Loveparade 2010: Automatic Video Analysis of a Crowd Disaster

Barbara Krausz, Christian Bauckhage

Fraunhofer IAIS, 53754 Sankt Augustin, Germany

Abstract

On July 24, 2010, 21 people died and more than 500 were injured in a stampede at the Loveparade, a music festival, in Duisburg, Germany. Although this tragic incident is but one among many terrible crowd disasters that occur during pilgrimage, sports events, or other mass gatherings, it stands out for it has been well documented: there were a total of 7 security cameras monitoring the Loveparade and the chain of events that led to disaster was meticulously reconstructed.

In this paper, we present an automatic, video-based analysis of the events in Duisburg. While physical models and simulations of human crowd behavior have been reported before, to the best of our knowledge, automatic vision systems that detect congestions and dangerous crowd turbulences in real world settings were not reported yet. Derived from lessons learned from the video footage of the Loveparade, our system is able to detect motion patterns that characterize crowd behavior in stampedes. Based on our analysis, we propose methods for the detection and early warning of dangerous situations during mass events. Since our approach mainly relies on optical flow computations, it runs in real-time and preserves privacy of the people being monitored.

Keywords: Crowd Behavior, Crowd Dynamics, Crowd Turbulence, Congestion, Video Analysis, Computer Vision, Optical Flow

Email address: {barbara.krausz, christian.bauckhage}@iais.fraunhofer.de (Barbara Krausz, Christian Bauckhage)

Year	Place	Deaths
2010	Loveparade, Duisburg	21
2010	Water Festival, Phnom Phen	>380
2006	Stadium, Yemen	51
2006	Pilgrimage, Mena	363
2005	Religious Procession, Bagdad	>640
1999	Subway Station, Minsk	53
1990	Pilgrimage, Mena	1426
1989	Stadium, Sheffield	96
1982	Stadium, Moscow	340

Table 1: Examples of recent deadly stampedes and crowd disasters (see [1]).

1. Introduction

Mass events are (and always have been) popular in human societies all over the world. Nowadays, typical examples include sports events, festivals, or concerts. Due to increasing populations and higher mobility, mass events attract ever growing numbers of visitors and adequate safety measures are becoming more and more important. Nevertheless, despite all precautions and the use of technology such as video surveillance, deadly stampedes and crowd disasters still occur rather frequently (see Table 1).

Work towards the prevention of crowd disasters is often based on experiments and the analysis of video footage from crowd catastrophes. Researchers try to understand the mindset of panicked people and develop physical models in order to simulate human behavior in crowded environments, in particular to devise evacuation strategies. Using such simulations, one can identify and defuse places in an environment that are potentially dangerous. For instance, a study of video footage from a crowd disaster in Mina during the annual muslim pilgrimage by Helbing et al. [2] led to structural alteration of the site and helped to re-organize the pilgrimage. Indeed, since then, there has not been another deadly stampede during the Hajj.

However, human reactions and mass behavior are often unpredictable, in particular, whenever alcohol consumption is an integral part of a mass events. Moreover, gathering areas are often large and become unmanageable if there are many thousands of visitors. In cases like these, automatic video surveillance systems may help to estimate the visitor density and to detect indicators of critical situations. This can buy crucial time in which security

personnel can be dispatched and streams of pedestrians can be redirected.

In this paper, we analyze video footage of the crowd disaster at the Loveparade 2010 in Duisburg, Germany. During that event, 21 people died and more than 500 people were injured in a stampede in a passage way that was too narrow to accommodate thousands of visitors. In our experiments, we observe critical motion patterns of the crowd that manifest shortly before the deadly congestion occurred. These motion patterns are in fact very characteristic for human behavior in congestion situations and can inform the automatic detection of over-critical densities of pedestrians.

We present a system for motion behavior analysis of masses that computes dense optical flow fields which can be calculated in real-time. By using optical flow, our system avoids the need for detection and tracking of individual pedestrians which is often impossible due to inappropriate camera viewpoints or occlusions due to the large number of people. Based on the results obtained from this approach, we propose methods to automatically detect congestions and shock waves. To the best of our knowledge, this is the first vision-based system that robustly detects dangerous situations like overcrowding and crowd turbulences in real-time.

1.1. Related Work

Pedestrian dynamics have been studied intensively for more than 40 years. In the past, empirical studies were mainly conducted in order to support planning of urban infrastructure, for instance, in the design of pedestrian facilities. More recently, knowledge about crowd dynamics has been used to improve evacuation strategies in emergency situations and to prevent congestions and overcrowding (see [3] for an overview). Simulations are a standard tool in the study of self-organizing effects of large groups of pedestrians. Physical models modeling pedestrians based on the analogy to gases, fluids or granulates have been developed in order to account for individual behavior. The *social force model* [4] as well as *cellular automata* [5] which both model pedestrian dynamics on a microscopic level are among the more widely used approaches.

The social force model assumes socio-psychological and physical forces that influence human behavior. It captures interactions between nearby pedestrians as well as effects caused by the environment by integrating repulsive and attractive effects into a force model that represents crowd behavior. In [6], Helbing et al. extend the basic model towards effects of escape panic by adding further random forces. Models based on cellular automata [5] represent the environment as a grid and superimpose static and dynamic force

fields that alter transition probabilities of the pedestrians moving on the grid. These fields model obstacles and regions being more attractive to pedestrians as well as influences of other pedestrians. This modeling approach has also been applied to evacuation situations [7]. Both models are frequently used in simulations and, indeed, they were shown to reproduce several phenomena that can be observed in crowds (see section 1.2).

In addition to simulations, experimental studies are conducted in order to understand human behavior and improve existing physical models. Parameters such as crowd density, speed, flow, and crowd pressure (see [2, 8] for definitions) are determined either manually [9] or by means of digital image processing [10, 11]. Usually the resulting representations are based on experimental data and do not consider real data. Moreover, video-based experiments are typically carried out using top-view cameras in order to avoid occlusions and to facilitate automatic video analysis. Techniques that are applied in this context usually detect and track individuals but there also exist holistic approaches that make use of optical flow features.

