MASTER THESIS

ASSESSMENT OF AUTOMATED CROWD BEHAVIOUR ANALYSIS BASED ON OPTICAL FLOW

Ingo Stijntjes

FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE
SERVICES, CYBERSECURITY AND SAFETY

EXAMINATION COMMITTEE
Prof.Dr.Ir. R.N.J. Veldhuis
Dr.Ir. L.J. Spreeuwers
Prof.dr.ir. C.H. Slump
C.G. Zeinstra MSc.

DOCUMENT NUMBER
SCS - 19

UNIVERSITY OF TWENTE.

10-1-14

Assessment of Automated Crowd Behaviour Analysis Based on Optical Flow

Ingo Stijntjes[1]*

***Abstract*—In visual surveillance, camera streams are often used to keep an eye on dense crowds. The examination of this data is mostly done manually by observers. When analysing multiple cameras some assistance is desirable. Computer vision methods can be used to assist observers in detecting crowd behaviours. Methods based on optical flow are particularly interesting since they can examine high density crowds with cluttering and (partial) occlusion without increasing computing costs. Not many methods can detect specific behaviour of dense crowds without the need of a learning stage. One promising method by Solmaz et al. uses the Jacobian stability of the optical flow field in the scene to detect five behaviour patterns viz. blocking, bottlenecks, fountainheads, rings and lanes. The method is implemented and a demo program is written with which experiments are performed on several datasets. The detection of three out of five behaviour patterns turn out to be promising, for the latter two improvements are proposed.**

I. Introduction

Everywhere crowds gather there is an increase in potentially dangerous situations, which clearly described by S. A. Velastin et al. [11], *"do not per se constitute an uncontrollable condition (an "incident"), but could become so with an unexpected event"*. One can see big crowds as entities having a high inertia; difficult to direct or to stop.

There is a trend of organising many events in western Europe, some of them attracting several thousands of people. At these big events crowds are monitored through video feeds by one or more observers. If needed security services can be deployed efficiently or crowds can be redirected to other areas. There are numerous incidents in the past where stampede or panic in dense crowds led to casualties, see [15] for a shortlist.

The analysis of crowds is a complex problem and it is very much dependent on the experience and concentration of the observer or simply on the observer to camera ratio. Since there is much data from the video feeds, computer vision methods are being used to assist an observer in doing his or her task. It should be kept in mind that some computer vision methods that are proven in scenes with individuals are likely to fail in high density situations due to cluttering, (partial) occlusion, computing costs or other factors. Analysis based on optical flow is particularly interesting, since this is a feature tackling these problems.

With respect to public safety, interesting behaviour patterns to detect are blocking, bottlenecks and panic. Blocking be caused by a stumbling person and bottlenecks by doorways exceeding their capacity. The method proposed by Solmaz et al. is able to detect behaviour patterns closely related to these.

The goal of this research is to develop a method that assists crowd observers by detecting the aforementioned behaviour patterns and extracting crowd features that are relevant for an observer. This is done by assessing current methods in crowd behaviour identification, use this assessment as base to propose a method that suits the needs of a crowd observer, implement the proposed method and perform experiments.

B. Zhan et al.[7] introduce a set of applications of interest for automated crowd analysis, for which this research can be valuable. These, non-disjunct applications are:

Crowd management - Where big crowds gather there will be an increased risk of potentially dangerous situations. Oppression and stampede in big crowds can easily lead to casualties or even death. It is valuable to analyse the crowd to develop crowd management tactics.
Public space design - When designing a public space it should be able to withstand the crowds that are expected to be there. Understanding crowd behaviour in several situations is a very important step in the final design.
Virtual environments - Having good crowd models can aid in simulating crowd behaviour in for example video games or animations.
Visual surveillance - The (autonomous) detection of anomalous behaviour or even specific crowd behaviours in crowds can aid for example public safety enforcers.
Intelligent environments - In intelligent environments the characteristics of a crowd can help in decisions to direct (part of) the crowd to a different place.

In this paper a short survey of relevant literature is presented in Section II. In Section III a promising optical flow based method is discussed and improvements to it are proposed. The implementation of this proposed method is presented in Section IV, after which experiments are defined and performed in Section V. The results of the implementation turn out to be promising for some of the behaviour patterns. During implementation and experiments possibilities for improvements are identified, some of which supported by an initial experiment. These additions can be found in Section VI. This work will close with conclusions and ideas for further work in Section VII.

*This work is part of the author's Msc project

[1]I. Stijntjes is with Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, 7500 AE Enschede, The Netherlands istijntjes at gmail.com

[2]Special thanks to my committee: Prof.Dr.Ir. R.N.J. Veldhuis, Dr.Ir. L.J. Spreeuwers, C.G. Zeinstra MSc. and external committee: Prof.dr.ir. C.H. Slump

Method	Training	Crowd Density	Behaviour Types	Keywords
Velastin et al.[11]	No	High	Occupancy level	Foreground/background classification
Andrade et al.[8][1]	Yes	Medium	Normal, abnormal, blocked exit	Optical flow, HMM
Rodriguez et al.[3]	Yes	High	Unknown	Motion pattern matching using optical flow
Wang et al.[4][10]	Yes	Low	Normal, abnormal	Feature point optical flow based motion pattern, historical data
Szczodrak et al. [5]	Yes	Low	Hold ups in doorway	Neural classifier, fuzzy logic
Liu et al. [9]	No	High	> 4 singular patterns	Applications in fluid analysis
Ali et al. [13]	No	High	Stable, unstable	Lagrangian coherent structures in flowmap
Solmaz et al. [2]	No	High	5 behaviours	Optical flow, Jacobi stability

Table I: Summary of methods discussed in Section II. Crowd density is defined as: Low < 50 pedestrians per scene, Medium < 100 per scene and High > 100 per scene. The latter can include several thousands of pedestrians.

II. Crowd Analysis in Literature

Visual surveillance is an active field of research in computer vision. There is much research done on topics including identification, access control, anomaly detection, tracking, behaviour understanding and so on and so forth.

In early work by Velastin et al. [11] a system is proposed that assists observers of a Closed Circuit TV (CCTV) system of the underground railway in Liverpool to detect situations of interest, including the occupancy level. This is achieved by simple background removal using a shifting mean implementation to get a model for stationary background pixels. After removal, the ratio of background and pedestrians pixels is an approximation for the occupancy level. At high levels, i.e. > 60 %, a texture based foreground background segmentation is used. Directions and velocities are derived by a dedicated system that calculates the optical flow of pedestrian pixels.

Andrade et al. [8] present work on a framework for analysing crowd videos using Hidden Markov Models (HMMs) to encode behaviour based on training data. In [1] they handle crowd scene modelling for event detection. They use unsupervised feature extraction from the optical flow of crowds to encode normal behaviour. This model is used to detect a blocked exit.

Rodriguez et al.[3] propose a data driven crowd analysis algorithm in their work. Behavioural priors are learned offline from a large database containing specific behaviours. During the analysis the video is divided into smaller blocks for which the motion pattern is matched to the learned priors. These matches are then used in a tracking framework that fuses this data with tracking algorithms using a linear Kalman filter.

