

PEOPLE TRACKING AND TRAJECTORY INTERPRETATION IN AERIAL IMAGE SEQUENCES

F. Burkert^{a,*}, F. Schmidt^b, M. Butenuth^a, S. Hinz^b

^a Technische Universität München, Remote Sensing Technology, 80333 München, Germany
(florian.burkert, matthias.butenuth)[@]bv.tum.de

^b Karlsruher Institut für Technologie, Institut für Photogrammetrie und Fernerkundung, 76131 Karlsruhe, Germany
(florian.schmidt, stefan.hinz)[@]kit.edu

Commission III, WG III/5

KEY WORDS: People tracking, people trajectories, event detection, aerial image sequences

ABSTRACT:

Monitoring the behavior of people in complex environments has gained much attention over the past years. Most of the current approaches rely on video cameras mounted on buildings or pylons and people are detected and tracked in these video streams. The presented approach is intended to complement this work. The monitoring of people is based on aerial image sequences derived with camera systems mounted on aircrafts, helicopters or airships. This imagery is characterized by a very large coverage providing the opportunity to analyze the distribution of people over a large field of view. The approach shows first results on automatic detection and tracking of people from image sequences. In addition, the derived trajectories of the people are automatically interpreted to reason about the behavior and to detect exceptional events.

1. INTRODUCTION

Monitoring the behavior of people in crowded scenes and in complex environments has gained much attention over the past years. The increasing number of big events like concerts, festivals, sport events and religious meetings as the pope's visit leads to a growing interest in monitoring crowded areas. In this paper, a new approach for detecting and tracking people from aerial image sequences is presented. In addition to delineating motion trajectories, the behavior of the people is interpreted to detect exceptional events such as panic situations or brawls.

A typical feature of current approaches is the utilization of video cameras mounted on buildings to detect and track people in video streams. Pioneering work on tracking human individuals in terrestrial image sequences can be found, e.g., in (Rohr, 1994; Moeslund & Granum, 2001). While this work focuses on motion capture of an isolated human, first attempts to analyze more crowded scenes are described in (Rosales & Scarloff, 1999; McKenna et al. 2000). Such relatively early tracking systems have been extended by approaches integrating the interaction of 3D geometry, 3D trajectories or even intentional behavior between individuals (Zhao & Nevatia, 2004; Yu & Wu, 2004; Nillius et al., 2006; Zhao et al., 2008). Advanced approaches, based on so-called sensor networks, are able to hand-over tracked objects to adjacent cameras in case they leave the current field of view achieving a quite comprehensive analysis on the monitored scene. The work of (Kang et al., 2003) exemplifies this kind of approaches. Instead of networks of cameras, moving platforms like unmanned airborne vehicles (UAVs) can be utilized, too, as e.g. presented in (Davis et al., 2000). An overview on the research of crowd modeling and analysis including all stages of a visual surveillance is given in (Hu et al., 2004; Zhan et al., 2008).

An important aspect of tracking a large number of people, as e.g. shown in (Rodriguez et al., 2009), includes the potential to not only analyze individual trajectories but also to learn typical interactions between trajectories (Scovanner & Tappen, 2009).

Hence, event detection has been an intensely investigated field of research in the last decade. A framework using two modular blocks to detect and analyze events in airborne video streams is presented in the work of (Medioni et al., 2001). The first module detects and tracks moving objects in a video stream, whereas the second module employs the derived trajectories to recognize predefined scenarios. A further event recognition system is based on two consecutive modules, namely a tracking and an event analysis step, in which complex events are recognized using Bayesian and logical methods (Hongeng et al., 2004). Video streams from close range surveillance cameras are used to detect events focusing on interactions between few persons. Further methods exemplify the emphasis on research in surveillance issues, as the scanning of video streams for unusual events (Breitenstein et al., 2009; Mehran et al., 2009). Additional related work in the field of people tracking and event detection is based on seminal research in crowd analysis and simulation (Helbing and Molnar, 1995; Helbing et al., 2002). Observed collective phenomena in moving crowds, like lane formations in corridors, have successfully been simulated using a social force model (SFM). The SFM considers interactions among pedestrians and between pedestrians and obstacles, resulting in a certain moving direction for each individual.