Over the years, various visual tracking approaches have been reported that were specifically developed for tracking pedestrians in crowded scenes [12, 13, 14]. More recently, ideas adopted from simulations of pedestrian dynamics were incorporated into the design of visual tracking systems. Ali and Shah [15] present a tracking framework inspired by the cellular automaton model [5]. They automatically calculate force fields that integrate information on human behavior as well as the locations of obstacles and important regions such as exit doors. In their previous work [16], Ali and Shah propose a flow segmentation framework which enables them to detect changes in flow patterns. Mehran et al. [17] adopt ideas from the social force model and estimate interaction forces in order to detect abnormal events. Both these works do not detect and track individuals. Instead, they apply the technique of *particle advection* that places particles onto a grid and moves them according to the underlying optical flow field. However, in our case, particle advection is not applicable due to the the camera view point and the resulting occlusions in situations of high pedestrian density.

1.2. Phenomena observed in Dense Crowds

At mass gatherings such as concerts, sports events or religious processions the density of the crowd easily becomes extremely high; studies report densities up to 11 people per square meter [18]. High pedestrian densities

usually come along with typical patterns of mass behavior such as *stop-and-go waves* or *crowd turbulences* [1]. These motions patterns in a crowd are first indicators of dangerous overcrowding that may get out of control and entail disaster (see Table 1).

In situations of high pedestrian density, jams and bottlenecks are building up rapidly and moving and passing becomes uncoordinated. In studies of pilgrim flows in Makkah [1], *stop-and-go waves* have been observed even in areas without any obvious bottlenecks. These waves are characterized by intermittent flow of pedestrians and can last over a long period of time (over 20 minutes). Essentially, stop-and-go waves show alternating forward pedestrian motion and backward gap propagation. They occur when the pedestrian density is critically high and unobstructed pedestrian flow becomes impossible [2]. Stop-and-go waves are first signs of critical and dangerous pedestrian densities. Researchers observed transitions from stop-and-go waves to even more critical motion patterns called *crowd turbulence*. This effect occurs in situations of extremely high densities and is characterized by movements into all possible directions. It is caused by people moving involuntarily inducing sudden movements of other people nearby. People are pushed around and fall down. They are trampled down and, moreover, they turn into obstacles for others leading to more stumbling people. As a result, most people die by suffocating due to dangerous pressure of up to 4500 N/m on their chests [1].

1.3. Outline

In the following section, we present background information on the Loveparade stampede. We describe the festival area and the locations of the surveillance cameras and give a detailed account of the incident. In section 3, we describe video analysis experiments carried out using the security camera footage recorded during the Loveparade. We use our findings in section 4 to develop a method for the automatic detection of situations of dangerous congestion. In section 5, we present an approach to detecting and tracking of shock waves that propagate through the crowd. A conclusion will close this contribution.

2. Loveparade Stampede

On July 24, 2010, the Loveparade in Duisburg ended in a terrible stampede where 21 visitors died and more than 500 were injured. As of this

writing, many questions as to the reasons of this catastrophe remain unanswered. However, it is unquestionable that at some locations of the festival area there were too many people present to guarantee their safety. At the main entry ramp to the festival area, the density of visitors was temporarily so extremely high that people who fell were trampled down by other visitors. All victims were crushed to death or suffocated due to enormous pressure on their chests.

2.1. Festival Area

The Loveparade took place at the compound of the former freight station of Duisburg which is located near the city center. It is situated about 2 km from the main station and lies between railway tracks and a highway. The former freight station has an area of about 230 000 square meters and about half of the total area was accessible for the visitors [19]. Figure 1 shows an aerial view of the festival area during the event and Figure 2 shows a schematic view of the compound. During the event, there were two entries for the visitors: The eastern tunnel which is about 250m long and the western bridge area. Both meet at the main ramp which simultaneously served as the main entry to- and the exit from the festival area. Another smaller exit ramp did exist but had been closed by the police during the event.

Figure 1: Aerial view of the festival area [20].

Figure 2: Schematic view of the festival area and locations of surveillance cameras [21].

2.2. Surveillance Cameras

The festival area was continuously monitored by seven cameras. Figure 2 shows the locations of each camera. Exemplary screen-shots of all cameras can be seen in Figure 3. The entrance area in the eastern tunnel and the western bridge area were monitored by 4 cameras (referred to as camera 13 – 16) where three of them were static cameras (camera 14–16). Cameras 4, 5 and 12 were non-static cameras and did monitor the main festival area in the upper part. Video footage recorded by all cameras can be downloaded from [21]. In order to protect the privacy of the victims, video recordings are only available for the time between 13:30h and 16:40h.

2.3. Chronology

In the following, we summarize the chronology of the Loveparade. For a full account see [21].