Wang at al. [4] cope with the dense crowd not by tracking individuals, but certain feature points, in this case Kanade-Lucas-Tomasi (KLT) corners. The video is divided into blocks of 16x16 or 32x32 for which motion information is extracted. This information is used to construct motion patterns for matching. In this work velocity normalization is included to handle image distortion from cameras that are not normal to the surface of the crowd but have a bird's eye view. The method is able to detect abnormal behaviour in datasets after a learning stage with normal data. The results of this method heavily depend on block size, thresholds and the weights used in the deviation measure. The same authors publish an extension in [10] where they propose to use historical data in the decision stage. Behaviours that occurred several times in the past at a certain location are more likely to be normal than other more rare behaviours.

Szczodrak et al. [5] focus on crowd behaviour analysis with in particular the detection of hold-ups in doorways. They propose and implement a method that uses feature point optical flow data of crowd behaviour to train a neural classifier for the detection of these behaviours. Based on these behaviour classifications a fuzzy logic framework aids in the detection of hold-ups. The algorithm is parallelized and implemented on GALERA, a computer cluster using KASKADE [6], a platform that handles cluster resource management resulting from previous work. This practical implementation adds much value due to the fact that often proposed methods in this field are tested on single videos and the reality is that there are a high number of data feeds at for example train stations and festivals. This clustered implementation improves computation time drastically and therefore realizes a framework for applications. The method employs corners in the images as features that are tracked between two consecutive frames to reduce computational costs. The crowd videos used as benchmark have a low density of people.

From another field of interest the analysis of flow fields is also getting attention. In fluid dynamics and the meteorological field much data is presented as vector flow fields in two or more dimensions. Liu et al. [9] present a method to find singular patterns in such flow fields invariant of scale and rotation using an orthogonal basis of multiscale singular patterns. The method is able to detect, for example, vortices in satellite images, but could be quite easily adapted to cope with flow fields of moving crowds.

Ali et al. [13] propose to use the optical flow field of a crowded scene to separate and detect changes in dynamic behaviour. The vector flow field is used to perform a particle advection. With the resulting particle flow map a Finite Time Lyapunov Exponent (FTLE) is calculated on which Lagrangian Coherent Structures (LCS) are identified. On the 2D flow map LCS indicate the regions that separate dynamic behaviour in a time varying flow field these are therefore valuable features in flow segmentation. The flow is monitored in time and any change in dynamic behaviour will give rise to new LCS which indicate the position where new flow segments can be developed.

Solmaz et al. [2] also utilize the optical flow of a crowded scene, but now in order to identify specific behaviours.

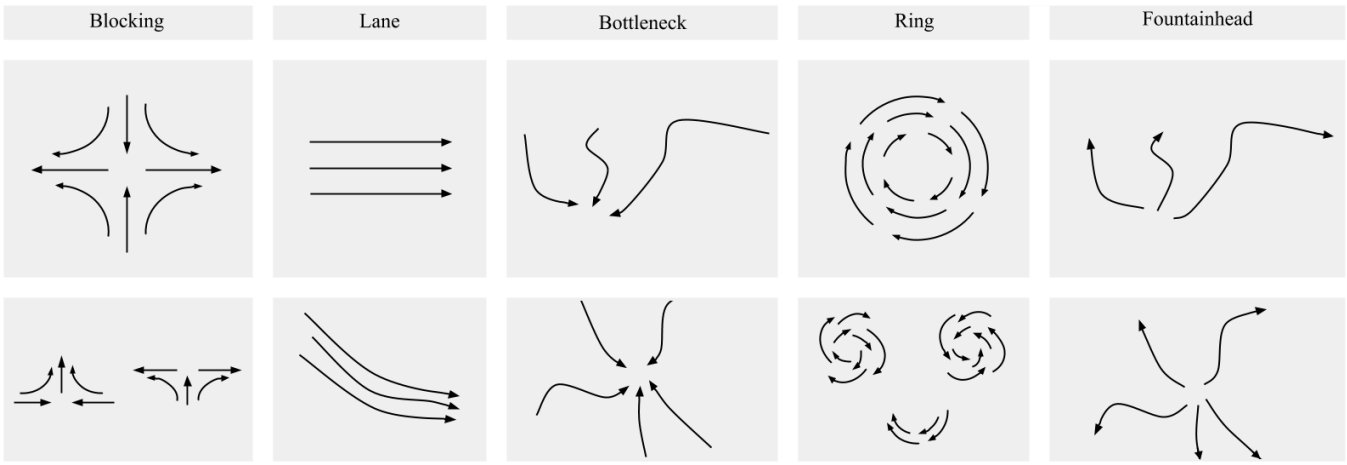

Figure 1: Behaviour types definitions, based on the stability analysis proposed in Solmaz et al. [2]. Each column is a specific behaviour. The first row represents the general behaviour from which variants are derived showed in the second row. Most variants are components of the general behaviour, except for some ring variants, these can flow inwards or outwards while rotating.

The method also performs a particle advection, where the particles can evolve towards accumulation points but can also contribute to regions of higher or lower particle density. All these points and regions are analysed based on some features of the trajectories and particles involved. Candidate points follow and the Jacobi stability is assessed in the region of those points, resulting in a possible detection of one of five basic behaviours; viz. bottlenecks, fountainheads, blocking, lanes and rings.

Methods based on optical flow have the advantage that computational costs will not increase when the number of people in the frame increases, which would be the case when using, for example, face detection to find crowd motion. Proposals that require learning will need an extensive training set and data acquisition is much work, especially for big crowds demonstrating a characteristic behaviour pattern. Because the behaviours detected by Solmaz et al. well match the goal of this research and the method relies on optical flow computations this approach is selected to be assessed in more detail. In Table I an overview of the literature study is given.

III. Optical Flow Based Method

The approach proposed by Solmaz et al. is designed to detect five different kinds of behaviours in video clips showing crowds of varying density. The behaviours are based on the influence of different stability features of the flow field. These features are defined by different eigenvalues, see Table II, and all cases represent a specific flow field. For example, in Table III it can be seen that a fountainhead behaviour depends on label R and A. These labels represent a source and outgoing swirl in the flow field respectively. When combining these patterns fountainheads are formed. The result of this procedure can be seen in Figure 1. These behaviour patterns definitions are slightly different from Solmaz et al., where only the top row is used. The names are adopted from Solmaz et al., but it can be debated whether the patterns are consistent with the expectations of the name.

The proposed algorithm here is inspired by Solmaz et al. and is based on defining a dynamical system by the optical flow of a crowd scene and checking the Jacobi stability at candidate points. It can be divided into three parts. The first part includes the definition of a dynamical system using the optical flow of the crowded scene. The second part handles the search of candidate points in the flow field, and the last part is assessing the stability at regions specified by these points through the eigenvalues of the Jacobian of the flow field linearized around a candidate. The method considers the influence of the candidate points on its surroundings.

