	47	3	94%	16.2%
$\alpha = 1.25$	50	46	4	92%	14.8%
$\alpha = 1.35$	50	45	5	90%	14.4%
$\alpha = 1.5$	50	45	5	90%	13.4%
$\alpha = 1.75$	50	43	7	86%	12%
$\alpha = 2.0$	50	43	7	86%	10.3%
$\alpha = 2.5$	50	41	9	82%	7.4%
$\alpha = 3.0$	50	38	12	76%	5.3%

Table 3: Detection of sudden movements in annotated video sequences with $\delta = 10$. Acronyms: see caption of table 1.

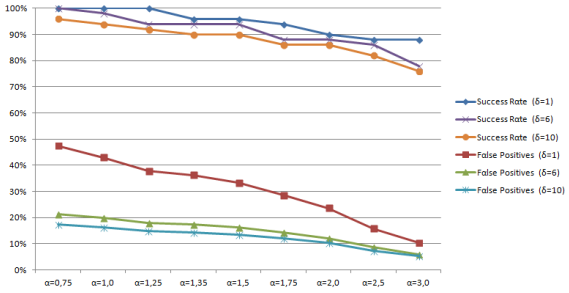

Figure 8: Graphical representation of results in tables 1, 2 and 3.

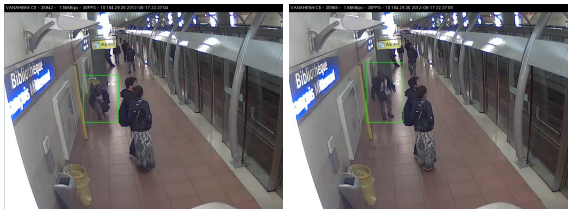

Figure 9: Example of an unusual movement annotated as sudden and detected with high sensitivity parameters.

Similarily, the given false positives rate (FP) in tables 1, 2 and 3 should be interpreted carefully. As mentioned above (section 4.2), the annotation process is subjective, it is impossible to establish an absolute Ground Truth. However, our goal is clearly defined as: detect in a live video stream rapid motion for enabling stream selection based on speed of motion. The 50 selected Ground Truth movements were significative, for the human annotator, of a known action, but the movements detected by the algorithm and counted as false positives were not necessarily errors. For example figure 6 shows a group walking fast or running and detected as a sudden movement. It was not annotated as such because most of annotated movements are movements of body parts and not movements of the whole body.

Tables 1, 2 and 3 allow to adjust the sensitivity of the algorithm, choosing between more or less alerts. With the highest tested sensitivity ($\alpha = 0.75$ and $\delta = 1$), more than half of the frames are filtered out. It is a mean of not showing to the operator a camera were people are very calm and move slowly. Setting the sensitivity to the lowest tested value ($\alpha = 3.0$ and $\delta = 10$), the algorithm detects 76% of what the annotator considered as sudden movements and has a false positive rate of only 5.3%.

Moreover, we can filter detections depending on their intensity. We can easily modify the algorithm to operate with several values of α at the same time, thus handling various intensity levels. The reactions of the system to movements detected at each level can be user defined. For instance, our algorithm can be integrated into an intelligent video surveillance system that uses stream selection based on various sources (*e.g.* algorithms for abnormality detection, crowd detection, abandoned luggage detection and so on) with the following rules:

Algorithm 1: Sudden movement events handling with 3 values of α.

input : 3 sets of detections for each value of α such as $\alpha_1 < \alpha_2 < \alpha_3$: $\{\mathcal{D}_{\alpha_1}\}$, $\{\mathcal{D}_{\alpha_2}\}$, $\{\mathcal{D}_{\alpha_3}\}$
output: Signals to the stream selection system
for $d \in \{\mathcal{D}_{\alpha_3}\}$ **do**
    RaiseWarningAlert($location(d)$);
    ShowWithPriorityLevelHigh($camera(d)$);
for $d \in \{\mathcal{D}_{\alpha_2}\}$ **do**
    ShowWithPriorityLevelNormal($camera(d)$);
for $d \in \{\mathcal{D}_{\alpha_1}\}$ **do**
    ShowWithPriorityLevelLow($camera(d)$);

In algorithm 1, the functions $location(Detection\,d)$ and $camera(Detection\,d)$ return respectively the bounding box in the image where the sudden movement was detected and the camera in which the detection happened. The functions RaiseWarningAlert, ShowWithPriorityLevelHigh, ShowWithPriorityLevelNormal and ShowWithPriorityLevelLow send signals to the stream selection system suggesting more or less strongly to show the given camera. It's up to the system to actually decide to select or not the given camera.