The festival area is opened for visitors at 12:00h and the festival begins at 14:00h. 30 minutes later the number of visitors starts to increase at the main entry/exit ramp. At 15:00h, the security guards at the tunnel entrances close some access control points thus reducing the number of visitors in the western bridge area. At 15:50h, a first police cordon is formed in the western bridge area. At the same time, the small exit ramp is closed. A crowd in front of the small exit ramp results from visitors trying to leave the area. Between 15:50h and 15:57h a second police cordon is formed at the entrance of the eastern tunnel. At 16:00h, the western access control points are closed, so that only a few visitors are visible in the western bridge area. Two minutes later, the access control point is re-opened resulting in a sudden increase of the number of visitors in the western bridge area. A crowd starts to form behind the police cordon which has been established near the small exit ramp in the western bridge area. At the same time, a third police cordon is built on the main entry/exit ramp stopping those visitors that want to leave the festival area as well as arriving visitors that want to enter the main festival area. Only a few visitors are visible in the lower part of the main ramp, whereas in the western bridge area the crowd increases. Starting around 16:09h, visitors being stopped by the third police cordon create a bottleneck on the main ramp. At 16:13h, the small exit ramp is opened as an entry ramp. Due to the police cordon in the western bridge area the large amount of visitors can hardly move. Some of the visitors begin to climb a fence. At 16:15h the second police cordon in the eastern tunnel dissolves and the number of visitors in the lower part of the ramp increases. Five Minutes later, the first police cordon in the western bridge area dissolves as well. The huge crowd that was blocked by this cordon, suddenly moves to the lower part of the ramp where they meet the visitors from the eastern tunnel. The visitor flows from both entries increase, but due to the third police cordon, they cannot enter the festival area. Thus, the pressure on the third police cordon increases and it is dissolved at 16:24h. Some visitors begin to escape by climbing a small flight of stairs at the bottom of the ramp as well as light poles. A fourth police cordon is formed at the upper part of the main ramp increasing the pressure in the lower part. Starting from 16:40h, the situation gets out of control. More and more people try to escape by climbing the stairs and light poles. In the following, 21 visitors die, more than 500 visitors

are injured.

3. Automatic Video Analysis

We conducted a series of video analysis experiments in order to determine to which extent typical motion patterns that characterize different phases of critical crowd behavior can be extracted automatically. For most of the results reported here, we concentrate on camera 15 (see Figure 3) which is a static camera that was located in the western bridge area. Next, we describe our feasibility experiments. Then, based on our findings, we develop methods for the detection of potentially dangerous situations in crowds. Since real-time capability is pivotal in our application domain, we focus on the automatic analysis of motion vectors. In contrast to traditional approaches to real-time surveillance, we did not attempt to detect and track individual people because the high densities of pedestrians and the challenging camera viewpoints that characterize our setting would necessitate considerable computational efforts. Instead, we consider dense optical flow fields to determine major motion patterns and motion directions in the crowd. In addition to computational efficiency, this also guarantees the privacy of the people being monitored.

3.1. Histograms of Dense Optical Flow Fields

In our first experiment, we investigate the utility of dense optical flow fields as a feature to represent global crowd movements. For optical flow computation, we apply the method proposed by Farnebäck [22] who uses quadratic polynomials to estimate translations of a local neighborhood and determines motion vectors from polynomial expansion coefficients. Figure 4 shows an example of the resulting optical flow fields for camera 15.

Given the dense optical flow field for a frame of a video sequence, we compute two-dimensional histograms ($36 \cdot 100$ bins) of motion magnitude and motion direction of the flow vectors of the entire frame. Next, we cluster all of the resulting two-dimensional histograms using the k-means algorithm. Figure 5 shows results for $k = 5$.

Figure 5(a) depicts how the individual frames of the video footage of camera 15 are assigned to different clusters. Up until about 15:50h, most frames (colored in blue) correspond to the cluster whose centroid is shown in Figure 5(b). This cluster of flow vector histograms represents a general motion to the right (250°) of a rather high magnitude: large numbers of people are

marching towards the festival area. Shortly after 16:20h, however, all histograms are assigned to the cluster (colored in green) whose centroid is shown in Figure 5(c). This cluster of motion vectors histograms is characterized by weak motion oscillating between right (270°) and left (90°). This pattern is indicative of congestion where there is no globally dominant motion direction anymore; instead of moving forward the crowd has come to a halt and people are stepping from one foot to the other in order to keep their balance.

From the results of these initial feasibility experiments, we conclude the 2D histograms that summarize magnitude and direction of optical flow vectors provide a suitable means to characterize basic motion patterns and phase transitions between general crowd behaviors.

3.2. Extracting Basis Histograms

In our second experiment, we consider complex motion that can be observed at a mass event as the superposition of various motion behavior performed by many different people. Obviously, it is very difficult or even impossible to extract motion behavior of every single human visible in the scene. Thus, we aim at extracting motion behavior of groups of people that perform similar actions. Thereto, we examine complex motion patterns and decompose them into basic motion patterns. Since motion varies over time and space, we divide the video sequence both spatially and temporally. We superimpose a grid of cells over the video frames where grid cells towards the back of the scene are smaller in order to account for perspective distortions. For each cell of the grid, we again compute two-dimensional histograms ($36 \cdot 100$ bins) of motion direction and magnitude for the whole video sequence. Then, we apply Non-Negative Matrix Factorization (NMF) [23] to sets of histograms obtained during time intervals of 30 seconds. Non-Negative Matrix Factorization and other matrix decomposition methods have successfully been applied to action recognition [24, 25], for example. By this means, complex motion patterns can be decomposed into a set of simpler motion patterns. The set of m optical flow histograms is regarded as an $n \times m$ matrix V in which all entries are non-negative and where each column vector corresponds to a flattened histogram of size n. NMF finds an approximate factorization of V as $V \approx WH$ with multiplicative update rules in an iterative procedure [23]. In our case, each column of W represents a basis histogram of motion magnitude and motion direction. In particular, we decompose complex histograms of magnitude and motion direction into a set of five basis histograms. Each histogram describes motion patterns that

are characteristic for certain parts of the scene as well as for particular time intervals.

Figures 6 to 8 show basis histograms for three different time intervals: The basis histograms shown in Figure 6 represent typical motion patterns at 14:40h. At that time, the flow of pedestrian flow is normal, visitors proceed to the main entry without encountering major obstacles. In the histograms this is reflected by a sharp peak corresponding to a motion direction of about 250°and a magnitude of 3 to 4 pixels.

Figure 7 shows basic motion patterns at 15:53h shortly after the police cordon had been established. Now, most visitors are blocked and just a few continue moving towards the festival area. In the extracted basis histograms, a distinct dominant motion is not visible anymore.

