JOURNAL OF SPATIAL INFORMATION SCIENCE
Number 2 (2011), pp. 87–117
doi:10.5311/JOSIS.2010.2.1

RESEARCH ARTICLE

Uncertainty-aware video visual analytics of tracked moving objects

Markus Höferlin[1], Benjamin Höferlin[2], Daniel Weiskopf[1], and Gunther Heidemann[2]

[1]VISUS, Universität Stuttgart, 70569 Stuttgart, Germany
[2]Intelligent Systems Group, Universität Stuttgart, 70174 Stuttgart, Germany

Received: March 31, 2010; returned: June 10, 2010; revised: July 22, 2010; accepted: September 1, 2010.

Abstract: Vast amounts of video data render manual video analysis useless while recent automatic video analytics techniques suffer from insufficient performance. To alleviate these issues, we present a scalable and reliable approach exploiting the visual analytics methodology. This involves the user in the iterative process of exploration, hypotheses generation, and their verification. Scalability is achieved by interactive filter definitions on trajectory features extracted by the automatic computer vision stage. We establish the interface between user and machine adopting the VideoPerpetuoGram (VPG) for visualization and enable users to provide filter-based relevance feedback. Additionally, users are supported in deriving hypotheses by context-sensitive statistical graphics. To allow for reliable decision making, we gather uncertainties introduced by the computer vision step, communicate these information to users through uncertainty visualization, and grant fuzzy hypothesis formulation to interact with the machine. Finally, we demonstrate the effectiveness of our approach by the video analysis mini challenge which was part of the IEEE Symposium on Visual Analytics Science and Technology 2009.

Keywords: visual analytics, video analysis, uncertainty, trajectories, interactive query, video processing, video visualization, video surveillance

1 Introduction

The amount of video data world-wide is growing tremendously and has already reached hardly manageable dimensions. The following examples help to understand the actual situation:

- According to Cisco's estimates on future Internet traffic, "the sum of all forms of video (TV, video on demand, Internet, and P2P) will account for over 91 percent of global consumer traffic by 2013."[1]
- In May 2009, the duration of videos uploaded to YouTube in every minute exceeded 20 hours.[2]
- The New York Times of January 11, 2010 notices that the US "Government agencies are still having trouble making sense of the flood of data they collect for intelligence purposes. ... Air Force drones collected nearly three times as much video over Afghanistan and Iraq last year as in 2007—about 24 years' worth if watched continuously."[3]
- With an estimated 40 million surveillance cameras worldwide [45], closed-circuit television (CCTV) is one of the major sources of video data.

For reasonable interpretation of video data, especially in order to draw valuable insights from it, comprehensive and reliable analyses are necessary. In consideration of the vast amount of video data, analysis methods should be scalable and efficient. In general, there is a trade-off between quality and efficiency, but in particular, there are a few problems that are well-suited to automated video analysis. Some examples for applications that are suited to automated video analysis are detection of shot changes in movies, or number plate recognition for traffic surveillance.

However, there are many problems that are unsatisfactorily handled by automated video processing. Imagine the challenge of searching for suspicious events within several hours of video data captured by a surveillance camera. While the task of detecting a specific person or a previously known event can be handled by an automated analysis process, the search for vaguely defined targets becomes unreliable (see Figure 1). We follow Leyk et al.'s definition of vagueness, which can be "defined as indeterminacy due to a lack of distinctness between ill-defined or fuzzy classes of objects" [37]. A search for such a vaguely defined target (which itself can consist of several objects) is an ill-posed problem, where we assume a smooth transition between well-defined and vaguely defined targets. In general, we observe that the reliability of automated video analysis depends on the complexity of the problem, which in turn can be expressed by i) the degrees of freedom inherent in the problem's definition; and ii) the degrees of freedom present in the problem's context or environment (e.g., projection, illumination, noise). The former (i) also depends on the semantic level of the problem (in other words the amount of abstraction involved), and is solely responsible for the classification if a problem is vaguely or well-defined. Automated video analytics systems that try to cope with complex problems often suffer from high false alarm rates [26]; there is a trade-off between recall and precision.

A non-exhaustive enumeration of domains that have to deal with such ill-posed search targets are:

- visual surveillance (CCTV, surgery documentation, aerial surveillance);

[1] "Cisco visual networking index: Forecast and methodology, 2008–2013," available at http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html

[2] "Zoinks! 20 hours of video uploaded every minute!", available at http://youtube-global.blogspot.com/2009/05/zoinks-20-hours-of-video-uploaded-every_20.html

[3] "Military is awash in data from drones," permalink: http://www.nytimes.com/2010/01/11/business/11drone.html

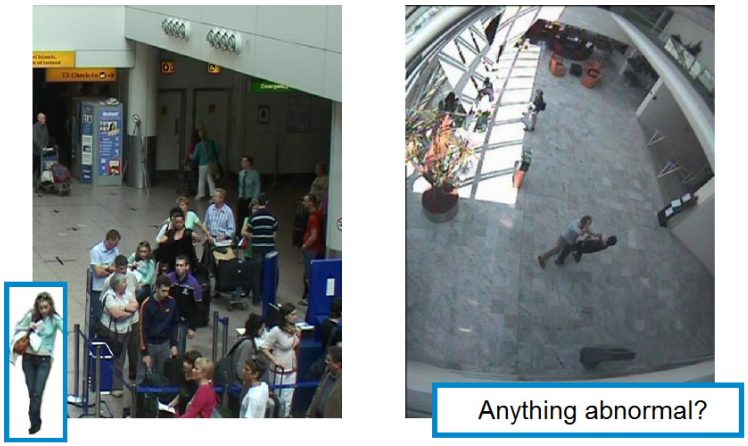

Figure 1: Example of well-defined search target (specific person in the left image) and vaguely defined target (search for "abnormal" behavior in right image). The images originate from PETS 2004 dataset [20] and PETS 2006 dataset [57].

- scientific video analysis (behavior analysis, biological surveillance/animal studies, motion analysis); and
- digital life streams (webcams, videoblogs).

Since automated video analysis in these fields is usually difficult and unreliable, users often fall back on the traditional method: manual video analysis (i.e., watching the complete video), which obviously lacks scalability. The situation in video analysis can be summarized by the words of John Naisbitt:

> "We are drowning in information but starved for knowledge." [44]

1.1 A visual analytics approach

To facilitate scalable video analysis with respect to vaguely defined search targets we introduce an approach based on the visual analytics (VA) methodology. The term visual analytics describes "an iterative process that involves information gathering, data preprocessing, knowledge representation, interaction, and decision making" [30]. Hence, the basic idea is to split the analysis task into two parts: automated low-level feature extraction and human-based high-level pattern recognition (see Figure 2 for a simplified VA process). Thus, we deploy both parts to the expert, since humans are adept at recognizing complex patterns and spatiotemporal events, whereas the machine is capable of performing large amounts of calculations reliably and in a short period of time. The connection between both worlds is established by visualization and human-computer interaction.

The goal of video analysis is to gain insight into the data provided by the capturing device. According to VA, insight is gained by a user-driven iterative process of exploration, hypothesis formulation, and hypothesis verification. The users are supported in this task by features automatically extracted from the input videos. We use trajectory features since they are suitable in many cases for visual surveillance and scientific video analysis. The challenge of visualization and interaction, however, is to assist the users in deriving these

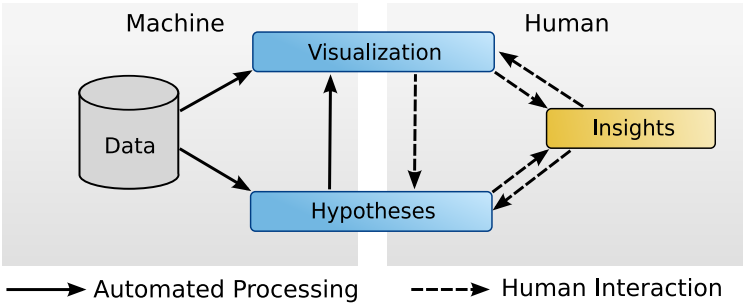

Figure 2: Simplified visual analytics process. The interfaces between machine and human (i.e., visualization and human-computer interaction) are fundamental parts to combine automatic low-level feature extraction and high-level pattern recognition by the user.

hypotheses, assumptions, and finally, insights on possibly relevant events by means of the video data and the provided features. The advantage of video visual analytics is that scalability is achieved by exploiting automatic feature extraction, while at the same time reliability is obtained by utilizing human experts.

The information retrieved by video analysis, especially in the case of CCTV, often provides the basis for more or less critical decisions. Thus, it is crucial for decision-making to be aware of and to consider uncertainties originating from different levels in order to draw the right conclusions. Uncertainty information has to be extracted and communicated to users in a comprehensible manner. International standards define how to aggregate, propagate, classify, and display uncertainty in measurement [56].

In this paper, we distinguish between uncertainties from different sources. We denote the uncertainty stemming from automatic analysis as *data quality*. This is the quality of trajectory data extracted from video, which is determined by the measurement process. Uncertainty is aggregated and propagated in every transformation step: video capturing, object tracking, etc. Additionally, not only automatic video processing introduces uncertainty to the analysis process: humans may be more or less confident about decisions and hypothesis they make, too. To communicate this *confidence* information to the machine, we take advantage of *fuzzy sets*.