The approach presented in this paper is aimed to complement the above work. The monitoring of people is based on aerial camera systems mounted on aircrafts, UAVs, helicopters or airships. The provided image sequences cover a large area of view allowing for the analysis of density, distribution and motion behavior of people. Yet, as the frame rate of such image sequences is usually much lower compared to video streams (only some Hz), more sophisticated tracking approaches need to be employed. Moreover, the interpretation of scenarios in such large scale image sequences needs to comprise an exceeding number of moving objects compared to existing event detection systems. Thus, the intention of the approach is to define a broader spectrum of identifiable scenarios instead of simply alerting a general abnormal event within a monitored area.

The paper is organized as follows: Section 2 sketches the concept underlying the new approach. Section 3 outlines algorithmic aspects and shows results of people tracking and trajectory interpretation including a performance evaluation, exemplified by a test scenario of a crowded entrance area of a soccer stadium. Section 4 gives concluding remarks and discusses possible future investigations.

2. APPROACH OF PEOPLE TRACKING AND TRAJECTORY INTERPRETATION

2.1 System

The underlying concept of the presented approach can be logically separated into two parts, as e.g. also (Medioni et al., 2001; Hongeng et al., 2004) do. The first part detects people and delineates their trajectories from a sequence of ortho-images and external geospatial information (Section 2.2). Additionally, macroscopic parameters describing typical patterns of a crowd scene like density, activity and systematic motion are calculated. The second part utilizes the obtained trajectories, the macroscopic parameters and model knowledge about the scene to analyze the people's behavior and to detect a certain instance of predefined scenarios (Section 2.3). The general overview of the system is shown in Figure 1.

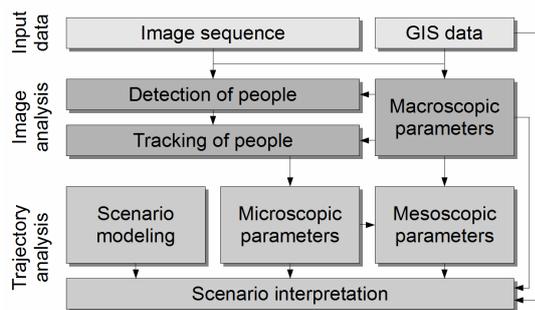


Figure 1. System overview.

2.2 Tracking of people in aerial image sequences

The detection and tracking scheme involves three different steps. First, macroscopic information about the imaged scene is determined. This is followed by detecting single people in each image and, finally, these detections are tracked by linking them iteratively over three images at the same time.

Often, geospatial information can be expected providing additional data about the application scene in advance. The combination with the georeferenced ortho-images reduces the region of interest in every image of the sequence. Further preprocessing is done by decreasing the number of image channels. For efficiency reasons, currently only the first principal component of every color image is utilized, although the additional channels might contain some supplementary information. Future investigations will be carried out to evaluate the benefit of using color data.

Estimation of macroscopic parameters: As shown in (Hinze, 2009), aerial image sequences can be used to estimate density, activity and motion of people in crowded environments. These parameters give coarse information about the behavior of people at a macroscopic level. They can also be used to improve detection and tracking of individuals and for trajectory analysis. The image analysis begins with the computation of a

general density map for each image. A simple region-growing algorithm is applied to detect large homogeneous regions which belong to the background. To exclude also small buildings and other man-made structures from the foreground, edge directions are calculated and analysed. The remaining foreground pixels are further denoised morphologically with a circular structure element.

The next processing step comprises dot detection in the foreground region, which is done by convolving and thresholding the image with a filter enhancing dark and light circular dots. The residual foreground pixels are likely to belong to a person or dot-like clutter. Afterwards the local object density is calculated by applying a medium filter to the resulting binary image. Finally the density image is divided into three regions of high, medium and low object density. Although this measure might still include some clutter, it is nonetheless helpful in further processing and to judge the results of detection and tracking correctly.