On the other hand, Figure 8 shows two motion patterns that are characteristic for the general crowd motion at about 16:28h when the visitor density is very high and causes congestion. The basis histogram in Figure 8(a) corresponds to rightward motion (270°); the histogram in Figure 8(b) reflects leftward motion (90°). Indeed, similar motion patterns showing both motion to the right and to the left can be observed for the whole period between 16:22h and 16:40h when then visitor density is over-critically high.

Given all the resulting basis histograms, we cluster them using the k-means algorithm ($k = 5$) and, for each time interval of 30 seconds, we determine the nearest cluster center. Figure 9(a) shows the cluster assignment for the entire video sequence. In time intervals of normal pedestrian flow, we observe motion patterns with a distinct peak at approximately 250°and a magnitude of about 3 to 4 pixels (see Figures 9(b) and 9(f)). At 15:50h when the first police cordon is formed, no distinct motion is visible in the corresponding histograms in Figures 9(c) and 9(e). Contrarily, the motion patterns after 16:22 are composed of rightward and leftward motions as depicted in Figure 9(d).

4. Detection of Congestions

The experiments described in section 3 reveal several phases of characteristic motion patterns: In the first phase (between 13:40h and 15:50h), we mainly observe motion directions of approx. 250°and a magnitude of 3-4 pixels. This pattern corresponds to movement of visitors into the direction of the main entry. The second phase begins at 15:50h when the first police cordon is formed in the western bridge area. At that time, we cannot observe

any motion into a distinct direction anymore. The last phase begins at about 16:22h shortly after the police cordon has been dissolved. The large crowd that was previously blocked from entering the festival area suddenly moves forward and quickly jams. In this congestion situation, the pedestrian density is very high and the only type of motion visible is to the right (270°) and to the left (90°) corresponding to a situation where people are stepping from one foot to the other in order to keep their balance resulting in oscillating motions.

In fact, such oscillating motions of a crowd are frequently observed in congestion situations and they result from the upper bodies of the people in the crowd swinging rightwards and leftwards. Liu et al. [10] reported experiments with several groups of pedestrians moving with different speeds from 0.26 m/s to 1.72 m/s. Their movements were filmed from above and the authors generally observed lateral oscillation in the trajectories. This is due to the fact that people do not move along a straight line, instead, it is a characteristic of human gait, that they tend to swing laterally. Liu et al. [10] also observed that while the amplitude of the lateral oscillation is higher for lower speeds, the frequency increases for higher speeds. The authors found linear relationships between the velocity and the amplitude as well as between the velocity and the frequency.

4.1. Feature for Congestion Detection

Given these observations and the findings of our feasibility experiments, we propose a method for the automatic detection of dangerous congestion situations. First, we compute dense optical flow and corresponding two-dimensional histograms of motion direction and magnitude. Then, we average the histogram over short time intervals. As described above, histograms that are indicative of congestion situations show small motion along two major directions (rightwards and leftwards in our case) which reflect lateral oscillation of the people's upper bodies. Such histograms show a high degree of symmetry (see Figure 9(d)) so that we measure the mirror symmetry of an optical flow histogram and consider the resulting value a feature for congestion detection.

We compute the symmetry measure by summing up the the absolute differences between the histogram and a flipped version of itself. As described above, we subdivide each frame into a set of cells with cells in the background of the scene being smaller to account for effects of viewing perspective. Let $H_{t,c}(dir, magn)$ be the two-dimensional histogram of direction and magnitude

of cell c at time t. Then, denoting by $\hat{H}_{t,c}(dir, magn)$ the flipped version of $H_{t,c}(dir, magn)$, we compute

$$sym_{t,c} = \sum_{dir,mag} \left| \hat{H}_{t,c}(dir, magn) - H_{t,c}(dir, magn) \right|. \quad (1)$$

Accordingly, low values of $sym_{t,c}$ indicate that $H_{t,c}(dir, magn)$ is highly mirror-symmetric and is indicative for a congested area. Figure 14 shows the result of a corresponding analysis for camera 15. To create this plot, we summed the values of $sym_{t,c}$ for the cells in the scene foreground.

4.2. Sequential Change-Point Detection

From the graph in Figure 14, one can clearly distinguish different phases of crowd motion. However, the scale of the values of $sym_{t,c}$ is not clear: It was stated above, that low values indicate that the optical flow histograms are highly mirror-symmetric which is indicative for congestions. But the scale depends on different factors, such as the camera view point, the number of people visible in the scene as well as the environment.

In order to overcome this problem, we learn typical values of $sym_{t,c}$ and detect deviations for identifying anomalies. Thereto, we apply a sequential change-point detection algorithm enabling us to detect unusual events as well as congestions in particular. The method proposed by De Oca et al. [26] extends the conventional cusum algorithm [27]. It is a non-parametric cusum algorithm that allows for distributions varying in time and uses historical data for obtaining suitable thresholds above which an alarm is raised. We extend this algorithm to compute an additional measure that characterizes the severity of an alarm.

Let Y denote the value of $sym_{t,c}$. Now, consider a sequence of observations $\{Y_i\}_{i=1}^{N}$. We use previous observations $\{Y_j\}_{j=i-k-l}^{i-k}$ to estimate a reference distribution where k is a fixed time interval and l is a fixed number of historical observations that are used for estimating the reference distribution. Next, we denote the upper and lower α-percentiles of the reference distribution as $Q(\alpha)$ and $Q(1-\alpha)$, respectively, where α is specified by the user and controls the degree a deviation from the reference distribution is considered as critical. See Figure 10 for an illustration.