1.2 Contribution

This paper is an extension of the work previously presented at the *Third Workshop on Behavior Monitoring and Interpretation 2009* [25]. There, we proposed a system for visual analytics of video data introducing a novel method for the interactive and exploratory analysis of videos. The visualization of the video sequences superimposed by automatically extracted trajectory data enables the analyst to keep track of the important parts of the video, while not losing the context. Further, the definition of relevant information by filtering for trajectories' movement description leverages scalability of the analysis.

The extensions in this paper additionally contribute to quality-aware analysis of videos. Among the calculation and fusion of uncertainty information in every step of automatic feature extraction (Section 4), visual feedback is provided to communicate associated data quality to the users, as frequently demanded [58]. Therefore, we extend the VideoPer-

petuoGram (VPG), a video stream visualization introduced by Botchen et al. [6], to provide positional uncertainty information, and to improve context information (Section 5).

Further, integration of fuzzy filter definition similarly enables users to formulate hypotheses and assumptions according to their confidence about these ideas (Section 6.3). To facilitate interactive trajectory query and to compensate inexperience, filter definitions are kept easy-to-use (Section 6.2) and analysts are supported by context-sensitive statistical graphics depicting the distribution of trajectories' movement properties (Section 6.4).

Hence, beyond scalability the proposed visual analytics approach allows for reliable video analysis, since high-level pattern recognition tasks are processed by human experts, aware of the uncertainty inherent in the data.

2 Related work

2.1 Video analysis

In the last decade, several approaches aiming to aid inspection of video streams were proposed.

Some video analysis systems guide users' attention by alarms to certain events in order to increase their efficiency and effectiveness. These events have to be modeled beforehand. Examples are approaches for left-luggage detection [3], fire detection [38], loitering pedestrians detection [28], and the detection of forbidden area violations [36].

Forensic video processing systems provide another type of video analysis. Such an environment is proposed by Jerian et al. [29]. They apply several processing functions to video footage for the purpose of enhancing the video quality and therefore enable the investigation of important regions, such as the license plate of a car. The processing functions include shifted lines correction, compression artifacts reduction, projective registration, and motion deblurring.

There is only little prior research on visual analytics for video surveillance. Related approaches in the literature are systems supporting exploratory search, such as the one proposed by Christel [11]. His system is capable of exploratory query in video libraries, such as broadcast news databases. He aims to retrieve relevant videos out of huge video databases by defining filters. The filters make use of text annotations and image classification algorithms to refine the set of relevant videos. This allows search for shots conveying different characteristics, such as "is an outdoor shot," "includes buildings," or "includes faces." The search is also supported by the annotations of time stamps, locations on a map view, and keywords in speech.

Ferguson et al. [19] integrates content-based analysis techniques into *interactive TV* (*iTV*) devices to explore videos and video archives. Therefore, they present keyframes that represent automatically segmented shots or scenes in the video, and a video review (the playback of these keyframes) for each video in the archive. Their system is capable of searching for video parts based on text (e.g., originating from subtitles), low-level visual features (e.g., color, texture, shapes), and faces. They propose to process specific video content, such as sports and news, differently to take advantage of their particular characteristics.

Video visualization techniques depict the content of a video in a different way than the traditional video representation: Daniel and Chen [14] apply volume rendering techniques to visualize videos; Chen et al. [9] propose a horseshoe visualization; and Botchen et al. [6] introduce a visualization called *VideoPerpetuoGram* (VPG). The VPG is a seismograph-like

visualization technique for continuous video streams. A 3D video volume, where time is extruded in the third dimension, combines keyframes for context information with extracted attributes in a single visualization. We have adopted the visual design of VPG in this work.

Video synopsis approaches [50] show several actions simultaneously even if they occur chronologically in succession. Showing these non-overlapping activity regions at the same time reduces the time requirement to watch a video, but may also be confusing. Caspi et al. [8] select and merge informative poses of objects to generate images or short video clips. To avoid occlusion they rotate and translate the poses in the video volume.

Other approaches introduce *video navigation* techniques to ease and speed up video exploration. For example, Schoeffmann and Boeszoermenyi [51] extend the time slider to show *navigation summaries*. Navigation summaries are content abstractions of the video highlighting different temporal aspects, such as visited frames, dominant colors, or motion. Another approach to enhance video browsing enables the user to interactively drag objects in video [15]. Here, moving an object causes the video to shift temporally to the frame where the object is located at the desired spatial position.

Adapting the fast-forward speed is another approach to watch video in less time. Irrelevant parts of the video are played faster than periods of interest. The definition of interesting parts vary: motion features [47], similarity to target clips [48], semantic rules and user input [10], and the information theory [23] are used as measures.

2.2 Trajectory analysis

The presented work also has substantial overlap with other research domains, such as geographical information science or spatiotemporal databases. Besides differences in terminology and emphasis on different issues the challenges are quite similar: acquisition and description of trajectories as well as their query, classification, or summarization. The latter often utilize an iterative process that involves context-dependent trajectory visualization.

In the field of geographical information science, Mark [40] termed "the continuous set of positions occupied by an object in geographic space over some time period" a *geospatial lifeline*, whose "data consist of discrete space-time observations, ... describing an individual's location in geographic space at regular or irregular temporal intervals." Geospatial lifelines often originate from GPS-tracking data and are analyzed in an interactive exploratory fashion involving their visualization (e.g., [1, 16]). These methods bridge the gap to visual analytics. There exist several visualizations of spatiotemporal data in geographical information science (see Andrienko et al. [2] for a survey). In particular, we point out the *space-time cube* [32, 33], whose foundations go back to Hägerstrand's *time geography* [22], since it is quite related to the VPG, which we apply for video visualization. Important progress was also achieved for the description of geospatial lifelines and their analysis with respect to the temporal context (e.g., [27, 35]).

Research in spatial databases was originally motivated by geographical information systems, which themselves served as foundation (together with temporal databases) for spatiotemporal databases. *Moving object databases* (MODs) are one particular line of research that "deal with moving objects whose position is recorded at, not always regular, moments in time" [34]. The database community has especially contributed to the representation of spatiotemporal objects (e.g., trajectories), their comparison, indexing, and query. Additionally, uncertainty inherent to the measurement of trajectory data as well

as query-imprecision support was in focus of research on moving object databases. For further introduction we refer the reader to the textbook of Güting and Schneider [21].

Wolfson presented in his vision paper [62] an overview of operators to query for trajectories. We integrate some of these operators into the filters introduced in Section 6.

2.3 Uncertainty

To derive reliable decisions, the process of decision making has to deal with data of different qualities. The quality is determined by the degree of uncertainty and the amount of missing values that are propagated and aggregated by the transformations applied to the data.

For handling uncertain information, Correa et al. [13] first generate a model for source uncertainty (*uncertainty modeling*) using parametric models. During data transformation, the uncertainties are propagated using sensitivity modeling and aggregated via error modeling. They denote the uncertainty of the derived data and its sources by *uncertainty-aware transformations* and finally map them to visual representations to complement the view.

Pang et al. [46] introduce a visualization pipeline that distinguishes between three types of uncertainty: data acquisition leads to *data uncertainty*; data transformation to *derived uncertainty*; and the visualization process to *visualization uncertainty*. Similar to Correa, they propose to display data and uncertainty in a combined visualization, too.

A comprehensive survey of uncertainty visualization is provided by MacEachren et al. [39]. They discuss several approaches ranging from uncertainty measurement and assessment, via coping with uncertainty in information analysis and decision making, through to uncertainty visualization. They reviewed literature both from the geographical information science community as well as from the scientific and information visualization community.

Uncertainty models of trajectories have been discussed within the moving object database community, too. Trajcevski et al. [59] and Wolfson [62] propose a threshold to define a cylindrical *uncertainty zone* around the linear interpolation of two sample points in which the real trajectory of the moving object is situated. Position updates are only sent to the database if the object leaves its uncertainty zone. Wolfson also demands query-imprecision support and introduces predicate pairs for spatiotemporal query such as always/sometimes, possibly/definitely, and somewhere/everywhere.

A more sophisticated method was proposed by Pfoser and Jensen [49]. They distinguish between measurement error and sampling error. The latter introduces positional uncertainty of the trajectory in time t_x between two consecutive sample points. In their approach the likelihood of the object's position in t_x is equally distributed over an area (termed *lens*) determined by the object's estimated maximum speed. The lenses of the time between two subsequent trajectory samples form two linked cones, a lower cone with its apex at the first sampling point oriented forward in time, and an upper cone with its apex located at the second sampling point and direction backward in time. This shape is known as *space-time prism* [42] or *lifeline bead* [27]. A whole trajectory, sampled in time, is then represented as chain of lifeline beads, called a *lifeline necklace*. Kuijpers and Othman report a three times higher precision for lifeline beads than the cylindrical method [34]. Our approach to describe the positional uncertainty of a trajectory is related to lifeline beads inspired by Pfoser and Jensen, but starts from a different data model.