Detection of individuals: At a resolution of few decimeters per pixel a single person is hardly visible. The body of one person covers a region of about 10 to 20 pixels and has the shape of a near circular dot. The proposed detection system is a cascade of image processing algorithms that is designed to find the circular dots in every image. In contrast to many other object detection problems, no search has to be accomplished through the scale space because the size of the objects is known and stays fixed over time.

The detection algorithm utilises the results of the previous processing step as input. A single person points out as a local maxima in the dot-filtered image. Hence a pouring-based segmentation algorithm is applied to find these maxima and to generate segments around them. These segments are further classified into clutter and possible individuals by examine shape features like area, convexity and compactness.

Tracking of people: Tracking of hundreds or thousands of people in a crowded environment from an airborne platform is a challenging task. If the density of people becomes too high, it is hardly possible to visually differentiate individuals. Additionally, the spectral signature of a single person in an aerial image has not many features to discriminate between neighboring persons. The appearance features can change significantly in a short period of time due to the movement of the airborne platform and the varying influence of shadows, clouds, or neighboring persons. For these reasons, a semi-global optical-flow tracking algorithm is employed which balances local features with global smoothness constraints to link detected regions in consecutive images.

Input data for the tracking algorithm are the segments from the previous detection phase. These segments are fed into an optical flow algorithm to compute the displacement vectors between two consecutive images. Typical maximum motion parameters of humans are involved to reduce the search space for flow calculation. If there is a segment in the second image at the predicted position and the estimated movement is realistic regarding the local object density, both segments are linked to represent the motion of one person. Because the results of the optical flow computation are affected not only by object motion but also changing lighting conditions, they can only serve as an approximate cue for the motion of a person through the sequence. Therefore, the same procedure is repeated in the

opposite direction and only those links that appear in both ways are kept.

The global smoothness constraint of optical flow allows to link object regions without an explicit matching of their unstable appearance. However the drawback of the proposed method is its dependency on a good and complete object detection result in each picture. To overcome situations when a single person could not be detected in one image of the sequence or when a link could not be established, images are processed additionally being two frames apart. These links are used to establish missing connections between the three consecutive images while the person’s location in the bridged image is interpolated.

The introduced procedure is applied to the entire sequence. The output of the tracking algorithm consists of trajectories which reflect the motion of individuals through the image sequence. They are used for further processing in the second module of the proposed system.

2.3 Interpretation of trajectories of people

The derived trajectories of moving people within the observed scene are used to initialize the second module analyzing the trajectories with regard to motion patterns. The trajectory interpretation system aims at bridging the gap between low level representation of single trajectories and high level people behavior analysis from image sequences. To achieve this goal, microscopic motion parameters of single trajectories as well as mesoscopic motion parameters of several trajectories have to be extracted. A graph is constructed containing microscopic and mesoscopic motion parameters to represent a neighborhood of trajectories. Additionally, GIS data and macroscopic parameters are utilized to recognize predefined scenarios. A hierarchical modeling of scenarios is feasible, as the interpretation of trajectories is based on the analysis of simple motion parameters of one or more trajectories. In the following, the module for trajectory analysis is presented in more detail.

Microscopic and mesoscopic parameters: Microscopic motion parameters concern the motion characteristics of one single moving person. Hence, the most important microscopic motion parameters to exploit are speed and motion direction. In addition, further parameters can be calculated from these two basic microscopic motion parameters. Figure 2 shows a single trajectory depicting some features which are used to calculate the following parameters.

The average speed v of a moving object is calculated using the relative distance d_{rel} of a trajectory which is given as the Euclidian distance between the points x_1 and x_n . Using this approach, v is the speed for the effectively covered distance for this object within the observed time frame, disregarding any multi-directional movements. In contrast, the absolute distance d_{abs} is derived from adding the segments d_i of one trajectory over all time steps i . The acceleration a of a moving object is computed by differencing the speeds of two consecutive line segments. A further microscopic parameter is straightness, calculated from the two different distances mentioned before by $s = d_{rel}/d_{abs}$. As d_{abs} always receives a bigger number than d_{rel} , s takes a value near 1 when the trajectory is very straight and a much smaller value towards 0 when the trajectory is very twisting or even self-overlapping.