A. Dynamical System

Since this method heavily relies on the optical flow of a crowd scene, defining a dynamic system based on this feature is the first step in the algorithm. Optical flow is the apparent movement of pixels in the video and is calculated using two consecutive frames. Using optical flow for the determination of motion makes it possible to analyse very dense crowds with high occlusion due to the fact that the crowd flows are treated as entities rather than individuals, this is sometimes referred to as a holistic approach. Furthermore, methods based on face recognition or texture analysis for example, will either fail or be too computational expensive. Another advantage of considering the optical flow instead of individual motion tracking is the inherent privacy of the individual. The optical flow of a video clip is integrated in time to construct a single vector flow field. The approach considers the continuous dynamical system:

$$\dot{w} = F(w) \tag{1}$$

Eigenvalues	Conditions	Label	Count
Real, $\lambda_1 > 0, \lambda_2 < 0$	$\Delta < -\epsilon^2$	Green	G
Real, positive	$\tau < -2\epsilon$ $\tau^2 > 4\Delta$	Red	R
Real, negative	$\tau > 2\epsilon$ $\tau^2 > 4\Delta$	Yellow	Y
Complex, positive real part	$\tau > 2\epsilon$ $\tau^2 < 4\Delta$	Magenta	A
Complex, negative real part	$\tau < -2\epsilon$ $\tau^2 < 4\Delta$	Cyan	C
Purely imaginary	$\lvert\tau\rvert < 2\epsilon$	White	W
at least one zero	$\lvert\Delta\rvert < \epsilon^2$	Blue	B
all zero	$J_F = 0$	Black	K

Table II: Labelling of candidate regions. This table is the corrected version of the one found in [2]. Count is the number of instances in the area under assessment.

with $\dot{w} = [u(w), v(w)]^T$ and $w(t) = [x(t), y(t)]^T$ denoting particle positions and particle velocities respectively. The first step is to find critical points w^* satisfying $F(w^*) = 0$. Critical points define the area in which the stability is assessed. Considering small disturbances denoted by $z = w - w^*$ rewriting and using Taylor's theorem yields:

$$F(w^* + z) = F(w^*)z + H.O.T. \tag{2}$$

Setting $F(w^*) = 0$ and discarding higher order terms of the Taylor expansion yields:

$$\dot{z} = J_F(w^*)z \tag{3}$$

With J_F the Jacobian matrix given by:

$$J_F = \begin{pmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix} \tag{4}$$

The solution space of (3) is determined by the initial conditions and eigenvalues of the Jacobian. In the stability analysis the determinant and the trace, given by $\Delta = \lambda_1\lambda_2$ and $\tau = \lambda_1 + \lambda_2$ respectively are used, with λ_1 and λ_2 the eigenvalues of J_F.

B. Candidate Selection Procedure

The selection procedure of candidate points is mainly based on particle advection and trajectory analysis. Particle advection is a technique where a grid of massless particles is overlaid on the flow field and is propagated along the flow resulting in a density map of the system. High density peaks are found and marked as accumulation points and trajectories of attributing particles are clustered and classified as small or big clusters. Accumulation points with a single narrow trajectory reaching it are labelled as a possible lane or ring. When an accumulation point is reached through several trajectories with different angles it is marked as a bottleneck or fountainhead. High density regions are labelled as blocking candidates. For the trajectory analysis the particles must meet some requirements; they have to reach an accumulation point and travel a minimum amount of distance.

Behaviour	Count ratio
Lane	$B/T > L$
Blocking	$G/T > L$
Bottleneck	$(Y + C)/T > L$
Fountainhead	$(R + A)/T > L$
Ring/Arch	$(W + A + C)/T > L$

Table III: Ratio conditions for count statistics. This table is adopted from [2] and corrected to follow the definitions stated in Figure 1. T is the total amount of pixels in the area considered. The threshold L is determined empirically.

C. Stability Analysis

As mentioned above the influence of a critical point on its surroundings is fully determined by the initial conditions and eigenvalues of J_F. To check which behaviour is coherent, Δ and τ are calculated and every point in a region around a candidate is labelled using Table II.

The labels each correspond to solutions of (3) and thus represent characteristic local flows that can contribute to one of the behaviours depicted in Figure 1. For example, real, negative eigenvalues match a stable sink in the flow field. This point clearly contributes to a bottleneck behaviour. The result of this assessment is shown in Table III.

IV. Implementation and Demo Program

The method presented in the previous section was implemented, during which some challenges had to be overcome. The paper presented by B. Solmaz et al. was not a detailed design report carefully treating every step in detail. Every part of the demo program is designed from scratch. From the start carefull attention is paid to the presentation of data, which can vary for different fields of application. During implementation some additional features are implemented. For example, the rectification of camera images allowing the projection of crowd flow on a two dimensional map.

A. Dynamical System

The first part of the algorithm concerns the construction of a dynamical system based on the optical flow. Before calculating the optical flow the user is prompted to create a mask and to add marker points. The first is used to discard uninteresting regions on the image plane and save CPU resources. The latter is used for image rectification. In the rectification step a projective transformation is applied to construct the flow field perpendicular to the ground plane in the scene. For this the algorithm needs matching point pairs to estimate the transformation matrix.

There are different approaches to estimate optical flow in an image sequence; Beauchemin et al. [16] present a survey on different techniques. The basic concept of optical flow estimation is pixel intensity conservation, also called brightness constancy, where it is assumed that an object does not vary in intensity in two consecutive frames. Let $I(x, y, t)$ be the pixel intensity at pixel location (x, y) and time t, when satisfying the constant intensity constraint it can be written

that:

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{5}$$

With δx and δy denoting the spatial displacement and δt the temporal displacement. The left hand side of (5) can be developed by a Taylor expansion, ignoring higher order terms and rewriting yields:

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0 \tag{6}$$

With u and v the pixel velocities that represent the optical flow. Equation (6) is known as the optical flow constraint [16]. This is an equation with two unknowns and needs an additional constraint to be solved. It is this additional constraint in which the different methods vary. In [16] various methods are described and Barron et al. [17] assess the performance of several methods. The fundamental problem in image based velocity approximation is that the images contain information from the optical domain; intensities, which are depended on illumination. The velocity, however, is a geometric quantity independent of lightning conditions thus making the method an estimation. However, acceptable results can be obtained when considering small motions, see [17] for a detailed performance analysis of different optical flow techniques. The appropriate method in this paper is chosen to be the Horn and Schunck method [14] due to its dense optical flow field generation, acceptable error and availability of implementations. The Horn and Schunck method applies as extra constraint a global velocity smoothness assumption, combining the two constraints in an error function yields:

$$\begin{aligned} E(u, v) = \int_D \Big(& (\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t})^2 \\ & + (\lambda(||\frac{\partial u}{\partial x} + \frac{\partial u}{\partial y}||^2 + ||\frac{\partial v}{\partial x} + \frac{\partial v}{\partial y}||^2)\Big) dxdy \end{aligned} \tag{7}$$

This is a least square minimization problem, solved iteratively, with u and v the velocities, x and y the components of the image plane D and λ the weight of the smoothness term. The optical flow is calculated over a video clip and is integrated in time to obtain a single vector flow field demonstrating the behaviour of the total clip. This field can be used to construct the Jacobian depicted in equation (4).