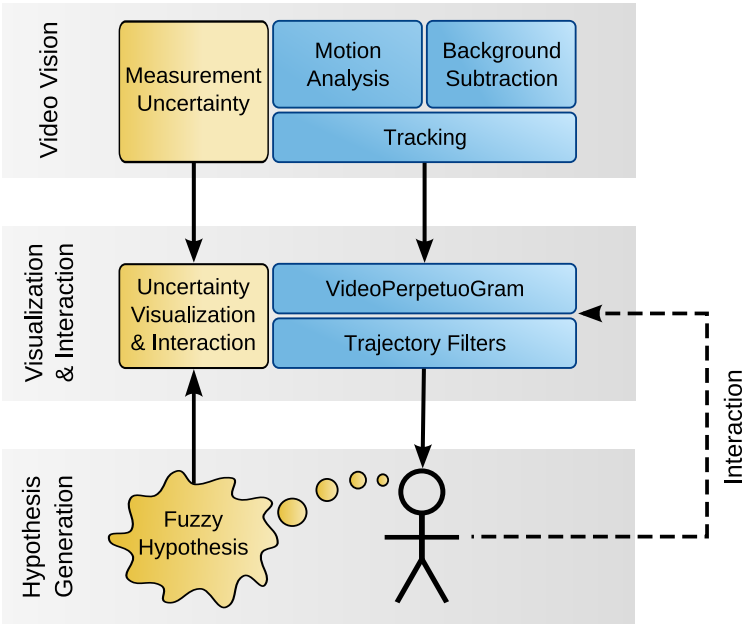

Figure 3: The structure of the proposed visual analytics framework.

3 Uncertainty-aware video visual analytics

In the introduction we briefly illustrated the basic concepts of visual analytics in general. This section explains our approach to video analysis by visual analytics.

The different stages involved in an iterative visual analytics process are sketched by Keim's mantra: "Analyze first; show the important; zoom, filter, and analyze further; details on demand" [31], cf. [53]. In the proposed approach, these stages are addressed in three parts: video vision, visualization and interaction, and hypothesis generation (Figure 3). Keim's first stage is covered by the video vision pipeline presented in Section 4, while the two latter stages largely involve human analysts for reasoning: creating assumptions and hypotheses and as a final goal, gaining insight. Visualization and interaction techniques are the connecting parts of visual analytics. They link automatic low-level feature extraction to high-level pattern-recognition by human analysts.

Among other tasks, video representation has primarily to support the navigation and orientation in spatiotemporal data space. Otherwise, users may lose spatial or temporal relationships between several events. Further, the abstraction capability of the visualization is important to enable fast and scalable exploration of a video sequence. It is fundamental to human-centered video analysis that the visualization supports the adaption of the level of detail for any particular region of the video and that it provides access to additional information, such as metadata, previously extracted features, or statistics.

Video analysis with vaguely defined search targets typically starts with the visual exploration of a huge amount of data of low detail and evolves into inspection of a small amount of data of high detail. In interactive query refinement, two principles turn out to be important: the analysis process focuses on relevant data and controls the level of detail.

This raises the question which part of a video sequence is relevant to the analysts. Further, it motivates the adaption of level of detail of the information presented by visualization. We will examine these issues in Sections 5 and 6.

Findings of analyses without knowing about their quality are generally of no avail. Especially for decision making based on the results of an analysis, it is indispensable to know about the reliability of the provided information. Uncertainty of information originates from the various transformations applied to the data, reaching from measurement via video processing through to visualization and perception. But also the definition of relevance, assumptions, and hypotheses by human analysts introduce uncertainty to the information, as depicted in Figure 3. A major contribution of this paper is to cope with uncertainties in all stages, utilizing uncertainty aggregation and propagation in the video vision part, fuzzy hypotheses definition on the analyst's side, and uncertainty visualization to communicate the quality of features to the human.

4 Video vision

Tracking moving objects is an elementary stage of the spatiotemporal analysis of video sequences. In our approach, we combine well-known, basic computer vision techniques to achieve simple but robust feature extraction (see Figure 4).

Background subtraction classifies the pixels of a video frame to foreground and background utilizing a background model. Assuming the background to be static to some extent, we apply the running Gaussian average method [63] that utilizes a single Gaussian luminance distribution per pixel to describe the background. The background model is updated by the currently processed video frame to cope with noise and gradually changing illumination. Foreground objects are retrieved by thresholding the subtraction of the current video frame and the background model.

By *motion segmentation* we identify areas of homogeneous movement, which are considered to be foreground objects. To calculate the required optical flow field of successive frames we apply the pyramidal *Lucas-Kanade method* [7]. We use these two segmentation approaches (background subtraction and motion segmentation) to avoid wrong trajectories and false alarms due to encoding artifacts or badly initialized background models. Finally, we refine the object regions by morphological operations utilizing prior knowledge of the video material (e.g., codec block size) and fuse the results of both segmentation methods. The fused object blob as well as the results of both segmentation methods are illustrated in Figure 5.

Object tracking integrates the observation of foreground regions of several frames. For this purpose we utilize a linear *Kalman filter* [61] with a dynamics model of second order derivative and trace the extracted blobs. To support multiple object tracking, the object's dynamics model represented by the Kalman filter is used to predict a target window (cf. red circle in Figure 5). Based on this prediction, target windows and object observations are associated by the *Hungarian algorithm* (also termed Munkres), according to their distance. The new observations are used to update the Kalman tracker. For trajectories without associated observation, for instance because of object occlusion, we apply an update rule according to Cipra and Romera [12].

After trajectories of moving objects within the video sequence are extracted, a selection of movement properties is calculated. These properties are a subset of the movement de-

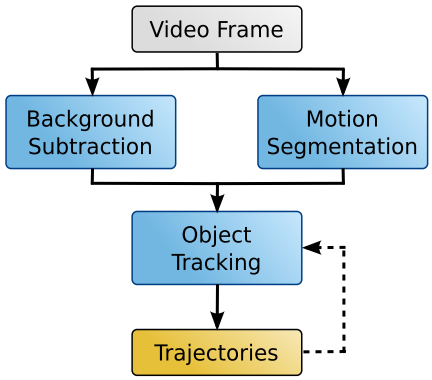

Figure 4: Workflow of automated trajectory extraction.

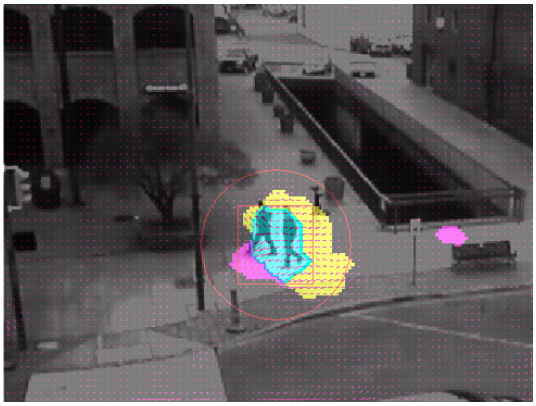

Figure 5: A video frame superimposed by the results of the different segmentation approaches. Blobs stem from background subtraction (magenta), motion analysis (yellow), and the fusion of both results (cyan). Small pink lines illustrate the optical flow. The object's position in previous frame is represented by red rectangle, while the red circle limits the prediction window.

scriptors by Laube et al. [35], which they use to specify and compare geospatial lifelines captured by GPS devices. We apply the movement descriptors of global trajectory context for interactive query introduced in Section 6. The subset of properties we use consists of location, speed, and movement azimuth.

To draw correct conclusions the analysts have to know about the quality of the information provided by the automated extraction process. Hence, we communicate the algorithmic uncertainty to the video visualization, which is an extension of our previous work [25]. Since every step of the trajectory extraction process introduces (or at least propagates) uncertainty, the quality of each result has to be evaluated and, finally, fused together to a combined uncertainty estimation.

Our uncertainty model is based on the following assumptions and simplifications. We neglect the uncertainty originating from the video capturing process, which basically re-

flects sensor noise and coding noise. We consider its influence as negligible since the effect of noise is alleviated by morphological post-processing of the segmentation result. Further, we only focus on data uncertainty and in general omit handling of missing values and contradictory data. Thus, it is possible that some objects (i.e., blobs, trajectories) will be missed due to inappropriate thresholds. Likewise, we do not consider any systematic error that leads to reduced *accuracy*; we solely regard random error that affects the *precision* of the result. Finally, we assume temporal precision of video sampling to be high enough for our purpose. Hence, we ignore it and focus only on spatial precision of the extracted features.

The different computer vision methods we apply introduce their own methods of how measurement uncertainty is estimated. To model the uncertainty inherent in the two segmentation approaches, we calculate each pixel's probability of being correctly assigned to foreground or background class. The uncertainty of both segmentation results is then fused by averaging. Regions where blobs of both methods overlap are considered as object areas and are further tracked (cf. Figure 5). Similar to Pfoser and Jensen [49], we only maintain trajectories of moving points, in contrast to traces of whole blob areas. Hence, we calculate the mean and covariance of the blob area weighted by the segmentation quality.

The classification uncertainty of a pixel is retrieved by widely accepted methods, according to the segmentation approach. For background subtraction, we exploit the standard deviation of the background model to specify the likelihood of being well classified. The quality of the optical flow field obtained by the Lucas-Kanade method is largely determined by the cornerness of the tracked image region, which is due to the aperture problem and exploited by the popular Kanade-Lucas-Tomasi tracking technique [52]. According to Barron et al. [4], we use the magnitude of the smallest eigenvalue of the system matrix to determine the uncertainty of the estimated optical flow.