The cusum algorithm continuously accumulates deviations of incoming observations from the reference distribution:

$$
\begin{aligned}
S_i^+ &= \max\{0, S_{i-1}^+ + Y_i - Q(\alpha)\}, S_0^+ = 0 \\
S_i^- &= \max\{0, S_{i-1}^- + Q(1-\alpha) - Y_i\}, S_0^- = 0
\end{aligned} \tag{2}
$$

It raises an alarm if either $S_i^+ > \Theta$ or $S_i^- > \Theta$, where in the first case, we detect an upward shift of the signal and a downward shift in the latter case. The threshold Θ is calculated from the reference distribution as follows: Suppose that the sequence of observations is drawn from the reference distribution, that is, no anomaly occurs. Using a bootstrap resampling method, Θ is selected so that the probability of a false alarm is equal to γ, a parameter specified by the user. For that purpose, M sequences are sampled from the reference distribution. For each sampled sequence m, cusum statistics according to equation 2 are computed and $\max\{S_{sampled,m}^+, S_{sampled,m}^-\}$ is determined. Next, for each sampled sequence m, we select the maximum value of $\max\{S_{sampled,m}^+, S_{sampled,m}^-\}$ and compute the threshold Θ as the $(1-\gamma)$-percentile from these maximum values.

Whenever either S_i^+ or S_i^- exceeds the computed threshold Θ, we raise an alarm (see Figure 11). De Oca et al. [26] also propose a method for detecting the end time of an alarm. They apply a slope testing technique for detecting a downward trend in the cusum statistics which indicates that the deviations from the reference distribution become smaller: Without loss of generality, we assume that S_i^+ exceeds Θ at time a (The same rationale holds for S_i^-.). Then, a linear regression model is continuously fitted to a sliding window of cusum values $\{S_i^+\}_{i=n-\nu+1}^n$ for $n = a, a+1, \ldots$ and ν being a fixed size of the sliding window, see Figure 12. The end time of an alarm is detected, when the slope of the linear regression model is less than or equal to zero. Then, cusum statistics S_i^+ or S_i^-, respectively, are set to zero. Additionally, we propose to measure the severity of the raised alarm as a value $L \in [0 \ldots 1]$ by computing the angle of the regression line in degrees and dividing it by 90°. This is motivated by idea that the slope of the linear regression model of the cusum statistics S_i^+ (or S_i^-, respectively) depends on the deviation of the current observation Y_i to the reference distribution: The higher the deviation is compared to the reference distribution, the larger the slope of the regression line is. If L is near to one, the slope is large and the situation is considered to be very critical.

In particular, congestions are characterized by low values of $sym_{t,c}$ as described above. Thus, an alarm raised by the system is very severe and

may indicate a congestion, if L is near to one and S_i^- exceeds the threshold indicating that $sym_{t,c}$ decreases due to optical flow histograms becoming more and more mirror-symmetric. Section 6 gives results obtained from video sequences showing congested areas.

5. Detection of Crowd Turbulences

In areas of extremely high pedestrian density, the movement of a person affects other nearby people. Shock waves might occur and propagate through the crowd. Situation like these are extremely dangerous since people cannot control their motion anymore but are moved by the crowd; people who loose their balance and fall down in a shock wave typically get crushed and suffocate.

The video footage from the Loveparade shows short sequences of shock waves in the vicinity of the main entry which was monitored by camera 13 (a non-static camera). Shock waves as they can be seen in the video footage are characterized by a sudden increase of the magnitude of the optical flow motion vectors as well as a high standard deviation of the magnitude of flow vectors as some people in a local region are already pushed by the wave whereas other nearby people do not move. Moreover, since several people in a local neighborhood move into the same direction, the standard deviation of local motion directions σ_{dir} is small. Therefore, in order to accomplish the automatic detection of shock waves, we divide the frame into C cells. For each cell c and each time t, we compute the average magnitude of optical flow motion vectors $\mu_{mag,t,c}$, the standard deviation of the magnitude of optical flow motion vectors $\sigma_{mag,t,c}$ and the standard deviaton of motion directions $\sigma_{dir,t,c}$. As described before, shock waves are characterized by a high magnitude and a high standard deviation of motion magnitudes as well as a small standard deviation of motion directions. Hence, we also compute a value $p_{t,c} = \dfrac{\mu_{mag,t,c} \cdot \sigma_{mag,t,c}}{\sigma_{dir,t,c}}$ for each cell c and each timepoint t. In those cells of a frame where a shock wave is observed to propagate, $p_{t,c}$ will be high. In order to normalize the feature $p_{t,c}$, we compute $\mu_{row,t}$, the average value of all $p_{t,c}$ values in a row as well as the standard deviation $\sigma_{row,t}$ of all $p_{t,c}$ values in a row. Then, we normalize each $p_{t,c}$ by subtracting the mean value of the corresponding row and dividing by the standard deviation all $p_{t,c}$ values in the row:

$$p_{t,c}^{norm} = \frac{p_{t,c} - \mu_{row,t}}{\sigma_{row,t}} \tag{3}$$

Now, we apply the sequential change-point detection algorithm described in section 4.2 to learn typical values of $p_{t,c}^{norm}$ for each cell c and detect deviations from that value in order to detect shock waves. Figure 13 shows an example of automatically detected shock wave regions colored in red. Here, in two cells near the left border of the frame, a shock wave can be observed.

6. Results

6.1. Congestion Detection in the Loveparade Video Footage

We tested the approach to detect unusual events and congestions described in section 4 on video footage from camera 15. Figure 14 shows the development of $sym_{t,c}$ measuring the mirror symmetry of the optical flow histograms. To create this plot, we averaged histograms of optical flow over a time period of 10 seconds and summed the values of $sym_{t,c}$ for the cells in the scene foreground. We automatically detect change-points in an online manner using the cusum algorithm presented in section 4.2. Here, we use $l = 90$ observations of historical data to estimate a reference distribution which corresponds to a time interval of 15 minutes whereas k is set to 30 observations (= 5 minutes). We set α, the parameter to control the degree a deviation from the reference distribution is considered as critical, to 0.95 and γ which controls the probability of false alarms to 0.1. Next, the parameter ν used in the detection of the end time of the alarm is set to 8 and M specifying the number of sampled sequences for computing a suitable threshold is set to 100. Alarms that have been raised by our system are colored in red, if the signal is at a low level, whereas a jump to a high value is marked in green. In the lower part, the alarm level L is depicted which measures the severity of the alarm.