5 Conclusions

We have presented a fully automatic method to discover unusually rapid motion in video surveillance videos. Our method detects *sudden movements* in an unsupervised way, without requiring any training or camera calibration, which makes the approach suited for large video surveillance systems. The method does not put any constraint on the camera placement and is adapted to views with strong perspective and people visible in all parts of the image. We have extensively evaluated our approach on 8 hours of video containing 50 Ground Truth sudden movements, obtaining very good results (up to 100%). Moreover, we have tested our approach on further 4:10 hours of the platform view, and 2:40 hours of the footbridge view, confirming the satisfactory results given by the algorithm.

The sample video chosen for the first phase of the algorithm (section 3.2) is random during normal occupation hours of the scene, and does not require any manual labeling.

One possible application of the proposed software is to be integrated in a stream selection system. The main goal is to avoid missing a dangerous situation because the operator was not observing the needed camera at the time. Our system, if parametrized to a high sensitivity, can detect 100% of the sudden events annotated in the Ground Truth and thus potentially dangerous. It can raise an alert to the global system if the intensity of the detected motion is the highest. In case of lower intensity, our algorithm can send a signal to the stream selection system, which will take the decision to show the stream or not. It avoids displaying to the security operator a camera with slow to normal motion, when a camera with rapid motion may be more relevant to show.

Finally, when integrated in an event detection framework such as the one described by (Zaidenberg et al., 2012), our sudden movement detector can be used to recognize behavior, and in particular group behavior. The *ScReK* system of (Zaidenberg et al., 2012) enables the definition of various scenarios of interest corresponding to group activities. The sudden movement detector triggers the recognition of one of these scenarios (called a primitive event) and can be part of a more complex scenario. Thus, the semantic understanding of the scene is enhanced.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Communitys Seventh Framework Programme FP7/2007-2013 - Challenge 2- Cognitive Systems, Interaction, Robotics - under grant agreement n 248907-VANAHEIM.

REFERENCES

Belshaw, M., Taati, B., Snoek, J., and Mihailidis, A. (2011). Towards a single sensor passive solution for automated fall detection. In *IEEE Engineering in Medicine and Biology Society*, pages 1773–1776.

Blank, M., Gorelick, L., Shechtman, E., Irani, M., and Basri, R. (2005). Actions as space-time shapes. In *ICCV*, volume 2, pages 1395–1402.

Brox, T. and Malik, J. (2011). Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3):500–513.

Castrodad, A. and Sapiro, G. (2012). Sparse modeling of human actions from motion imagery. *International Journal of Computer Vision*, 100(1):1–15.

Daniyal, F. and Cavallaro, A. (2011). Abnormal motion detection in crowded scenes using local spatio-temporal analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, pages 1944–1947.

Efros, A., Berg, A., Mori, G., and Malik, J. (2003). Recognizing action at a distance. In *ICCV*, pages 726–733 vol.2.

Emonet, R., Varadarajan, J., and Odobez, J.-M. (2011). Multi-camera open space human activity discovery for anomaly detection. In *AVSS*.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In *Scandinavian Conference on Image Analysis*, LNCS 2749, pages 363–370.

Gaidon, A., Harchaoui, Z., and Schmid, C. (2011). A time series kernel for action recognition. In *BMVC*.

Jouneau, E. and Carincotte, C. (2011). Particle-based tracking model for automatic anomaly detection. In *ICIP*, pages 513–516.

Kellokumpu, V., Zhao, G., and Pietikinen, M. (2008). Human activity recognition using a dynamic texture based method. In *BMVC*.

Mahadevan, V., Li, W., Bhalodia, V., and Vasconcelos, N. (2010). Anomaly detection in crowded scenes. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 1975–1981.

Raptis, M. and Soatto, S. (2010). Tracklet descriptors for action modeling and video analysis. In *ECCV*.

Wang, H., Kläser, A., Schmid, C., and Cheng-Lin, L. (2011). Action Recognition by Dense Trajectories. In *CVPR*, pages 3169–3176, Colorado Springs, United States.

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., and Bischof, H. (2009). Anisotropic Huber-L1 Optical Flow. *BMVC*, pages 108.1–108.11.

Wu, S., Oreifej, O., and Shah, M. (2011). Action recognition in videos acquired by a moving camera using motion decomposition of lagrangian particle trajectories. In *ICCV*.

Xiang, T. and Gong, S. (2005). Video behaviour profiling and abnormality detection without manual labelling. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1238–1245 Vol. 2.

Xiang, T. and Gong, S. (2008). Activity based surveillance video content modelling. *Pattern Recognition*, 41(7):2309 – 2326.

Zaidenberg, S., Boulay, B., and Bremond, F. (2012). A generic framework for video understanding applied to group behavior recognition. In *9th IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS 2012)*, Advanced Video and Signal Based Surveillance, IEEE Conference on, pages 136 –142, Beijing, Chine. IEEE Computer Society, IEEE Computer Society.