B. Candidate Selection

Selecting candidates is very important in this method; this procedure generates some points with a high likelihood of exposing one of the five behaviours. Without this pre-selection higher misclassification rates will occur due to false interpretation of the flow field, leading to the stability assessment of regions that had insufficient influence on the particle field. This particle field is a key feature in this pre-classification phase and is the result of a simple particle advection as depicted by:

$$w(t + 1) = w(t) + F(w(t)) \tag{8}$$

Basically, the implementations looks for three different features from the particle advection, viz. accumulation points, high dense areas and trajectories.

1) Accumulation Points: After some advection time a density plot is generated by replacing each particle with a Gaussian function. This map is used to find local maxima. These points correspond to high density points that might be accumulation points depending on the advection time and the size of the Gaussian producing the density map. An increased advection time will lead to efficient sink seeking, since many particles find their equilibrium in an accumulation point, opposed to shorter advection times that will more likely reveal high density areas of particles before they reach an accumulation point. There are several local maxima found in the density map which are filtered on height relative to the highest peak, resulting in strong accumulation points. Accumulation points are used in finding ring and lane candidates. Detecting lanes can be very valuable for observers to get insight in the preferred paths of people. This data is, for example, especially important in the event of a train station; the civil engineers involved in developing the structure can use this info in their design choice.

2) High Density Areas: High dense areas are found in a similar way to the previous feature, except with a decreased advection time. This way regions are revealed with, for example, due to blocking, temporal higher particle density which would be invisible in a saturated advected field. Furthermore, the size of the Gaussian is slightly increased to find high density regions rather than high density points. Due to the similarity of these regions and accumulation points it is expected that they are also found near each other. In this case, both of the points are taken into account; one labelled as accumulation point and the second as high density region.

3) Trajectories: Among other particle features, the trajectories are observed. The length of the trajectories determines whether a particle and its trajectory is taken into account in the selection. When an advected particle is close to its starting point it does not belong to a characteristic behaviour. This length limit is empirically set to 40 pixels, but it has to be taken into account that this number is highly depended on, for example, video resolution and the scale of the scene. Another check is the endpoint of a particle. When particles end up isolated after being evolved on the flow field it is an indication that this particle does not belong to a dominating flow. Particles are of interest when their end point lies near an accumulation point, the limit is experimentally set to 50 pixels, but again this number greatly depends on the resolution and should be related to scene properties, i.e. a distance in meters. The directions of particles reaching accumulation points are clustered and used to separate unidirectional accumulation points from omnidirectional accumulation points.

4) Selection: The features described in the previous sections are used to label located candidates. This labelling process is described in Figure 3. With the clustering of trajectories by their angles a distinction between an omnidirectional and unidirectional accumulation point is made. The first case

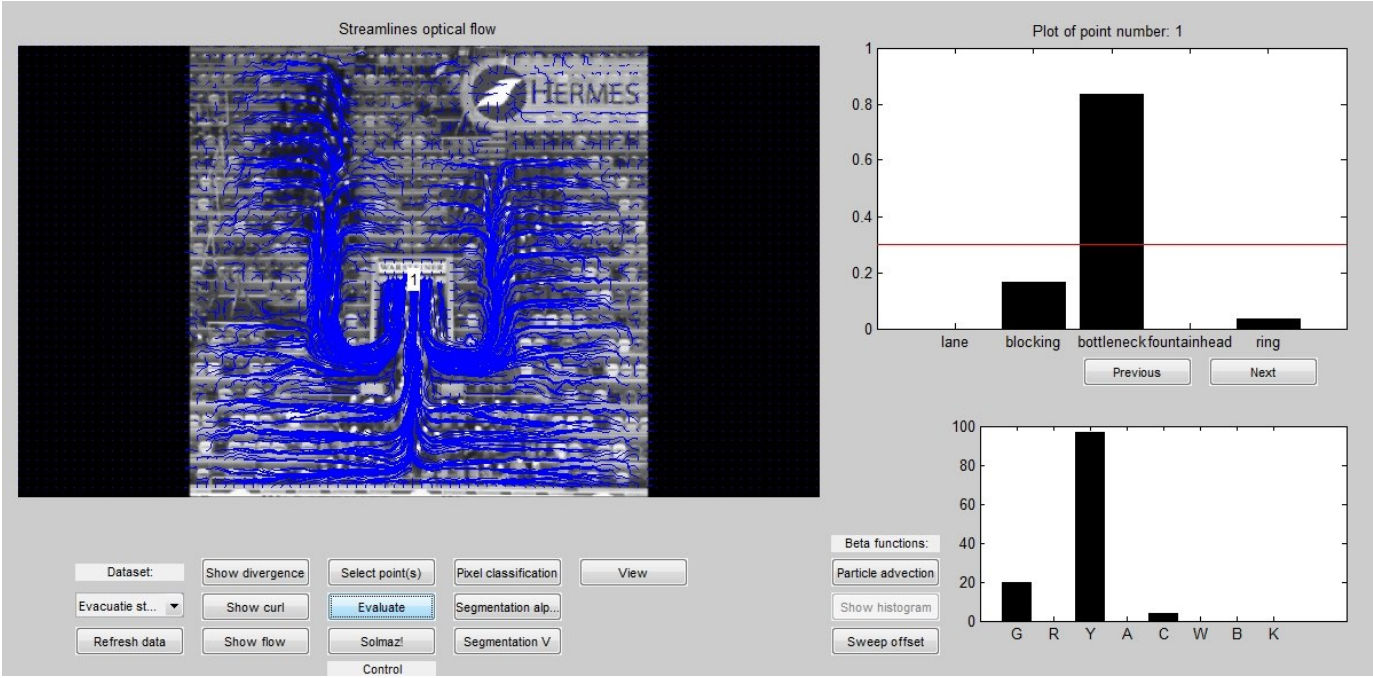

Figure 2: Front panel of demo tool. The left view is used to visualize the flow field and particle advection, the upper right figure shows the method results and the lower right the stability features of each candidate.

occurs when several small and big clusters are contributing to the accumulation point, the latter case when there is only a single cluster.

C. Stability Analysis

The last features that are used in the classification are given by the stability analysis of the area of a candidate point. At every position where a behaviour is expected an area around this point is evaluated. For all pixels in this area the Jacobian is constructed, the eigenvalues considered and a label assigned following Table II and III. This results in a score for each behaviour in the area under consideration, this score has to exceed a given threshold and in combination with previous features a behaviour is assigned. The threshold level is empirically defined to be 0.3.

D. Graphical User Interface

In order to maintain and demonstrate the program a Graphical User Interface (GUI) is designed and programmed. It is build modular to be of use for testing. The GUI consists of two figures embedded in one window with its controls at the lower side, see Figure 2, the left figure plots mainly end results of algorithms while the smaller right plots show intermediate results of the algorithm. The demo program is used for testing parts of the implementation and for demonstrating purposes. Since the flow fields are very dense, one vector per pixel, the field is divided into blocks in which the flow is averaged, threshold filtered, scaled and plotted in its centre resulting in a user friendly flow view. Besides the main application, also some debug tools are developed. For example a graphical tool that allows manual inspection of the flow field stability.