Motion direction is the second basic microscopic motion parameter: the direction $z(x_i)$ at a point x_i is the direction of

the tangent at this point defined by the points $x_{(i-1)}$ and $x_{(i+1)}$. The motion direction is specified counterclockwise with reference to a horizontal line. Similar to straightness, the standard deviation σ_z of the motion directions indicates the degree of the twists and turnarounds within one trajectory.

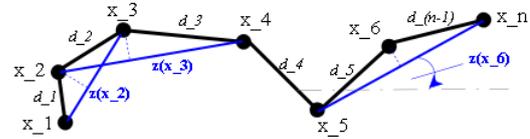


Figure 2. Features of a trajectory to calculate microscopic motion parameters: points x_i and line segments d_i (black), direction at point with reference to horizontal line $z(x_i)$ (blue).

Mesoscopic motion parameters represent the interaction between several individuals. Therefore, it is necessary to evaluate the proximity of a trajectory with respect to the number of neighboring trajectories, their motion directions and potential interferences. Figure 3 shows an example of two neighboring trajectories. The detection of neighbors is accomplished by scanning the surrounding area of existing trajectory points at every time step i . For each detected neighbor, the offset o_i of each pair of points x_i and y_i is stored. Comparing length and direction of these offsets during the entire image sequence, robust information can be derived if neighbors come closer or even touch each other. In addition, the motion direction at each point is inspected to detect intersections of trajectories.

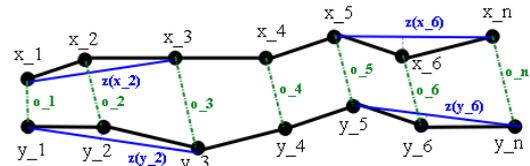


Figure 3. Two neighboring trajectories with offsets o_i (green) between pairs of points x_i and y_i (black).

Scenario modeling and scenario recognition: Scenarios are modeled hierarchically to recognize complex motion patterns based on the extraction of simple microscopic and mesoscopic motion parameters, similar to the event detection systems mentioned in Section 1. Hence, predefined scenarios consist of trajectories and local GIS information in the lower level which represent simple image features by coordinates (Figure 4). Microscopic motion parameters follow in the next level of motion parameters which give a more abstract representation of the trajectories. Additionally, mesoscopic motion parameters are embedded in this level because they are closely linked to microscopic motion parameters and directly derived from the trajectories. In the subsequent level, simple events are modeled resulting from beforehand defined parameters. These events concern single trajectories or try to model information from mesoscopic motion parameters. In the highest level of the hierarchical scenario modeling, simple events are combined with GIS data to complex scenarios representing complex motion patterns within the observed scene.

The goal of the proposed system is to recognize scenarios which are predefined as described before. Based on the tracking in the first module of the system, motion parameters are extracted. These parameters are evaluated to compute probabilities of simple occurring events. The combination of several simple events leads to the recognition of a predefined scenario.

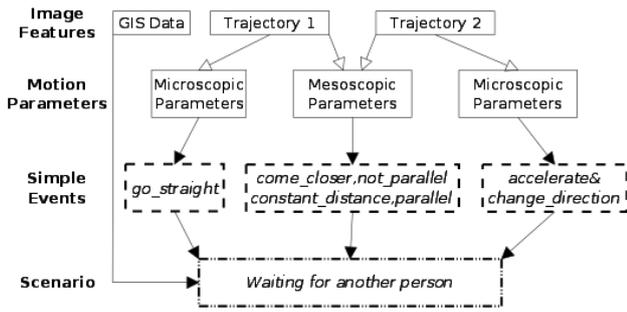


Figure 4. Example for the scenario “waiting for another person” consisting of four hierarchical layers.

3. EXPERIMENTAL RESULTS

3.1 Test scenario

For developing and testing the presented new approach, aerial image sequences provided by DLR’s 3K multi-head camera system are used (Kurz et al., 2007). This system consists of three non-metric off-the-shelf cameras, with one camera pointing in nadir direction and two in oblique direction. The basis for near-realtime mapping is provided with a coupled realtime GPS/IMU navigation system which enables accurate direct georeferencing.