In the Kalman filter, we treat the mean of the uncertainty weighted blob area as the observation of the object's location, while its covariance is considered as the normal distributed measurement noise. In consequence, the trajectory is modeled by the location (filter state) and the positional precision (a posteriori error covariance matrix) of the moving point in each frame. The square root of the error covariance is used for visualization of trajectories' positional precision in Section 5 according to standards of measurement uncertainty [56]. We assume the temporal sampling of video frames to be dense enough to ignore positional deviation of the trajectory between two consecutive samples. If observations were missed for some frames, for instance in cases where the object is occluded, the trajectory is extrapolated according to its dynamics model. Inaccuracy of the dynamics model in combination with the process noise (which covers the simplifications of the dynamics model) leads to increasing uncertainty of the predicted positions.

In case of missing observations, the proposed method of modeling the trajectory uncertainty is quite similar to lifeline beads [49]. The main difference between lifeline beads and the proposed approach is that the former uses a uniform distribution to model the object's likelihood of being at any location within a bead, which is constrained by the object's estimated maximal velocity. In contrast, the approach based on Kalman filtering uses a multivariate Gaussian distribution with respect to the recent dynamics model and its accuracy to define the probability density function of the moving object's location.

Based on the ground plane projected locations of the moving object, further movement descriptions are extracted. In contrast to different levels of granularity [35], we evaluate the descriptions of the whole trajectory, since the length of trajectories extracted from video

footage is generally short. Movement descriptions, such as speed:

$$T_{\text{speed}} = \frac{N}{fps} \sum_{i=1}^{N-1} \sqrt{\left(T_{\text{pos}}^{(i)} - T_{\text{pos}}^{(i+1)}\right)^2} \tag{1}$$

or movement azimuth:

$$T_{\text{azimuth}} = \frac{180}{\pi} \cos^{-1} \left(\left[\frac{T_{\text{pos}}^{(1)} - T_{\text{pos}}^{(N)}}{\left\| T_{\text{pos}}^{(1)} - T_{\text{pos}}^{(N)} \right\|} \right]^T \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right) \tag{2}$$

are solely based on the object's locations. Hence, we use error propagation [41] (also called *Gaussian error propagation* or *linear error propagation*) to calculate their uncertainties. Since, we assume the uncertainties to be Gaussian distributed, they are expressed as the standard deviations σ_{speed} and σ_{azimuth} (cf. Appendix for derivation):

$$\sigma_{\text{speed}}^{(1,N)} = \frac{N}{fps} \sqrt{\sum_{i=1}^{N-1} \frac{(v^{(i,i+1)})^T C_v^{(i,i+1)} v^{(i,i+1)}}{(v^{(i,i+1)})^T v^{(i,i+1)}}} \tag{3}$$

$$\sigma_{\text{azimuth}}^{(1,N)} = \frac{180}{\pi} \sqrt{\frac{1}{1 - \frac{(x^{(1,N)})^2}{(v^{(1,N)})^T v^{(1,N)}}} \left[\frac{var_x^{(1)} + var_x^{(N)}}{(v^{(1,N)})^T v^{(1,N)}} + \frac{(x^{(1,N)})^2 ((v^{(1,N)})^T C_v^{(1,N)} v^{(1,N)})}{((v^{(1,N)})^T v^{(1,N)})^3} \right]} \tag{4}$$

with vector:

$$v^{(i,j)} = \begin{pmatrix} x^{(i,j)} \\ y^{(i,j)} \end{pmatrix} = T_{\text{pos}}^{(j)} - T_{\text{pos}}^{(i)}$$

and the according covariance matrix representing the uncertainty:

$$C_v^{(i,j)} = \begin{pmatrix} var_x^{(i,j)} & cov_{x,y}^{(i,j)} \\ cov_{y,x}^{(i,j)} & var_y^{(i,j)} \end{pmatrix} = C_{\text{pos}}^{(i)} + C_{\text{pos}}^{(j)}$$

where N is the number of equidistant trajectory fixes, sampled at the rate of fps (frames per second). $T_{\text{pos}}^{(i)}$ denotes the trajectory's position vector at sample i with the according covariance matrix $C_{\text{pos}}^{(i)}$. The superscript represent the time index of the sample or the vector between two samples, e.g., in the case of (i, j). Note that we assume the errors in $T_{\text{pos}}^{(i)}$ and $T_{\text{pos}}^{(j)}$ (for $i \neq j$) to be independent and uncorrelated.

5 Video visualization

Video visualization aims at displaying relevant parts of video, features extracted by video vision, and consequences of filter definitions. Uncertainty-aware video visualization has additionally to communicate uncertainties originating from the video vision stage, and confidences from the filter definitions. For the visualization of videos, we utilize the

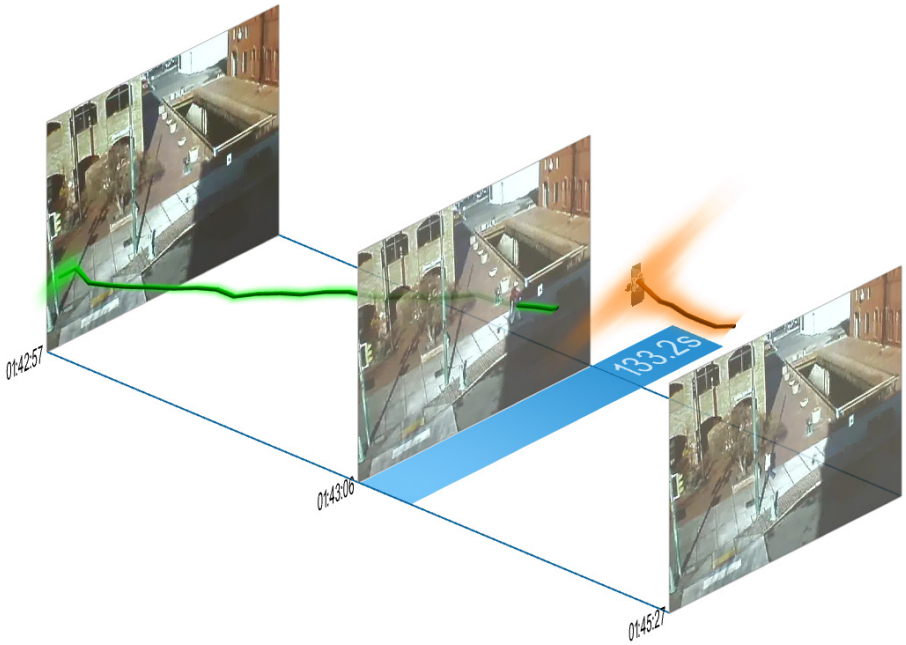

Figure 6: The video is visualized by VideoPerpetuoGram (VPG) containing the two spatial axes and the temporal axis of a video. Three keyframes with their time stamps are depicted to convey context information. Although there are just few keyframes, the video volume is dense and shows detailed trajectories (in the space-time volume) with their degree of membership (DOM) related to filters mapped to color (green: high; red: low). Their positional uncertainty is illustrated by semi-transparent blur. The green trajectory, featuring high DOM, shows a person walking from left across the footway. The orange trajectory, featuring moderate to low DOM, represents a car appearing for a short period of time in the upper part of the video. Trajectories with a DOM below a user-defined threshold are hidden, and periods without trajectories are skipped. Between the second and third keyframe, no trajectory is left. Therefore, this period of time can be skipped and is cut off the (afore dense) video volume. The blue bar indicates the gap in the video volume and the amount of time skipped.

VideoPerpetuoGram (VPG) [6] and enhance it to provide positional uncertainty information of trajectories.

The VPG visualizes continuous video streams similar to the illustration of seismographs and electrocardiograms (ECGs). Features, such as trajectories of moving objects, are displayed together with a couple of keyframes that convey the context of the original video data. The VPG depicts a video volume composed of the time axis and two spatial axes of the video frames. This representation supports the navigation and orientation in the spatiotemporal video space. Hence, users are enabled to keep track of spatial and temporal relationships between several events. This abstract visualization of the video helps to overview a long video sequence at a glance and allows interactive exploration of trajectories.

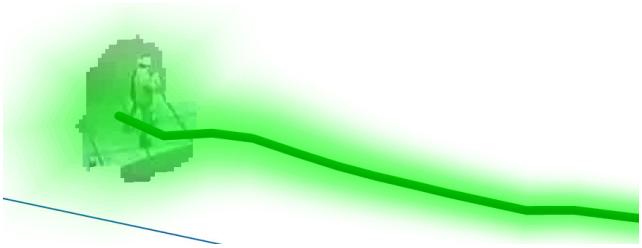

Figure 7: The trajectory of a moving object is depicted by a geometric tube. A blob in the beginning conveys context information, i.e., to which object this trajectory belongs. DOM with respect to filters and positional uncertainty is depicted by color and semi-transparent blur, respectively.

Figure 6 shows our VPG visualizing a part of the video stream from the IEEE VAST Challenge 2009 [60].

There is a trade-off between using abstract visualizations and visualizations similar to standard video playback. The former provide faster exploration with more details than the latter, but are unfamiliar. We decided to use an abstract visualization since it was demonstrated by a controlled user study that users are able to learn abstract visual signatures [9]. Even visualization novices are able to understand the VPG [24] after introducing simple related 2D visualization designs.

To convey data quality, uncertainty information has to be provided by the VPG. There exist different visual mappings of uncertainty discussed in literature: adding glyphs, adding or modifying geometry, modifying attributes (e.g., color and shading), using animation [46], and applying transparency [13]. Botchen et al. [6] use two levels of saturation to indicate the relationship plausibility of trajectories in the VPG. We use color to encode the degree of membership (DOM) of the trajectories with respect to the defined filters (see Section 6). In addition, quantitatively expressed trajectory quality as well as further information of the trajectory can be acquired on demand (cf. Section 6.6). Trajectories with a DOM below a user defined threshold are hidden, and periods without trajectories are skipped. The blue bar indicates the amount of time skipped.