V. Experiments and Results

The goal of the experiments is to tune the parameters in the method and to assess if detected behaviour patterns match the definitions from Figure 1. Some video material of crowds is acquired to perform the experiments. These videos have to meet some requirements involving an appropriate resolution, some characteristic behaviour and a steady camera. The first two are evident, the necessity of a steady camera is due to the fact no optical flow correction is implemented to compensate for camera movements, since this lies outside the scope of this research. With these requirements five datasets are selected and used to assess the performance of the implementation. Since the experiments are done on different datasets than the results presented in [2], the results can only be compared to a ground truth generated by hand obeying the definitions in Figure 1. The results of the method will be assessed by answering the following two questions: *Are the resulting behaviours to be expected at that position?* and *Are additional behaviours expected in the dataset?* Of course when answering the first question a position error can be obtained by measuring the distance between the ground truth and the found behaviours, due to the different scales of the data, common sense is used to determine if an error is too high.

A. Datasets

When dealing with behaviour patterns in dense crowds it is difficult to obtain much data satisfying the condition depicted above. There are some datasets containing crowds that depict several normal behaviour but when dealing with abnormal or

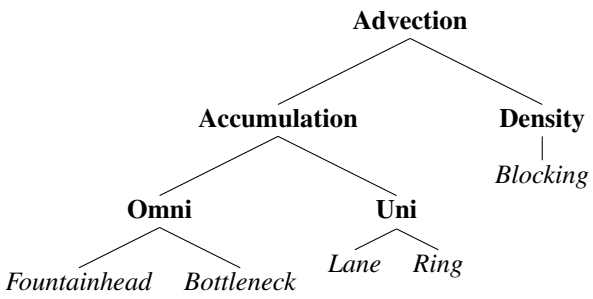

Figure 3: Decision tree for candidate selection.

extreme situations data is scarce. In this paper three different ground truth sources are used to verify the method viz.:

- **Real data**, publicly available, e.g. panic at the memorial service 2010 in Amsterdam, The Netherlands;
- **Synthesized data**, e.g. instructed evacuation of a stadium;
- **Computer generated data**, e.g. computer animations of crowds evacuating a building.

This data is retrieved from public sources and from non public sources with courtesy of other researchers in this and related fields. In Table IV the used data is enumerated with a short description.

B. Parameter Tuning

The results of the method are highly dependent on some user defined parameters. The threshold value L, for example, is dominant in detecting the behaviour since the score of the behaviour has to exceed it. This value has to be determined by means of experiments. In the conducted experiments a threshold of $L = 0.3$ is maintained. The size of the area in which the stability will be assessed will be dependent on the scale of the video. It would make no sense to make this region a great deal smaller than, for example, the size of one person. The method is designed to cope with global behaviour rather than individual behaviour. The area should be scaled likewise. What these sizes should be has to be determined empirically. In the conducted experiments an area of several people is used. Also noticeable are the steps taken in the advection phase, the maximum distance parameter and travel limit. When advecting too many steps, saturating the particle field, temporal higher density regions are missed, but accumulation points are more clear. This effect yields a multi stage advection where the flow field is advected in several stages. Each stage represents an advection length and each stage should be assessed by the method and results should be fused. An informal experiment conducted on the *Complex* dataset illustrates some promising results leading to the detection of some more ground truth bottlenecks. The maximum distance and travel limit parameters define the distance between a particle and an accumulation point after advection and minimum travelled distance of every particle respectively. This way particles that end up alone or reside close to their starting position, stating a weak characteristic flow, will be discarded and have no influence on the candidate selection. These parameters heavily depend on the scale of the video and could be related to a real world measure, so instead of using pixels one can use meters. Overall the tuning of these parameters is complex and moreover dependend on the observer's need. Beside some parameters that mainly depend on video and location properties, there will be a trade-off between a missed detection and misclassifications. It depends on the application which one should be appointed a higher weight.

Dataset	Description	Type	Density
Stadium[18]	Staged evacuation stadium	S	292
Damschreeuwer [19]	Panic in dense crowd	R	>60k
Complex [20]	Evacuation of building	C	>1800
Bottleneck [21]	Escalator entrance at station	R	>300
Stampede [22]	Evacuation of room	C	186

Table IV: Some datasets used in experiments. Types: R = Real, S = Synthesized, C = Computer generated. Density is measured in people and in some cases approximated

C. Results

All these datasets are processed by the above mentioned method. Only the *Damschreeuwer* and the *Complex* datasets are rectified. The processing results are shown in Figure 4. To answer the first of the aforementioned questions with regard to all the datasets, the location of detected behaviours is assessed. It is found that these behaviours are at reasonable positions. It can be debated whether some of the bottlenecks in the *complex* dataset are at the right positions, especially the top two can be expected to be closer to the doorway. The second question is more difficult to answer. There are clearly formed lanes in the *complex* dataset, it is expected the method finds them. However, this is not the case. This is mostly due to the selection of the candidates for which, in the next section, some alternative selection algorithms are discussed. Although data is scarce, the results are evaluated and published in Table V. Since parameters had to be tuned on the data the *Stampede* dataset was left outside this scope to verify the obtained parameter values. The scale is roughly comparable to the *Stadium* and the parameter values from the tuning on this set are adopted. The result of this test is given in Figure 5

1) Bottleneck: In the case of bottleneck detection the implementation yields a relative high misclassification. When looking closer to the results it can be seen that instead of finding one bottleneck, as the ground truth states, several bottlenecks are found at the vanishing line of the people. Two of them fairly close to the ground truth, the others are labelled as misclassifications, since their distance to the ground truth is significant. When taking a closer look at the flow field it can be debated whether the ground truth obeys the definitions, different streams can be appointed that lead to a bottleneck. Taking this all into account the detection of bottlenecks is rather strong. On the other hand there are some missed detections. There are some parameters that have significant