The aerial image sequence used in the experiments was captured at a soccer match with a few thousand people heading for the gates of the stadium. The height of flight was 1500m resulting in a ground sampling distance of about 20cm. In spite of the low resolution, people can be recognized clearly by their long shadow. The camera system has been operating in continuous mode which resulted in image sequences with a length of 40 frames at a sampling rate of 2 Hz. Every image covers an area of approximately 1000m × 600m and with an in-track overlap of about 90%. For the evaluation a smaller area has been selected, completely visible in 16 consecutive frames. Figure 6 shows the test area in every third frame of the image sequence.

3.2 Detection and tracking of people

The detection and tracking algorithms are evaluated comparing the achieved results with reference trajectories, collected manually using the same image sequence. This reference data is not complete in regions where the density of people is too high to differentiate between individuals. Therefore, the evaluation is limited to a region of about 40m × 50m south of the entrance



Figure 6. Test area in the aerial image sequence used for evaluation of the tracking and interpretation system; frames no. 1, 4, 7, 10, 13 and 16 out of 16 consecutive images are shown [3K-images are provided by the German Aerospace Center (DLR)].

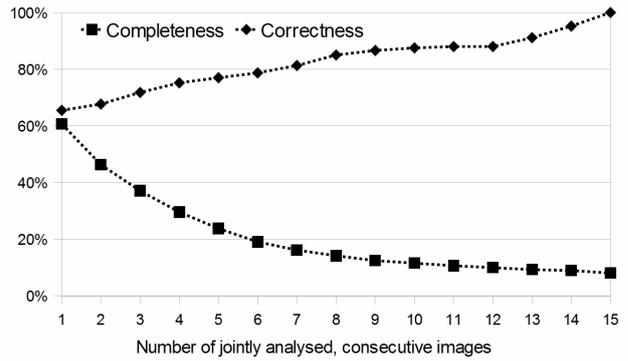


Figure 5. Comparison of manually tracked persons with the results of the algorithm over a sequence of 15 aerial images with about 130 persons visible.

which is not too crowded. Here, 130 persons could be marked manually in average through a sequence of 15 frames. It is important to know for a correct interpretation of the evaluation that the reference data might not be free of errors. Occasionally, manually tracked persons merged with others so that their position had to be estimated for some frames. In other situations, the contrast became too low to define the accurate position of a person due to clouds passing by.

The evaluation results of the detection and tracking algorithms are shown in Figure 5. An automatically generated segment is considered as a correct detection if the distance between its center and the next reference position is within a tolerance radius of 3 pixels corresponding to 45cm on the ground. The same criterion is applied to evaluate the tracking results. Though, in this case every point of a generated trajectory has to be close enough to one of the reference trajectories. For the evaluation of the tracking results all possible links between two up to 15 consecutive frames are compared. Figure 7 visualizes a result of detection and tracking in comparison to the reference.

Averaging the results over all 15 images, the detection module has achieved a completeness of 61% and a correctness of 66% (cf. Figure 5, length 1). The completeness of the generated trajectories increases almost linearly with growing length while the completeness drops down quickly. Several reasons are possible: one effect still to investigate is the influence of the tolerance radius during evaluation. The center of the detected segments could be more than 3 pixels away from the manually marked position of the head of a person. This can happen when the body of one person merges with its shadow to a uniform dot due to low contrast, cf. Figure 7 (left). Another effect stems