A contribution of this paper is to superimpose the positional uncertainty of the video vision stage in the VPG by semi-transparent blur. This represents the probability density function of the object's location and is realized by an elliptical tube with according transparency distribution. The uncertainty visualization enables the users to be aware of the trajectory's quality and facilitates reliable conclusions.

Another extension of the VPG aims to improve context information. The trajectory's start position is therefore augmented with the blob of the moving object (cf. Figure 6, orange trajectory, and Figure 7).

6 User interaction

The objective of human-computer interaction in video visual analytics is mainly to enable users to explore video data and to verify the derived hypotheses. Therefore, the user interaction has to be closely connected with the visualization.

In general, users start the analysis process of a video sequence by exploring it to derive the first hypotheses. Basic concepts for navigation, selection (to access additional information), and magnification (adjustment of information granularity) are widely established and adopted for the special needs of video analytics.

6.1 Relevance definition for video parts

After exploration has led to first findings, a second type of interaction gains in importance: the definition of relevance of certain video parts. This is the key element to achieve scalability for large amounts of video data. Both interaction schemes—the exploration and the feedback of user-defined relevance based on hypotheses—are fundamental elements of video visual analytics and define the analysis process: users generate insight by iteratively deriving hypotheses, verifying, adapting, accepting, or rejecting them.

In the proposed framework, relevance is defined by the formulation of filters. An introductory example should help to understand the relationship between hypothesis, relevance, and filter: if we have the hypothesis that the person searched for is an employee traveling by bus, only those trajectories are relevant. Hence, we define filters to constrain the set of trajectories to those leaving the company and walking into the direction of the bus stop (cf. Figure 8). Please note that the term relevance is used here related to users' interests. Of course, other users, searching for other events, may be interested in different parts of the footage and therefore define relevance differently. The set of supported filter types stems from the trajectories' movement description calculated in the video vision stage and covers various types: location filter (cf. Figure 9), filter for movement azimuth (cf. Figure 10), lifetime filter, speed filter (cf. Figure 11), black list, and white list. Additionally, a relationship filter (cf. Figure 12) allows the user to filter trajectory relations. The particular filters will be discussed in Section 6.2.

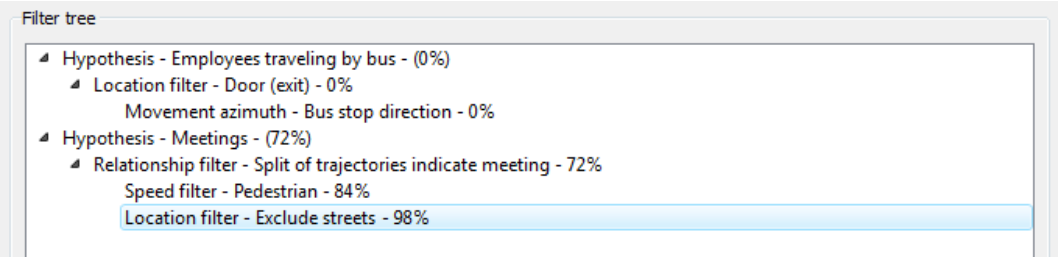

Figure 8: Exemplary filter tree. A particular filter is represented by its type (e.g., "Location filter") and its user-defined description (e.g., "Exclude streets"). The union operator is applied to fuzzy sets of filters of the same level (e.g., "Speed filter—Pedestrian" and "Location filter—Exclude streets"). Sets of filters at successive levels are intersected (e.g., "Location filter—Door (exit)" and "Movement azimuth—Bus stop direction"). Filter containers are used to aggregate filter subtrees in hypotheses. The percentage values next to the filters denote the DOM of a selected trajectory to this filter. The percentage values in brackets next to filter containers express the DOM to the whole hypothesis (i.e., the resulting DOM regarding all intersection and union operations of the filters in this filter container).

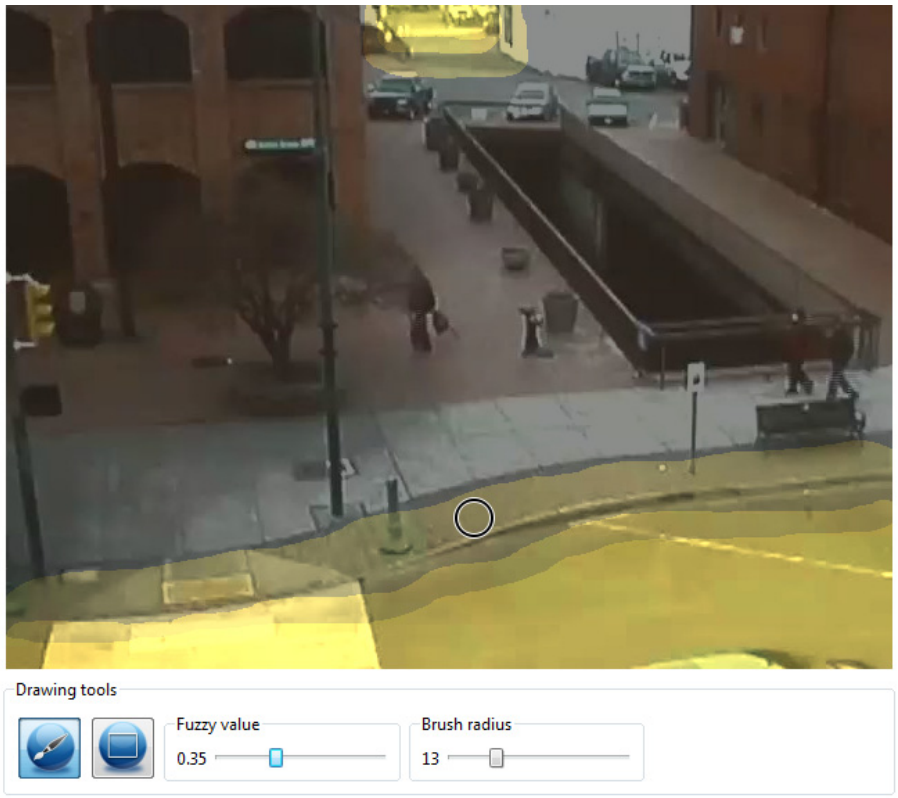

Figure 9: Interface of the location filter. Spatial areas of interest (yellow) can be brushed on a keyframe similar as done in applications such as Photoshop. The fuzzy value between 0 and 1 that will be brushed is selected by the "Fuzzy value"-slider. The fuzzy value of an area of interest is depicted by saturation of the yellow color. This particular location filter can be used to exclude trajectories situated on streets (cf. Figure 8 "Location filter—Exclude streets").

To enable users to keep track of the semantics and purposes of the filter definitions, filters can be annotated. These descriptions are given by users to explain the intention of the filter (such as "Exclude streets" or "Meeting") and provide an overview.

The filters can be arbitrarily combined using intersection, union, and complement operators on fuzzy sets to form complex filter trees. Filter containers are used to aggregate filter subtrees in hypotheses. An exemplary filter tree is illustrated in Figure 8. In fuzzy sets, intersection can be applied by any *t-norm* and union by the corresponding *t-conorm*. The most common t-norm and t-conorm for fuzzy sets are Zadeh's *min* and *max* [64] operators, which are applied by default. However, the users are not restricted to them and may apply any t-norm and t-conorm. For an introduction to fuzzy sets and their operators we refer to the textbook of Siler and Buckley [55].

A major contribution of this work is to support users in their filter definition process. Therefore, we designed a user interface that complies with the following five essential interaction guidelines:

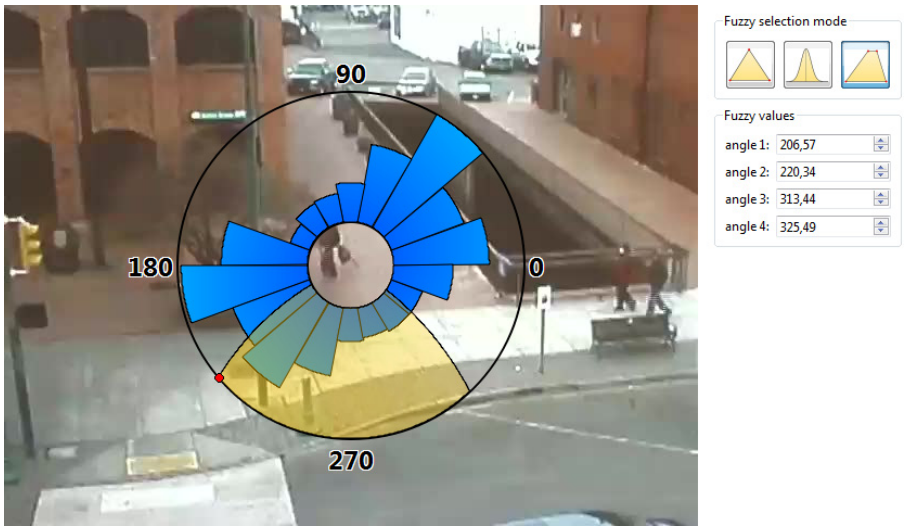

Figure 10: Interface for movement azimuth selection. A keyframe is depicted for context information. The histogram (blue bins) shows the distribution of trajectories' directions. The yellow area represents the selected directions. On the right side, users can choose to model their uncertainty by different fuzzy selection modes, described in Section 6.2. This particular selection of the movement azimuth will focus on trajectories walking from top to bottom.