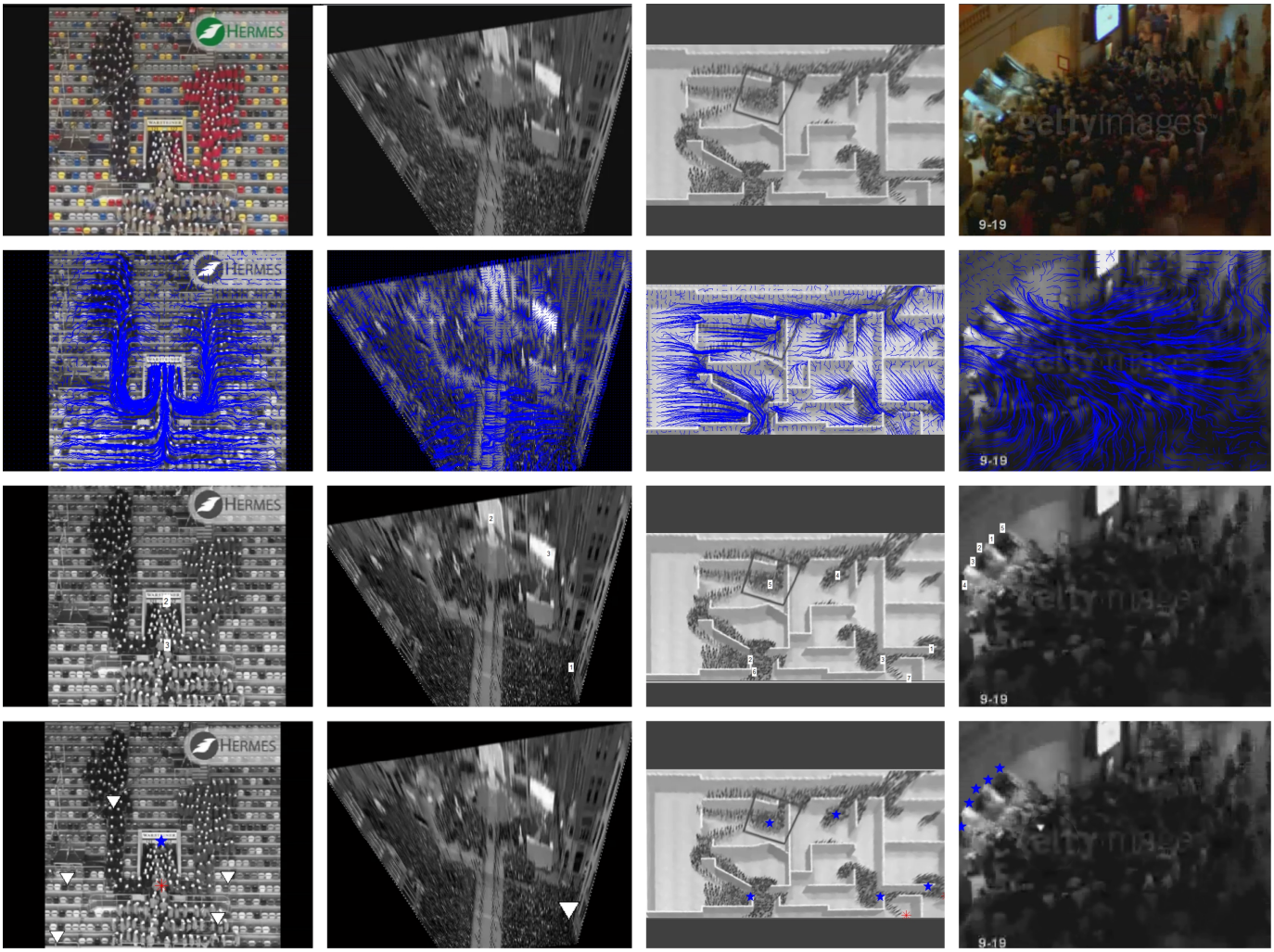

Figure 4: Results of the implementation, each column is one of the datasets depicted in Table IV. From left to right: *Stadium, Damschreeuwer, Complex* and *Bottleneck*. The first row is a frame of the video. In the second row the advection trajectories are plotted. The third row is the output of the candidate selection stage for blocking, lanes, rings and bottlenecks. The fourth row are the final results; with bottlenecks marked with $\star$, blocking with $*$ and fountainhead with $\bigtriangledown$.

influence on the output of the method, in this case the amount of advection steps taken is of great influence. Depending on the application, the user can tune this and other parameters to get an acceptable balance between misclassifications and missed detections.

2) Blocking: The blocking detection has detected the ground truth blocking accurately. However, two misclassifications are generated. These can be attributed to the event where particles are following a flow to the end of the image. The particles are slightly drifting at that position, due to noise and small flows, as though they are blocked. It is therefore not unexpected to find this behaviour, but it still leads to a misclassification. This effect should be detected and compensated for.

3) Fountainhead: From the two fountainheads in the ground truth data both are detected. However, there is a relative high misclassification. This all occurs in the *stadium* dataset. The positions may not be surprising since they all are positions where people come from. However, in the definition of the method it is stated that a fountainhead is a single point from which particles, in this case people, diverge. It can be debated whether the candidate selection or the ground truth is correct, but overall the result is promising.

4) Lane and Ring: Taking a closer look at Table V will reveal the weak performance of the lane and ring detection. The two positions where rings are to be expected by the ground truth are assessed by hand and indeed the stability of this area is showing a majority of ring features. This misclassification can be attributed to a weak candidate selection. For the lanes this is not exactly the case. Even manual inspection of the paths does not result in a convincing majority in stability features yielding a lane. This can mean that there is a definition mismatch between the ground truth and the method; it can also mean that the stability features are weak for the classification

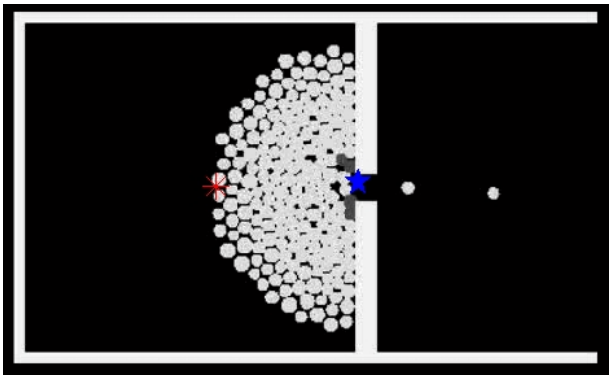

Figure 5: Results of the method with tuned parameters on a new dataset provided with courtesy of I. Farkas [22]. This dataset illustrates panic behaviour when exiting a room. The results are comparable to the *Stadium* dataset.

of lanes. Stronger lane features are proposed in the next section.

VI. Improvements

The defined behaviours in Figure 1 can be valuable to an observer, however other types of behaviour like stumbling could be too. The five behaviours can be used to define even more complex behaviours, this is discussed in Section VI-A. While doing experiments it became apparent that lane and ring behaviours are missed frequently by the method. This is mostly caused by the candidate selection. Improvements to detect lanes are proposed in Section VI-B. To make the method usable for application in the field a higher-level view of the results is preferred, this high level observer is described in Section VI-C. The last improvement on this work, Section VI-D is the visualization of some flow features besides the detected behaviours. This is an open discussion that can be used by professionals in the field of crowd management.

A. Behaviour Definition

The method described in Section III defines five different behaviours. With respect to public safety, the most interesting behaviours are blocking, bottleneck and fountainhead. The first indicates objects that block the flow (stumbling of people, blocked exits and so on), the second indicates high risk areas for crushing and trampling of people due to increased density and pressure and the latter can indicate panic or a run for emergency exits. Behaviours that can be relevant to detect are: panic, fights, emergency exit, stumbling/trampling, exit blockage. These behaviours might be defined by one or more of the behaviours proposed in Figure 1. See Table VI for some possible combinations of basic behaviour.