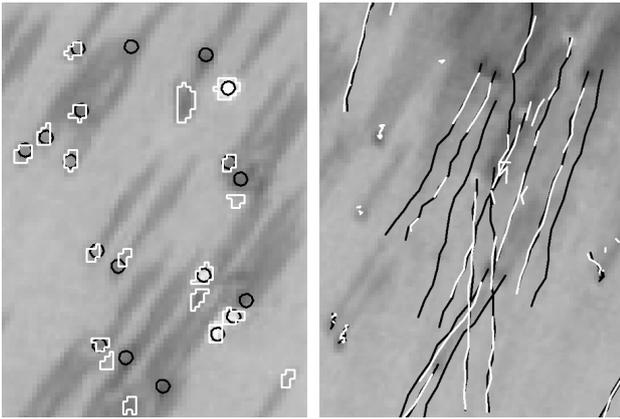


Figure 7. Detection results (left): black circles show the position of manually marked persons, white regions were generated by the algorithm; tracking result (right): reference trajectories are black, automatically generated trajectories are white.

from the optical flow which can occasionally lead to a wrong displacement vector. Although the limitations of the reference data and the low contrast of the image sequence corrupt the results to a certain amount, the numerical evaluation shows clearly the potential for further improvements of the detection and tracking system. A more sophisticated detection algorithm based on a machine learning approach will certainly reduce the effect of clutter and improve the results significantly.

3.3 Interpretation of trajectories

Experimental results for the computation of microscopic motion parameters in the trajectory analysis module are presented in this section. The computation is performed on manual reference data used in the image analysis module to focus at this point to the new approach of trajectory interpretation. Two example scenes of the image sequence are used for event detection, both with eight trajectories visualized on the last image of the sequence, respectively (Figure 8, Figure 9).

Figure 8 shows a scene nearby the queue. Table 1 lists the related microscopic parameters of the eight trajectories. The microscopic parameters are meaningful with regard to the characterization of the trajectories. Trajectories 1 and 2 turn out to cover the longest distance and to have the smoothest and the least twisting path. This facts are clearly depicted by the parameters d_{rel} and d_{abs} , which result in a high value for $s \approx 0,98$, as well as by the small deviation $\sigma_z \approx 10^\circ$. Caused by the turn, trajectories 3 and 4 receive a smaller $s \approx 0,75$ and a higher deviation $\sigma_z \approx 50^\circ$. The shorter trajectories 7 and 8 are characterized by small values for s as the paths are very twisting. Modeling the motion patterns of trajectories 1-4 result in a complex event called “waiting for other person”, because the persons 3 and 4 obviously accelerate their motion when persons 1 and 2 are next to them, afterwards walking alongside each other. Additionally to the shown microscopic parameters, evidence for parallelism can be given by mesoscopic parameters and, thus, result in the complex event similar described in Figure 4. Parallelism is visualized by the chronologically colored crosses of frames no. 1, 4, 7, 10, 13 and 16 in Figure 8.

Figure 9 shows another more crowded scene 2 representing a different event. Table 2 lists the microscopic parameters of eight trajectories. This scene is located at the right boundary of the queue, depicted by a macroscopic border of high density

(red), next to a wall derived from GIS data (blue), cf. Figure 9. Resulting from the trajectory’s characteristics, microscopic motion parameters again receive values as expected. In this scene, the motion pattern shows a possibly dangerous “bottleneck” event, because the faster walking persons are pushed aside by the queue. These faster persons have to sidestep to a small gap between the queue and the wall.

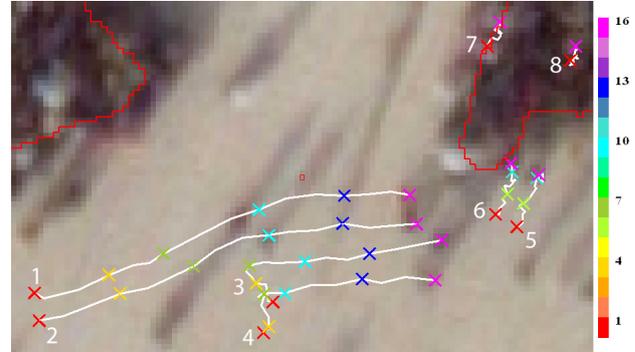


Figure 8. Results of the trajectory analysis module of scene 1: high density borders depicted in red; the color bar shows the time steps within the sequence.