1. easy-to-use filter definition;
2. confidence-incorporated filter definition;
3. decision-guided filter definition;
4. filter feedback; and
5. details-on-demand.

6.2 Easy-to-use filter definition

Since filter specification is an important element in interactive exploration, it should be easy-to-use. The International Organization for Standardization (ISO) describes a set of usability heuristics for dialogs. One of these is the "Conformity with user expectations."[4] This principle demands that an application should be consistent and in accordance with the users' expectations. In our case, this implies that the filter formulation has to support the users in defining filter parameters according to their real world associations. The input of locations, directions, distances, and so on has to be done in a manner that matches the knowledge and experience of the users with these attributes.

In particular, the location filter (see Figure 9) is defined by drawing in the image similar to applications such as Photoshop. A keyframe is depicted in the background to convey the spatial context information to the user.

[4]ISO 9241-110:2006 Ergonomics of human-system interaction—Part 110: Dialogue principles; conformity with user expectations

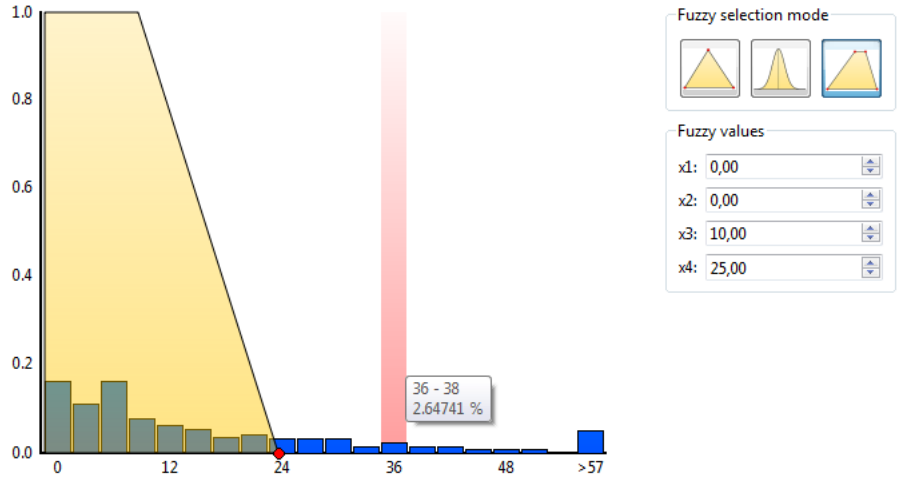

Figure 11: Speed filter. A histogram (blue bins) shows the distribution of the mean speed of the trajectories. The yellow area represents the selected speeds. This particular speed filter can be used to focus on trajectories at walking speed (cf. Figure 8 "Speed filter—Pedestrian").

For the selection of the movement directions it is important to put the movement azimuth in context to the image space. By a compass-like visualization we illustrate how directions are embedded, while a keyframe contributes scene context. The highlighted parts of the circle represent the desired and supported directions (see Figure 10). Hence, a direct match between the movement azimuth and the scene is established. This facilitates that movement directions of desired trajectories can be easily estimated.

Interactions between moving objects can occur if they share nearby spatial locations at a similar time. Thus, two attributes, the spatial distance and the temporal distance, are of interest for a relationship filter. Figure 12 shows the interface to provide the spatial distance of the relationship filter. To increase users' awareness of their defined distances at different locations, we project a distance circle on the ground plane, as the mouse hovers over the keyframe. Please note that the relationship filter is independent of the absolute spatial location: only the relative distance between trajectories is considered. The temporal distance of the relationship filter, as well as filters for lifetime or speed of trajectories, are fed into the system with a dialog similar to Figure 11.

Endert et al. [18] reported on observations of analysts working with the IEEE VAST 2009 Challenge data. One of the outcomes was that professional analysts like to start with intuitive GUIs to sketch filters and in this way explore data rapidly. Later, they prefer to enter exact values. We support both interaction modes: users can drag and drop values (e.g., movement azimuth and speed filter, red points in Figures 10 and 11) and brush their desired locations (location filter, Figure 9) to sketch filters roughly as well as enter the exact values (cf. right panels of Figures 10 and 11).

Wolfson [62] introduced operators for retrieving trajectories that stand in certain relationships to a region: *always/everywhere* and *sometimes/somewhere*. For the location filter and the spatial distances of the relationship filter, we require trajectories to *start at*, *end*

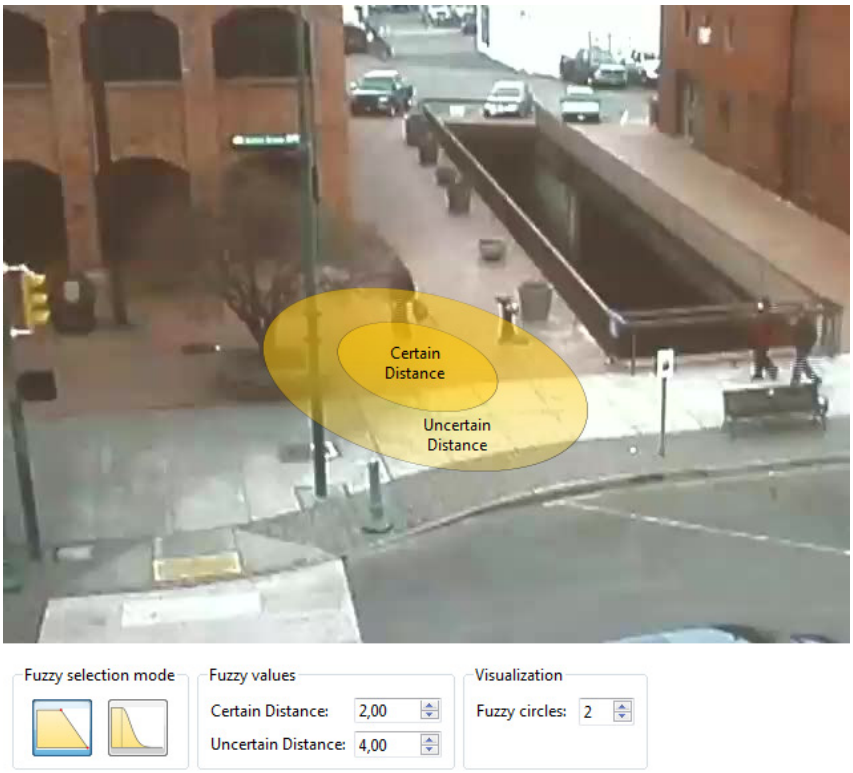

Figure 12: Interface for spatial distance selection of the relationship filter. A keyframe is depicted for context information. Yellow ellipses indicate the certain and uncertain spatial distances projected on the ground plane at the mouse position helping users to estimate distances. This particular relationship filter can be used in combination with the property *start at* (property selection not visible here, cf. Section 6.2) to detect meetings (cf. Figure 8 "Relationship filter—Split of trajectories indicate meeting" and Section 7).

at, be *partially in*, or to be *completely in* a region. Wolfson uses the two discrete states to model uncertainty: *possibly* and *definitely*. In contrast, we assign trajectories to filters by a continuous DOM.

6.3 Confidence-incorporated filter definition

The human-centered and hypothesis-based relevance filtering introduces a new level of abstraction. Since filters are associated with assumptions, they are a source of uncertainty, too. Filter definitions have to enable analysts to incorporate their confidence about hypotheses into the relevance feedback. In consequence, the relevance feedback must not strictly filter for trajectories, but has rather to cope with fuzzy decisions based on user-defined confidence values.

To address this issue, we represent filters by fuzzy sets and allow analysts to model their confidence by fuzzy membership functions. We support three different membership

functions: Gaussian, triangle, and trapezoid set functions. These functions are the most common fuzzy sets functions [43], and they are modeled by only few parameters: two (Gaussian: mean and deviation), three (triangle), or four (trapezoid) (cf. Figures 10 and 11). Additionally, these functions are also very common to represent measurement uncertainty distribution in calibration applications, for instance. For the relationship filter, we use a single-sided fuzzy membership function (termed *Z-function*), such as the one in Figure 12. For the location filter (see Figure 9), the fuzzy values of arbitrary areas can be set separately.

A trajectory's DOM to a filter is calculated by the uncertainty originating from the video vision stage (data quality) and the confidence provided by the users. Therefore, we integrate the product of an uncertain attribute of a trajectory and the fuzzy set function of a filter. The resulting interval of the DOM is between 0 and 1.

The filters differ in the complexity and discretization in calculation of the DOM. While the location filter has to regard all three spatiotemporal dimensions, other filters consider only a single attribute, for example the mean speed of a whole trajectory. The discretization is either introduced in the filter evaluation step (i.e., the uncertain attribute is evaluated at sample points) or is already available (e.g., pixels and frames).

6.4 Decision-guided filter definition

Thomas and Cook identified the importance of "visual analytics systems to support the analyst in executing sound analytic technique routinely, facilitating insight and sound judgment in time-pressured environments and compensating for inexperience wherever possible" [58].