B. Lane Detection

Since lane detection is a potentially important behaviour to detect some alternative candidate selection features are proposed.

1) Trajectory Smoothness: There are two types of lanes; lanes that are forced by the environment and lanes that form spontaneously. The latter case would be of interest to detect. The spontaneous lane formation can best be explained by the least-effort hypothesis proposed in G.K. Still [15], in short: it takes less effort for an individual to follow a path. This work suggests some other parameters that could be responsible for lane formations, i.e. focal routes, routes defined by the field of view of the individual, and the dynamical (and maybe even social) attractiveness of a cue. It makes sense that the majority of lanes formed are rather smooth in principal. A step-shaped lane would probably not obey the least-effort hypothesis in general. It should be noted that not the instantaneous lane is considered but an equilibrium state of a lane. This smoothness can be exploited to detect paths evaluating particle trajectories resulting from the particle advection. Lanes should not only obey the smoothness constraint, but there should also be some parallel trajectories, since not every path of an individual should be labelled as a lane.

2) Hough Transform: It could be useful to detect straight lanes, or the straight parts of a lane in the trajectories of particles after advection. One way to implement this is using the Hough transform [23]. This technique projects points on a Hough space representing all possible lines through that point defined by the angle and the shortest distance from the line to a constant origin. This transformation, yielding a sine form in the Hough space for every point, representing all possible straight lines through that point, can be easily integrated in a voting framework. In the Hough space of two points there will be two sine forms present that differ in phase and amplitude. The intersection of these two sine waves represent lines passing through both points, since each intersection corresponds to a line passing through both points. There are two intersection since the 180 degrees rotated version is also a valid solution in the framework. When additional points are processed this method generates a density map where at high densities many intersections occur. These points correspond to strong lines in the initial point cloud. In the case of considering points of advection trajectories, this

Behaviour	Total	Correct	Missed	Misclassification
Bottleneck	10	7	3	4
Blocking	1	1	0	2
Fountainhead	2	2	0	5
Lane	2	0	2	0
Ring	2	0	2	0

Table V: The analysis of the result with respect to the total behaviours in the manually generated ground truth.

Behaviour	Pseudo Definition
Panic	Instantaneous fountainhead
Fighting	Blocking/fountainhead
Emergency exit	Bottlenecks/sudden lane forming
Stumbling/trampling	(Instantaneous) blocking
Exit blockage	Blocking/bottlenecks

Table VI: Some interesting behaviours and their possible representation in terms of behaviours defined by Figure 1.

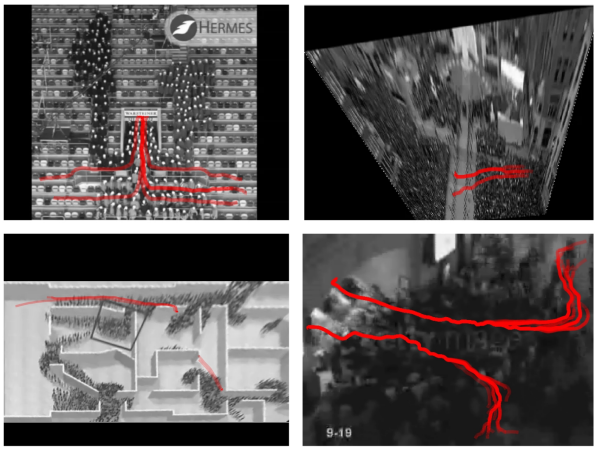

Figure 6: Results of the trajectory stability algorithm performed on four datasets, see Section VI-B3. The likelihood threshold used in this experiments varies between 0.3 and 0.5. The length threshold of a trajectory varies between 75 and 300 pixels, depending on scale and flow velocity.

method can output pieces of straight trajectories. Additionally, in the Hough space parallel lines are easily identified, since they appear as intersections on an imaginary vertical line. This information can be combined to find lanes.

3) Trajectory Stabililty: Let every trajectory, neglecting the ones which travelled insufficient distance, be a candidate lane. The stability theory as depicted in Section III-C can be applied to every pixel of these trajectories, treating the total amount of pixels of the path as the area assessed. Pixels corresponding to a trajectory should demonstrate a high score for the lane behaviour. This process is implemented and an initial experiment is performed. The results are shown in Figure 6. The likelihood is calculated by labelling every pixel of the trajectory according to Table II and III, with L in Table III the total amount of pixels in the trajectory. The resulting lanes of this algorithm are actually quite near where one would expect them.

4) Angular Segmentation: The method by B. Solmaz et al. is limited in the usage of angular information of the flow field, despite the fact that this can possibly contain some helpful details for lane detection. In lane forming it is evident that participants of the lane are heading the same way. By finding wide paths through assessment of the optical flow direction at every pixel lanes can be indicated. When two lanes are parallel, but opposite, this method can produce some sharp edges in the angular plot, resulting in clear boundaries.

C. High Level Observer

This concept involves an algorithm that uses the detected behaviours in a higher level. While the proposed method will process small pieces of video to retrieve behavioural data, a high level observer will continuously monitor these results, this way it is able to detect changes in both time and space. For example, the aforementioned sudden velocity change possibly indicating stumbling of a person is an application that should be implemented at this level. This approach also enables the prediction of behaviour especially when this high level observer has access to multiple simultaneous feeds evaluated by the method. This observer will have to detect relations between different features of the flow field. Testing an initial implementation of this observer requires some correlated datasets. The datasets processed in this work demonstrate at most two behaviours at a time, are not correlated and are thus not suited to test such an observer.

D. Data Representation

The different features of the flow field can be represented using different approaches. Since it is not possible to demonstrate the results to professionals working in the field of crowd management and surveillance an open discussion is presented.

It might be valuable for an observer to see the flow of the crowd represented by arrows, but others could prefer a colour scheme, where the colour is depended on the direction and the intensity of the colour on the velocity.

Most of the field of interest is monitored by numerous cameras. It would be valuable to fuse this data from multiple cameras to one flow field which is processed with the method. To facilitate this every camera needs to be calibrated; they have to provide a synchronized image stream and their real world position and pose should be known. An intuitive presentation of flows in big crowds is proposed in the form of a top view two dimensional map.

To show the detection of behaviour markers are used, each corresponding to a certain behaviour. This way it is easy to recognize the algorithm results. It is also possible to use transparent colours and overlay them on the video feed. The observer should be able to choose the behaviours he would like to see.

A feature that can give much insight in the crowd is the velocity of the flow. From the optical flow a velocity map is easily created and can be overlaid, with or without colour map, on the video feed or optical flow field. It might also be meaningful to calculate some contour lines of this velocity representing matching velocities in the crowd.

VII. Conclusions

he goal of this research is to develop a method usable for aiding a crowd observer. This is accomplished by evaluating a state of the art method in the field of crowd analysis. Methods handling dense crowds that are based on optical flow are interesting since computational cost does not increase drastically when crowd density is increasing. Furthermore, cluttering and occlusion are not an issue as this technique regards the crowd as an entity rather than a group of individuals. The method proposed and implemented in this work is inspired by Solmaz et al., because of the fact that no learning stage is required and the method enables specific behaviour identification. Behaviour definitions are derived from the stability features proposed by Solmaz et al., but the method proposed here contains more variations of behaviour patterns.