ID	d_{abs} (m)	d_{rel} (m)	v (km/h)	s	z_{mean} ($^\circ$)	σ_z ($^\circ$)
1	11,22	10,91	5,24	0,972	13,27	10,95
2	11,17	10,94	5,25	0,980	13,95	9,97
3	7,05	5,05	2,42	0,715	51,30	57,31
4	6,55	5,05	2,42	0,771	42,88	44,67
5	2,50	1,60	0,76	0,641	29,66	92,03
6	2,14	1,52	0,72	0,710	67,99	50,63
7	1,71	0,78	0,37	0,454	56,72	35,34
8	1,67	0,43	0,21	0,257	31,47	84,53

Table 1. Microscopic parameters for trajectories in Figure 8. Results of the trajectory analysis module of scene 1.

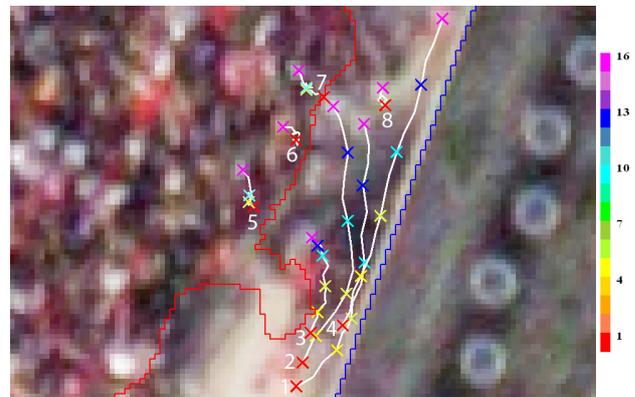


Figure 9. Results of the trajectory analysis module of scene 2: high density border depicted in red, GIS data depicted in blue; the color bar shows the time steps within the sequence.

ID	d_{abs} (m)	d_{rel} (m)	v (km/h)	s	z_{mean} ($^\circ$)	σ_z ($^\circ$)
1	8,73	8,17	3,92	0,936	72,99	18,31
2	8,27	7,78	3,74	0,941	81,11	20,75
3	3,29	2,86	1,38	0,871	82,03	60,34
4	9,77	9,72	4,66	0,994	72,19	3,84
5	2,31	1,15	0,55	0,500	67,68	85,22
6	1,78	0,56	0,27	0,315	34,84	112,24
7	2,01	1,10	0,52	0,549	89,24	86,78
8	1,66	0,54	0,26	0,324	36,12	102,24

Table 2. Microscopic parameters for trajectories in Figure 9.

4. CONCLUSIONS

The presented new approach for detecting and tracking people from aerial image sequences shows very promising first results. In addition, the achievements interpreting the trajectories demonstrate the potential of event detection. Several further developments and investigations are of interest: Haar-like features and AdaBoost classification (Smal et al., 2010) is planned to be used in the future to improve the object detection component. Besides detection also tracking can be improved: although the algorithm can handle situations of a person being missed in a single frame, it fails completely when it happens in two or more consecutive frames. This drawback cannot be dissolved with the proposed optical-flow algorithm. Bridging more than one image would allow to construct longer trajectories, whose completeness increases significantly as the currently derived results. The trajectory interpretation module is exemplarily shown by two different events: obviously, the modeling of further scenarios is aimed to get a more overall monitoring of possible occurring events. The automatic detection of predefined events using statistical methods, similar to (Hongeng et al., 2004), is intended to be accomplished in the near future. In addition, a backward-loop is strived to be integrated in the system: results derived from the interpretation of the trajectories could be integrated in the strategies to improve the tracking model. Obviously, the dependent interpretation module will benefit afterwards from more reliable tracking results.