Comparing the automatically detected alarms with the video footage reveals that seven out of ten alarms correspond to anomalies in the video, e.g. ambulances or police cars crossing the scene. Figure 14 gives interpretations for these alarms. In particular, at about 16:27 the system raises a severe alarm ($L > 0.7$) and reports low values of $sym_{t,c}$ which is indicative for a congestion. In fact, at this time point the crowd is densely packed and has come to a halt. In this situation, our system would have detected a very critical situation and alarmed the security personnel to take necessary actions in order to prevent a deadly stampede.

Only two false alarms at the beginning of the video recordings are reported which can be explained by the fact that our system has not yet integrated

enough observations of normal crowd behavior. A third false positive alarm is raised at 16:17, but has a low severity ($L = 0.1$) and lasts for just a few seconds.

6.2. Congestion Detection in the Hermes Dataset

We also tested our approach on a dataset recorded by the Hermes project [28] under laboratory conditions. Here, pedestrians walk through a corridor with a bottleneck at the end of the corridor. The scene is recorded by two overhead cameras with a small overlapping field of view, see Figure 15. The camera viewpoint differs significantly from the camera viewpoint of the Loveparade videos.

The feature for congestion detection described in section 4.1 has been designed to capture lateral body oscillation of people that are going towards the camera by measuring the mirror symmetry of optical flow histograms. However, due to a different viewpoint in the Hermes dataset, lateral body oscillation of pedestrians do not result in leftwards and rightwards motion. As a consequence, we do not make use of the original feature for congestion detection (section 4.1). However, from an overhead viewpoint, we can directly make use of the magnitude of optical flow motion vectors to estimate the velocity of the pedestrians. Thus, we adapt the feature for congestion detection by computing the center of mass of the two-dimensional optical flow histogram where the histogram is obtained by averaging optical flow histograms over a time interval of one second. Let $m_{dir,mag,t}$ denote the histogram value (or the mass) of the histogram bin with direction dir and magnitude mag at time t. The center of mass $(c_{dir,t}, c_{mag,t})$ of the histogram is computed as the average of the histogram bins weighted by their masses:

$$c_{dir,t} = \frac{\sum_{dir,mag} m_{dir,mag,t} \cdot dir}{\sum_{dir,mag} m_{dir,mag,t}} \tag{4}$$

$$c_{mag,t} = \frac{\sum_{dir,mag} m_{dir,mag,t} \cdot mag}{\sum_{dir,mag} m_{dir,mag,t}} \tag{5}$$

which can be simplified to

$$c_{dir,t} = \sum_{dir,mag} m_{dir,mag,t} \cdot dir \tag{6}$$

$$c_{mag,t} = \sum_{dir,mag} m_{dir,mag,t} \cdot mag \tag{7}$$

since the optical flow histogram is normalized and, hence, $\sum_{dir,mag} m_{dir,mag,t}$ sums up to 1. Now, we take $c_{mag,t}$ as a feature for congestion detection for topview cameras. Low values of $c_{mag,t}$ indicate that the velocity of the people is low which might be indicative for a congestestion.

Figure 16 shows two optical flow histograms obtained from the Hermes dataset. The center of mass is marked in red. The histogram in Figure 16(a) was computed at a timepoint, when people were moving fast. Here, $c_{mag,t}$ has a value of 4.5. Contrarily, the histogram in 16(b) corresponds to slow motion and $c_{mag,t} = 1.06$.

Figure 17(a) shows the development of $c_{mag,t}$ over time for all cells of camera 1. See Figure 15(a) for the locations of the cells. One can easily identify different phases in this plot. After approximately 18 seconds, people enter the camera's field of view. They are going quite fast which results in a high value of $c_{mag,t}$. This value decreases to a low value at timepoint 00:40 min when the corridor is congested. This phase of low velocity lasts until timepoint 01:50 min (for cell 0) or 02:20 min (for cell 7), respectively, when people leave the field of view of camera 1. Note the time offset of the curves for different cells.

Now, we apply the algorithm described in section 4.2 to automatically detect change-points in these curves. We use the same parameters that were described previously in section 6.1. We set k to 10 observations and use $l = 10$ observations for estimating a reference distribution. Figure 17(b) shows resulting alarm levels that describes the severity of an alarm. Here, alarm levels greater than zero indicate an increasing value of $c_{mag,t}$, whereas values less than zero indicate a decreasing signal. Our method successfully detects three change points: At timepoint 00:20 min, when people enter the field of view, next at 00:40 min, the beginning of the congestion and around 02:00 min, when people leave the camera's field of view. In particular, our method detects the phase of congestion and reports an alarm level of -0.3, which indicates a decreasing signal and thus is a sign of a congestion. Two false alarms are reported after 12 seconds in cell 0 and cell 2, because the system has not yet integrated enough observations. Note, that similar to Figure 17(a), the time offset between different cells is clearly visible.

We also apply the described procedure to video footage from camera 2 (see Figure 15(b)) with the same parameters as described previously. Our method successfully detects all three change-points without reporting any false alarm.

6.3. Shock Wave Detection

We also tested our approach to detect shock waves on videos obtained from camera 13 (see Figure 3(d)) monitoring the main entry ramp to the festival area. The video shows short sequences of shock waves propagating through the densely packed crowd. Regions of high pressure are automatically detected by the method presented in section 5 and are depicted in Figure 13. Two cells at the border of the frame are correctly classified as alarming cells showing a shock wave. In total, we extracted approximately 4 minutes of video sequences of camera 13 showing a densely packed crowd. Three shock waves can be observed which are correctly detected by our method. However, our methods also reports two false alarms.