To support the analysts, we supply background information for decision guidance by presenting context-sensitive graphical statistics. Context sensitivity denotes the selection of the provided statistics according to the filters. For different filters and when appropriate, we depict normalized histograms of the filters' attributes.

To conform to users' expectations, the appearance of the histogram is adapted to the filter. For example, the histogram of the movement azimuth filter is arranged in a circle in order to ease derivation of assumptions of trajectories' behavior. Investigating the histogram in Figure 10, one assumption may be that the amount of people walking on the footway from right to left outnumbers the people walking into the opposite direction. This supports users to define relevance, derive hypotheses, and to create appropriate filters.

Further, such context-sensitive graphical statistics help also to reject previously made incorrect assumptions, for example that there are two predominant speed intervals: one for people and one for cars. The histogram of the speed filter shown in Figure 11 tells us that there are no sharp speed intervals. This distribution arises from cars decelerating and stopping at the signal light and bicycles at intermediate speed. Therefore, the assumption has to be rejected.

As already mentioned above, we complement the context-sensitive graphical statistics with keyframes for spatial context information where appropriate (Figures 9, 10, and 12).

6.5 Filter feedback

Feedback is essential for users to verify their hypotheses by filter definitions. Without filter feedback, users cannot know whether their filter formulations are too weak or too restrictive. More than that, users are left in the dark whether the filters operate as intended.

They are also not aware if the filters cause side-effects. We provide filter feedback of three different kinds.

During the filter definition, we do not remove the trajectories from the VPG, but immediately map their DOMs to the color of tubes (cf. Section 5). The changes have to be explicitly applied. In this way, users are made aware of how the filters affect the trajectories.

The second filter feedback shows for a selected trajectory the DOM to each single filter. Additionally, the combined DOM to each hypothesis (including the intersections and unions of the filters) can be monitored in the filter tree as illustrated in Figure 8.

The third kind of filter feedback displays the amount of remaining trajectories in contrast to the number of all trajectories extracted from the video. Hence, users get an impression of how restrictive their filter formulations are.

6.6 Details-on-demand

Details-on-demand techniques provide an overview of a larger region of interest (in our case the video part displayed in the VPG), while users may select parts of data to be visualized in more detail. This combination allows for in-depth analysis without losing the big picture. Since the proposed framework is mainly based on the trajectories of moving objects, detailed information of particular trajectories are indispensable.

Therefore, our framework supports the selection of a particular trajectory. The selected trajectory will be highlighted in the VPG and a window with detailed information about this trajectory appears. The detailed information includes all features extracted by the computer vision stage, such as the mean speed, movement azimuth, location, and temporal constraints, each with the inherent uncertainties. With this detailed information, the user can investigate particular trajectories to search for characteristic attributes. Having these characteristic attributes in mind, the user can define or refine filters that permit or exclude these trajectories.

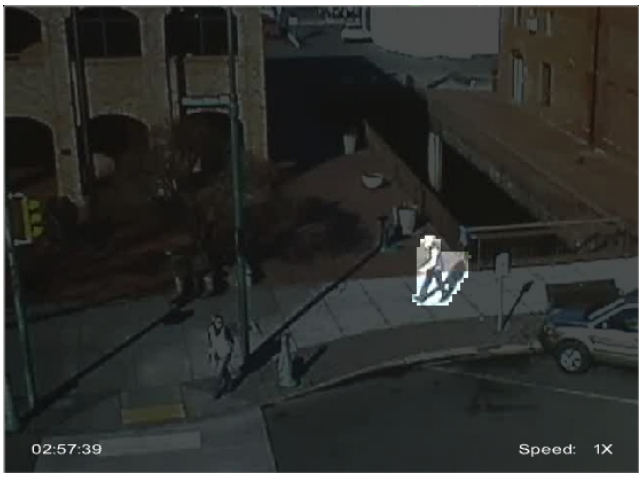

Figure 13: The part belonging to a trajectory can be played as conventional video. The region of the trajectory is highlighted.

Sometimes, whether for novices to learn the VPG or for further investigation of a trajectory, analysts want to see the native video stream of a trajectory. We support this necessity by a conventional internal video player that is able to highlight the region that belongs to a trajectory (Figure 13). Of course, it is also possible to play arbitrary parts of the video even if no trajectory is situated there.

7 Case study and evaluation

For the validation of our work we follow Ellis and Dix [17]. They claim that validation should consider two parts: justification and evaluation. The justification of our approach can be found in the particular sections. For the evaluation of our approach we participated at the IEEE VAST Challenge 2009 [60].

Shneiderman and Plaisant claim the IEEE VAST Challenge to be a good evaluation method: "the Visual Analytics VAST contest will combine both qualitative and quantitative metrics by using synthetic dataset which provides ground truth against which insights can be evaluated" [54].

The IEEE VAST Challenge has been part of the IEEE Symposium on Visual Analytics Science and Technologies since 2006. The IEEE VAST Challenge 2009 covered three mini challenges: *Badge and Network Traffic*, *Social Network and Geospatial*, and the *Video Analysis* mini challenge. Additionally, participation in the grand challenge, which combined the mini challenges, was possible. We participated in the Video Analysis mini challenge with our framework, and collaborated with two other groups who worked on the other mini challenges to solve the grand challenge.

The background story of the challenge was the suspicion that an embassy employee is corrupt and transfers data to an outside criminal organization. For the video analysis mini challenge, 10 hours of surveillance footage was provided and it was suspected "that at least one, perhaps more, meetings of persons associated with this case took place at locations captured by this security camera" [60].

According to the discussion in the introduction about well-defined and vaguely defined search targets, this is a vaguely defined search target: the detection of suspicious meetings requires analyis of the behavior as well as interpretation of the intentions of the involved persons. Although we have some constraints (search target: meetings), the problem definition includes many degrees of freedom, such as an arbitrary number of associated persons and the undefined appearance of the persons. Thus, solving this task solely by automated computer vision techniques is hard. In our approach, we address this issue by involving the analysts in the exploration of the video and delegate the responsibility to evaluate potential scenes of interest to them.

We will now explain how our filter framework was applied to solve the task. We started the VA process on the whole video containing 809 trajectories. In the beginning, we explored the video to find irrelevant parts. We detected that a lot of trajectories belong to cars driving on streets. Since we are searching for encounters of people, we made the assumption that meetings will not take place on streets. Thus, we created a location filter similar to Figure 9 to get rid of trajectories originating from cars. Cars are typically faster than pedestrians. Hence, the application of a speed filter similar to Figure 11 would be an option, too. After further exploration, we came up with the hypothesis that an encounter of people includes a discriminatory trajectory relationship. The trajectories involved in a meeting

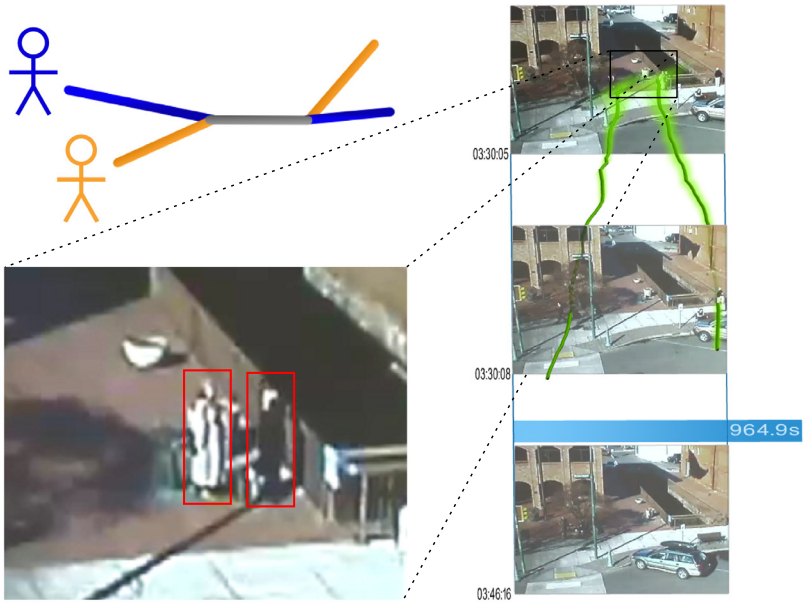

Figure 14: The split and merge of trajectories indicate encounter of people (upper left). At the split of the trajectories at the end of the meeting (right) we find the woman and the man involved in briefcase exchange (lower left, marked).

first merge (arrival of people) and then stay together while the meeting takes place. After the meeting, the trajectories of the participants finally split. To detect meetings, filtering for one of these stages is appropriate. Therefore, we formulated a relationship filter similar to Figure 12 with the condition *starts at*. That means, we are searching for the last stage: the end of the meeting. With these simple filters, the amount of remaining trajectories was decreased to 22, a manageable amount to investigate manually. Within the remaining video, we found a woman and a man involved in a briefcase exchange, the solution of the challenge (see Figure 14).

The task of deriving hypothesis, such as that meetings will not take place on streets and that meetings can be described by a split and merge of trajectories, is solely the responsibility of the analyst. It requires a large amount of knowledge (e.g., that persons do not meet on busy streets since it is dangerous) which is collected by humans through daily experiences. It is impractical to model the whole knowledge in the machine, and therefore the machine alone cannot solve this task. Nevertheless, the machine is appropriate to aid the user in the particular parameter definitions of filters, such as "Where are the streets located?", "How fast are the trajectories on the pedestrian walk in common?", or "How far is a distance of 5 meter in the projected image?" This is the kind of information that is provided by the proposed system. This also points out that the given solution may not be the optimal one. The filters are applied exactly in the order the analysts derived their hypothesis. There may be better and worse filter combinations that lead to the solution, depending on the ideas and skills of the analysts.