REFERENCES

- Breitenstein, M., Grabner, H., van Gool, L., 2009. Hunting Nessie – Real-Time Abnormality Detection from Webcams. *Proceedings International Conference on Computer Vision*, pp. 1243-1250.
- Davis, L., Philomin, C., Duraiswami, R., 2000. Tracking Humans from a Moving Platform. *Proceedings International Conference on Pattern Recognition*, pp. 171-178.
- Helbing, D., Farkas, I.J., Molnar, P., Vicsek, T., 2002. Simulation of Pedestrian Crowds in Normal and Evacuation Situations. In: Schreckenberg, M., Sharma, S. D. (eds.), *Pedestrian and Evacuation Dynamics*, Springer, pp. 21-58.
- Helbing, D., Molnar, P., 1995. Social Force Model for Pedestrian Dynamics. *Physical Review E* 51(5), pp. 4282-4286.
- Hinz, S., 2009. Density and Motion Estimation of People in Crowded Environments Based on Aerial Image Sequences. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(1-4-7/W5), on CD.
- Hongeng, S., Nevatia, R., Bremond, F., 2004. Video-based Event Recognition: Activity Representation and Probabilistic Recognition Methods. *Computer Vision and Image Understanding* 96(2), pp. 129-162.
- Hu, W., Tan, T., Wang, L., Maybank, S., 2004. A Survey on Visual Surveillance of Object Motion and Behaviors. *IEEE Transactions on Systems, Man and Cybernetics* 34(3), pp. 334-352.
- Kang, J., Cohen, I., Medioni, G., 2003. Continuous Tracking Within and Across Camera Streams. *Proceedings Conference on Computer Vision and Pattern Recognition*, pp. 267-272.
- Kurz, F., Müller, R., Stephani, M., Reinartz, P., Schröder, M., 2007. Calibration of a Wide-angle Digital Camera System for Near Real-time Scenarios. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 36(1/W51), on CD.
- McKenna, S.J., Jabri, S., Duric, Z., Rosenfeld, A., Wechsler, H., 2000. Tracking Groups of People. *Computer Vision and Image Understanding* 80(1), pp. 42-56.
- Medioni, G., Cohen, I., Bremond, F., Hongeng, S., Nevatia, R., 2001. Event Detection and Analysis from Video Streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(8), pp.873-889.
- Mehran, R., Oyama, A., Shah, M., 2009. Abnormal Crowd Behavior Detection Using Social Force Model. *Proceedings International Conference on Computer Vision and Pattern Recognition*, pp. 935-942.
- Moeslund, T.B., Granum, E., 2001. A Survey of Computer Vision-based Human Motion Capture. *Computer Vision and Image Understanding* 81, pp. 231-268.
- Nillius, P., Sullivan, J., Carlsson, S., 2006. Multi-Target Tracking - Linking Identities Using Bayesian Network Inference. *Proceedings Conference on Computer Vision and Pattern Recognition*, pp. 2187-2194.
- Rodriguez, M., Ali, S., Kanade, T., 2009. Tracking in Unstructured Crowded Scenes. *Proceedings International Conference on Computer Vision*, pp. 1389-1396.
- Rohr, K. 1994. Towards Model-Based Recognition of Human Movements in Image Sequences. *CVGIP: Image Understanding* 59(1), 94-115.
- Rosales, R., Sclaroff, S., 1999. 3D Trajectory Recovery for Tracking Multiple Objects and Trajectory Guided Recognition of Actions. *Proceedings Conference Computer Vision and Pattern Recognition*, pp. 117-123.
- Scovanner, P., Tappen, M., 2009. Learning Pedestrian Dynamics from the Real World. *Proceedings International Conference on Computer Vision*, pp. 381-388.
- Smal, I., Loog, M., Niessen, W., Meijering, E., 2010. Quantitative Comparison of Spot Detection Methods in Fluorescence Microscopy. *IEEE Transactions on Medical Imaging* 29(2), pp. 282-301.
- Yu, T., Wu, Y., 2004. Collaborative Tracking of Multiple Targets. *Proceedings Conference Computer Vision and Pattern Recognition*, pp. 834-841.
- Zhan, B., Monekosso, D. N., Remagnino, P., Velastin S. A., Xu L., 2008. Crowd Analysis: A Survey. *Machine Vision and Applications* 19(5-6), pp. 345-357.
- Zhao, T., Nevatia, R., 2004. Tracking Multiple Humans in Complex Situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26(9), pp. 1208-1221.
- Zhao, T., Nevatia, R., Wu, N., 2008. Segmentation and Tracking of Multiple Humans in Crowded Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(7), pp. 1198-1211.