7. Conclusion

The stampede at the Loveparade 2010 is a recent example of a terrible crowd disaster where, in spite of all precautions, many people died during a mass event. Research on crowd dynamics and behavior that aims at improving security has a history of more than four decades. The two major directions of research in this context are: first of all, experimental studies that are conducted in order to understand human behavior in crowds and mass evacuation situations and that are used to develop physical models of pedestrian dynamics. Second of all, physical models of crowd behavior are used in simulations that are intended to identify potentially hazardous locations in a mass event. However, in spite of the considerable efforts that went into crowd dynamics research, automatic video analysis systems that would detect dangerous situations such as overcrowding have not been reported yet.

The extensive video footage of the Loveparade stampede enabled us to analyze the dynamics of a crowd disaster using computer vision techniques. We were able to develop automatic approaches to the detection of congestion situations and shock waves that are propagated through a crowd. We conducted feasibility experiments to assess the utility of computationally efficient features. Our approach based on histograms of flow vector magnitude and direction allowed us to automatically detect a re-occurring motion pattern that is characteristic for congestion situations and is due lateral oscillations of people's torsos. This oscillation is a result of the low velocity of people in a congested crowd and can be identified by computing symmetry measures from histograms of optical flow vectors. In case of topview cameras, we di-

rectly make use of the magnitude of optical flow vectors to estimate velocities of pedestrians.

Our method is applicable in many different situations and is mostly independent of local conditions and camera viewpoints. Moreover, since the proposed approach is based on dense optical flow fields and simple measures derived therefrom, it works in real-time and thus meets a crucial prerequisite in video surveillance. Finally, since our method does not necessitate identification or tracking of people, it preserves the privacy of the pedestrians being monitored. Our analysis also revealed motion patterns that occur when shock waves are propagated through a crowd. Here, we developed an approach that detect and locates shock waves again based on the characteristics of flow vectors only. To the best of our knowledge, our system is the first video-based system to detect overcrowding and shock waves in real-time.

Acknowledgements. This work has been supported by the EU funded Research Project PRONTO (FP7-ICT 231738).

References

[1] D. Helbing, A. Johansson, Pedestrian, crowd and evacuation dynamics, in: R. A. Meyers (Ed.), Encyclopedia of Complexity and Systems Science, Springer, 2009.

[2] D. Helbing, A. Johansson, H. Z. Al-Abideen, Dynamics of crowd disasters: An empirical study, Physical Review E 75 (4) (2007) 046109–1–7.

[3] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, L. Xu, Crowd analysis: a survey, Machine Visions and Applications 19 (5–6) (2008) 345–357.

[4] D. Helbing, M. P., Social force model for pedestrian dynamics, Physical Review E 51 (5) (1995) 4282–4286.

[5] C. Burstedde, K. Klauck, A. Schadschneider, J. Zittartz, Simulation of pedestrian dynamics using a 2-dimensional cellular automaton, Physica A 295 (3–4) (2001) 507–525.

[6] D. Helbing, I. Farkas, T. Vicsek, Simulating dynamical features of escape panic, Nature 407 (6803) (2000) 487–490.

[7] A. Kirchner, A. Schadschneider, Simulation of evacuation processes using a bionics-inspired cellular automaton model for pedestrian dynamics, Physica A 312 (1-2) (2002) 260–276.

[8] B. Steffen, A. Seyfried, Methods for measuring pedestrian density, flow, speed and direction with minimal scatter, Physica A 389 (9) (2010) 1902–1910.

[9] A. Seyfried, B. Steffen, W. Klingsch, M. Boltes, The fundamental diagram of pedestrian movement revisited, Transportation Science 39 (1) (2005) 1–24.

[10] X. Liu, W. Song, J. Zhang, Extraction and quantitative analysis of microscopic evacuation characteristics based on digital image processing, Physica A 388 (13) (2009) 2717–2726.

[11] A. Johansson, D. Helbing, H. Z. Al-Abideen, S. Al-Bosta, From crowd dynamics to crowd safety: A video-based analysis, Advances in Complex Systems 11 (04) (2008) 497–527.

[12] T. Zhao, R. Nevatia, Tracking multiple humans in crowded evnironment, in: CVPR, 2004.

[13] G. Brostow, R. Cipolla, Unsupervised bayesian detection of independent motion in crowds, in: CVPR, 2006.

[14] D. Sugimura, K. Kitani, T. Okabe, Y. Sato, A. Sugimoto, Using individuality to track individuals: Clustering individual trajectories in crowds using local appearance and frequency trait, in: ICCV, 2009.

[15] S. Ali, M. Shah, Floor fields for tracking in high density crowd scenes, in: ECCV, 2008.

[16] S. Ali, M. Shah, A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in: CVPR, 2007.

[17] R. Mehran, A. Oyama, M. Shah, Abnormal crowd behavior detection using social force model, in: CVPR, 2009.

[18] K. Still, Crowd dynamics, Ph.D. thesis, University of Warwick (2000).

[19] S. Utler, Die wundersame Schrumpfung der 1,4 Millionen, http://www.spiegel.de/panorama/0,1518,709372-7,00.html (2010).

[20] Spiegel.de, Katastrophe bei der Love Parade, http://www.spiegel.de/panorama/0,1518,708399,00.html (2010).

[21] Lopavent, Dokumentation Loveparade, http://www.dokumentation-loveparade.com (2010).

[22] G. Farnebäck, Two-frame motion estimation based on polynomial expansion, in: SCIA, 2003.

[23] D. Lee, H. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.

[24] B. Krausz, C. Bauckhage, Action recognition in videos using nonnegative tensor factorization, in: ICPR, 2010.

[25] C. Thurau, V. Hlavac, Pose primitive based human action recognition in videos or still images, in: CVPR, 2008.

[26] V. M. De Oca, D. Jeske, Q. Zhang, C. Rendon, M. Marvasti, A cusum change-point detection algorithm for non-stationary sequences with application to data network surveillance, Journal of Systems and Software 83 (2010) 1288–1297.

[27] E. Page, Continuous inspection schemes, Biometrika 41 (1954) 100–115.

[28] Hermes - Investigation of an evacuation assistant for use in emergencies during large-scale public events, BMBF Research Programme, http://www2.fz-juelich.de/jsc/appliedmath/ped/projects/hermes (2011).