Experiments are performed to validate the implementation and asses performance. In addition some improvements are proposed to the candidate selection stage for finding lanes and rings.

The discussion about which features would be important to an observer is left open, since professionals in the field should attribute in this conversation. Velocity, direction and density of a crowd and especially variations of these properties in time are very interesting to an observer. The results obtained throughout this paper are very promising in the light of aiding an observer. Not all defined behaviours can be detected with the same confidence, but bottleneck and blocking, both obviously valuable to an observer, score very well. It should be noted that more data has to be processed to conclude the experiments. Data is scarce and the data found varies significantly, making it hard to compare the results. This work however has led to a framework that can be used to process new data efficiently.

A. Future Work

As mentioned before more data has to be acquired to make the data more reliable. The proposals made in Section VI are only initially tested, a thorough evaluation requires a detailed implementation accompanied by experiments. In most practical applications it is expected to have more than one camera on the terrain. It would be valuable to fuse this data and plot it on a two dimensional map of the field. Groundwork for a modular framework that allows this fusion is presented by the rectification step described earlier. The resulting flow maps can easily be stitched together to acquire a flow map of the whole location. To facilitate this implementation some sort of calibration of each camera should be performed. This calibration can also facilitate the tuning of location depended parameters. It should be noted that when rectifying cameras some areas of the video can be stretched out resulting in large flows caused by little noise. Therefore, it is recommended to accompany the rectification with a masking step discarding error prone regions. It is mentioned that stable camera feeds are preferred, however in practice there can be vibrations from the crowd, music or wind that make the camera drift. One can make the camera fixture more stable, but it is also possible to compensate these movements by post processing the optical flow. This also enables the use of moving cameras, which are not uncommon in field observation. As extra input of the optical flow compensation encoder readouts can be used. The edge effects caused by the advection of particles to the edge of the flow field can be detected and compensated for. It is also possible to replace the massless particles in the flow field with particles incorporating some inertia. This will make the particles move outside the flow field where they will not drift anymore due to zero flow. It is also interesting to evaluate the influence on the total advection since inertia enabled particles are a better model for people. A feature that is not used in the method, but can be important as feature for both the method and the observer, is the velocity of the flow field and in particular the change in velocity. Sudden changes in velocity can imply specific events; it can mean, for example, blocking by a stumbled person. But it can also simply give an overview of the lane velocity to the observer, assisting in, for example, deciding which exit to open or close. This method does not require a learning stage, however the flow features can be used to learn a classifier. This will be particularly interesting when deploying the method as part of a permanent system, like in a train station. Classifiers can be trained to predict characteristic crowd flows. In Section VI-B2 the Hough transform is suggested to assist in lane detection, there are variations on this methods proposed to find circles and arcs, which could be promising in detecting (part of) rings.

References

[1] Andrade E. L., Blunsden S. and Fisher R. B., "Modelling Crowd Scenes for Event Detection", IPAB, School of Informatics, University of Edinburg, IEEE ICPR'06.

[2] Solmaz B., Moore B., and Shah M., "Identifying Behaviors in Crowd Scenes Using Stability Analysis for Dynamical Systems", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2012.

[3] Rodriguez M., Sivic J., Laptev I. and Audibert J., "Data-driven Crowd Analysis in Videos",Proceedings of the IEEE International Conference on Computer Vision (2011), Oral.

[4] Wang S., Miao Z., "Anomaly Detection in Crowd Scene", Institute of Information Science, Beijing Jiaotong University, IEEE ICSP2010.

[5] Szczodrak M., Kotus J., Kopaczewski K., Lopatka K. and Czyzewski A., "Behavior analysis and dynamic crowd management in video surveillance system", Multimedia Systems Department, Gdansk University of Technology, Gdansk, Poland, IEEE 2011 22nd International Workshop on Database and Expert Systems Applications.

[6] Krawczyk H., Proficz J., "KASKADA - multimedia processing platform architecture", Proceedings of the International Conference on Signal Processing and Multimedia Applications, SIGMAP 2010 page 26-31.

[7] Zhan B., Monekosso D. N., Remagnino P., Velastin S. A. and Xu L., "Crowd analysis: a survey" Digital Imaging Research Centre, Kingston University, Kingston upon Thames, UK, Machine Vision and Applications 2008, Spinger-Verlag.

[8] Andrade E. L., Scott Blunsden and Robert B. Fisher, "Hidden Markov Models for Optical Flow Analysis in Crowds", IPAB, School of Informatics, University of Edinburg, IEEE ICPR'06.

[9] Liu W. and Ribeiro E., "Scale and Rotation Invariant Detection of Singular Patterns in Vector Flow Fields", Computer Vision and Bio-Inspired Computing Laboratory, Department of Computer Sciences, Florida Institute of Technology.

[10] Wang S., Miao Z., "Anomaly Detection in Crowd Scene Using Historical Information", Institute of Information Science, Beijing Jiaotong University, IEEE International Symposium on Intelligent Signal Processing and Communication Systems, December 6-8, 2010.

[11] Velastin S. A., Boghossian B. A. and Lazzarato A.,"Detection of potentially dangerous situations involving crowds using image processing", Department of Electronic Engineering, King's College London, Intelligent Industrial Automation June 1999.

[12] Narain R., Golas A., Curtis S., and Lin M. C., 2009. "Aggregate Dynamics for Dense Crowd Simulation", In ACM Transactions on Graphics (Proceedings of SIGGRAPH Asia), vol. 28, no. 5, pp. 122:1-122:8.

[13] Ali S. and Shah M., "A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis", IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[14] Horn B. K. P. and Schunck B. G., "Determing Optical Flow", Massachusetts Institue of Technology Artificial Intelligence Laboratory, A.I. Memo No. 572.

[15] Still G. K., "Crowd Dynamics", PhD Thesis, University of Warwick, August 2000.

[16] Beauchemin S.S. and Barron J.L., "The Computation of Optical Flow", ACM Computing Surveys, Vol 27. No. 3, September 1995.

[17] Barron J. L., Feet D.J.,Beauchemin S. S. and Burkitt T. A., "Performance of Optical Flow Techniques", 0-8186-2855-3/92 IEEE, 1992.

[18] Vitracom.de, "Staged Stadium Evacuation", Permission requested.

[19] NOS, Dutch news station, "Damschreeuwer", 2010.

[20] Berkeley University, ”Aggregate Dynamics for Dense Crowd Simulation”, Permission requested.

[21] UCF dataset, Saad Ali and Mubarak Shah, ”A Lagrangian Particle Dynamics Approach for Crowd Flow Segmentation and Stability Analysis”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), 2007.

[22] Helbing D., Farkas I., and Vicsek T., ”Simulating dynamical features of escape panic”, Nature 2000

[23] Herout A., Havel J. and Dubská M., ”Real-Time Detection of Lines and Grids”, SpringerBriefs in Computer Science.