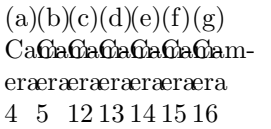

Figure 3: Views of the scenes as recorded by the surveillance cameras [21]. See Figure 2 for the locations of the cameras. In the work reported here, we concentrate on camera 15 for the detection of congestions and on camera 13 for the detection of shock waves.

Figure 4: Dense optical flow computed using the method in [22]

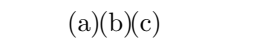

Figure 5: k-means clustering ($k = 5$) of two-dimensional histograms of motion direction and magnitude in an entire frame. Figure (a) shows the cluster assignment (encoded in colors) for camera 15. Figure (b) shows considerably more motion mostly in a direction of 250°. Shortly after 16:20h, the frames are assigned to the cluster centroid depicted in Figure (c). This cluster contains vectors of little motion, mostly in directions 90°and 270°.

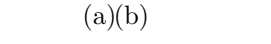

Figure 6: Two basis histograms extracted by applying NMF to 2D histograms of motion direction and magnitude obtained during a time interval of 30 seconds at 14:40h for a grid cell in the foreground.

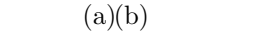

Figure 7: Two basis histograms extracted by applying NMF to 2D histograms of motion direction and magnitude obtained during a time interval of 30 seconds at 15:53h for a grid cell in the foreground. At that time, a police cordon stopped the visitors, so that only little motion can be observed. Just a few visitors are not stopped by the police resulting in optical flow histograms such as Figure 7(b).

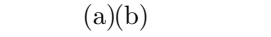

Figure 8: Two basis histograms extracted by applying NMF to 2D histograms of motion direction and magnitude obtained during a time interval of 30 seconds at 16:28h for a grid cell in the foreground. Figure (a) shows motion to the right (270°), whereas Figure (b) shows motion to the left (90°). At that time, the visitor density was very high.

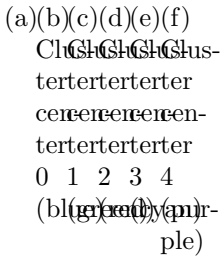

Figure 9: K-Means Clustering ($k = 5$) of basis histograms obtained by applying NMF. Figure (a) shows the cluster assignment for the entire video sequence. Figures (b) to (f) show the five cluster centers. The cluster centers shown in Figures (b) and (f) show distinct motion peaks with a high magnitude and a motion direction of approx. 250 °. This motion pattern is characteristic for time intervals of normal pedestrian flow. Figures (c) and (e) show no distinct motion; these basic motion patterns correspond to the situation observed after the police cordon has been established. The cluster center in Figure 9(d) shows left- and rightward motion. It is characteristic for motion observed between 16:22h and 16:40h when the pedestrian density is very high and people are stepping from one foot to the other in order to keep their balance.

Figure 10: Given the current observation Y_i, a reference distribution is obtained from historical data $Y_{i-k-l}, \ldots, Y_{i-k}$. The upper and lower α-percentiles of the reference distribution ($Q(\alpha)$ and $Q(1-\alpha)$) are used in equation 2 for calculating cusum statistics. Here, α controls the degree a deviation from $Q(\alpha)$ or $Q(1-\alpha)$ is considered as critical.

Figure 11: Cusum Statistics S_i^- for camera 15. Whenever S_i^- exceeds the threshold Θ, an alarm is raised. The threshold Θ has been determined automatically by using a bootstrap resampling method.

Figure 12: Slope test for cusum statistics. At each timepoint a, a linear regression model is fitted to a sliding window of cusum values $\{S_i^+\}_{i=n-\nu+1}^{n}$ for $n = a, a+1, \ldots$, here: $\nu = 8$. The end time of the alarm is detected at timepoint $t_a = 12$, when the slope of the regression line (labeled as regression line 10) is less than zero for the first time. Additionally, we measure the severity of an alarm as the angle of the regression line divided by 90°.

Figure 13: Detection of Crowd Turbulences. Camera 13 which monitored the main entry ramp shows short sequences of shock waves that are propagated. We automatically detect regions of high pressure which are colored red in the above Figure. Note that the saturation of the alarming cells have been increased for the sake of visibility.

Figure 14: Detection of anomalies in the pedestrian flow at the Loveparade 2010 recorded by camera 15. Alarms raised by our system are colored in red for a abnormally decreasing signal, whereas jumps to a high level are depicted in green. In the lower part, the computed severity $L \in [0 \dots 1]$ of the alarm is shown. Clearly, our system has detected abnormal events that deviate from typical observations. In particular, our system raises a severe alarm at about 16:27 when the crowd is jamming up.

(a)(b)

Figure 15: Setting of the Hermes dataset [28]. The scene is observed by two cameras. A grid of cells is placed onto the frame, where each cell covers an area of approximately 1m x 3m except for cell 0 in (b) which covers approximately 1m x 1.2m. Note that the fields of view of the cameras overlap: Cells 5 and 6 seen from camera 2 (Figure (b)) correspond to cells 0 and 1 observed by camera 1 (Figure (a)).

(a)(b)

Figure 16: Two different optical flow histograms obtained from camera 1 of the Hermes dataset. In (a), people are moving fast, whereas in (b), they are going slowly. Instead of using the feature described in section 4.1, we directly make use of the magnitude of optical flow motion vectors and compute the center of mass of the optical flow histogram which is marked in red.

(a)(b)

Figure 17: The value of $c_{mag,t}$ over time for all cells (see Figure 15(a)) of camera 1 are shown in (a). One can easily identify different phases in this plot. Interestingly, the time offet of the curves for different cells is clearly visible. Figure (b) shows alarm levels reported by our change-point detection algorithm. Alarm levels greater than zero correspond to an increasing value of $c_{mag,t}$, lower values indicate a decreasing signal.