As mentioned above, we also participated at the grand challenge. This required us to check further events. For example, the collaborating groups came up with the hypothesis

that a particular building in the video is the embassy. To check this hypothesis, we defined a location filter with the condition *ends at* and marked the door to exclusively select trajectories entering the building. We then compared the entrance timestamps of these trajectories with those of the provided by badge data. They did not match. Therefore, the hypothesis was rejected: the building is not the embassy.

Another meeting we found included a cyclist. Here, a man arrived with a bicycle and had a short talk with another man. A query of badge traffic database for this timestamp indicated that a few minutes later someone arrived in the embassy. Since the embassy is in walking distance (this information was provided by the challenge description), this caused the hypothesis that this cyclist may be of interest. Hence, we defined two filters to extract trajectories of cyclists: a location filter selecting the pedestrian walk (the cyclist was solely driving on the pedestrian walk) was combined (by intersection) with a speed filter discriminating between pedestrians and cyclists. Now, we could investigate these trajectories further and found out that there were no suspicious event including the cyclist.

During analysis, we experienced several advantages, as well as some drawbacks of our system. The approach scales well for huge amounts of video data and a wide variety of tasks. The interactive filter formulation including background information for decision guidance helps to define relevant parts of the video. Decisions about how the filter parameters should be chosen are especially supported by context information and information on how a particular filter parameter is applied in the concrete video scenario. Additionally, the visualized statistics provide a good summary of trajectories' properties and therefore supports users to get a feeling about these parameters and helps them to formulate the filters.

The visualization of the positional uncertainty in the VPG helps especially in the definition of the location and relationship filter. Being aware of these positional uncertainties provides confidence to the users that their definitions of filters are not wrong because of inaccurate location calculations of the video vision stage and enables them to create the filters with respect to positional uncertainties. The color coded DOM of trajectories to the filters in the VPG indicates not only how well this trajectory matches hypotheses, rather it is also a signal for users that helps to validate the filters and often initiates further refinement. As expected, the possibility to model a fuzzy membership function simplifies the definition of filters because users do not need to derive sharp constraints from uncertain assumptions. We experienced that filter definitions are often formulated coarsely with a large contingent of uncertainty in the beginning. Here, the users "play safe" (i.e., they do not discriminate too many trajectories) and do not spend too much time on the initial filter definition. After a while, as they closer investigate the left trajectories and their DOMs, they reduce the modeled uncertainty and refine the filters to be sharper and more discriminative. Calculation and communication of uncertainty in all stages enables a reliable analysis process. Without data quality information from the video vision stage (communicated via video visualization to users and used for DOM calculations of filters) and the input of a user's confidence during fuzzy filter definition, the basis of decisions would be wrong. Considering uncertainty in combination with the expertise and the excellent high-level pattern recognition ability of the human analyst leads to the reliability of our system.

Fuzzy filter definition and uncertainties of the automatic processing step also support higher recall rates. This is due to color coding of trajectories' DOM indicating whether filters are badly adjusted or measured properties of trajectories are uncertain and thus do not match the intended filters. In contrast to sharp filtering process without uncertainty

consideration, poorly fitting trajectories are not immediately removed and filters can be adapted. Finally, this increases the analyst's trust in the analysis process, too.

At the moment, the system lacks of integration of data sources other than video. For instance, the grand challenge necessitated the combination of insights from different data sources (e.g., comparison of video data with badge data). This required the manual transfer of data, assumptions, and insights from different analysis tools.

Our experiences were also shared by the jury of the IEEE VAST Challenge 2009, which consisted of two visualization experts and a professional analyst. For the proposed video visual analytics approach we received the award: "outstanding video analysis tool" [5]. We also received the grand challenge award "excellent example of analytic tradecraft" [5], which honored the flexibility of our visual analytics approach that allowed for checking and searching for different events.

8 Conclusion and future directions

Following the methodology of visual analytics, we introduced a system for video analysis that is both scalable and reliable. We identified the importance of visualization and interaction techniques to connect automatic video vision with pattern recognition of humans. We covered both uncertainty-aware visualization of video and trajectory features as well as easy-to-use filter definition for relevance feedback, guiding the analyst by graphical statistics. Finally, we presented an evaluation of the proposed system by the IEEE VAST Challenge 2009 and pointed out how filter feedback and uncertainty-aware video visualization support scalability and reliability.

Future work includes consideration of the movement description of a trajectory in different granularities [35], rather than only the description of a whole trajectory. Since videos are embedded in certain context, other time-dependent data streams have to be integrated to provide a capable and useful analysis system.

Acknowledgments

We thank the reviewers for their useful comments and advice. This work was funded by German Research Foundation (DFG) as part of the Priority Program "Scalable Visual Analytics" (SPP 1335).

References

[1] ANDRIENKO, G., ANDRIENKO, N., AND WROBEL, S. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter 9*, 2 (2007), 38–46. doi:10.1145/1345448.1345455.

[2] ANDRIENKO, N., ANDRIENKO, G., AND GATALSKY, P. Exploratory spatio-temporal visualization: An analytical review. *Journal of Visual Languages and Computing 14*, 6 (2003), 503–541. doi:10.1016/S1045-926X(03)00046-6.

[3] AUVINET, E., GROSSMANN, E., ROUGIER, C., DAHMANE, M., AND MEUNIER, J. Left-luggage detection using homographies and simple heuristics. In *Proc. 9th IEEE Interna-*

tional Workshop on Performance Evaluation in Tracking and Surveillance (PETS'06) (2006), pp. 51–58.

[4] BARRON, J., FLEET, D., AND BEAUCHEMIN, S. Performance of optical flow techniques. *International Journal of Computer Vision 12*, 1 (1994), 43–77. doi:10.1007/BF01420984.

[5] BOSCH, H., HEINRICH, J., HÖFERLIN, B., HÖFERLIN, M., KOCH, S., MÜLLER, C., REINA, G., AND WÖRNER, M. Innovative filtering techniques and customized analytics tools. In *IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST'09* (2009), IEEE Computer Society, pp. 269–270. doi:10.1109/VAST.2009.5334300.

[6] BOTCHEN, R. P., BACHTHALER, S., SCHICK, F., CHEN, M., MORI, G., WEISKOPF, D., AND ERTL, T. Action-based multifield video visualization. *IEEE Transactions on Visualization and Computer Graphics 14*, 4 (2008), 885–899. doi:10.1109/TVCG.2008.40.

[7] BOUGUET, J. Pyramidal implementation of the Lucas Kanade feature tracker: Description of the algorithm. *OpenCV Documentation, Intel Corp., Microprocessor Research Labs* (2000).

[8] CASPI, Y., AXELROD, A., MATSUSHITA, Y., AND GAMLIEL, A. Dynamic stills and clip trailers. *The Visual Computer 22*, 9 (2006), 642–652. doi:10.1007/s00371-006-0046-y.

[9] CHEN, M., BOTCHEN, R., HASHIM, R., WEISKOPF, D., ERTL, T., AND THORNTON, I. Visual signatures in video visualization. *IEEE Transactions on Visualization and Computer Graphics 12*, 5 (2006), 1093–1100. doi:10.1109/TVCG.2006.194.

[10] CHENG, K., LUO, S., CHEN, B., AND CHU, H. SmartPlayer: User-centric video fast-forwarding. In *Proc. 27th International Conference on Human Factors in Computing Systems (CHI)* (2009), ACM New York, NY, USA, pp. 789–798. doi:10.1145/1518701.1518823.

[11] CHRISTEL, M. G. Supporting video library exploratory search: When storyboards are not enough. In *CIVR '08: Proc. 2008 International Conference on Content-based Image and Video Retrieval* (New York, NY, USA, 2008), ACM, pp. 447–456. doi:10.1145/1386352.1386410.

[12] CIPRA, T., AND ROMERA, R. Kalman filter with outliers and missing observations. *Test 6*, 2 (1997), 379–395. doi:10.1007/BF02564705.

[13] CORREA, C., CHAN, Y., AND MA, K. A framework for uncertainty-aware visual analytics. In *IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST'09* (2009), IEEE Computer Society, pp. 51–58. doi:10.1109/VAST.2009.5332611.

[14] DANIEL, G., AND CHEN, M. Video visualization. In *Proc. 14th IEEE Visualization 2003 (VIS'03)* (2003), IEEE Computer Society Washington, DC, USA, pp. 409–416. doi:10.1109/VISUAL.2003.1250401.

[15] DRAGICEVIC, P., RAMOS, G., BIBLIOWITCZ, J., NOWROUZEZAHRAI, D., BALAKRISHNAN, R., AND SINGH, K. Video browsing by direct manipulation. In *Proc. 26th International Conference on Human Factors in Computing Systems (CHI)* (Florence, Italy, 2008), ACM, pp. 237–246. doi:10.1145/1357054.1357096.

[16] DYKES, J., AND MOUNTAIN, D. Seeking structure in records of spatio-temporal behaviour: Visualization issues, efforts, and applications. *Computational Statistics and Data Analysis 43*, 4 (2003), 581–603. doi:10.1016/S0167-9473(02)00294-3.

[17] ELLIS, G., AND DIX, A. An explorative analysis of user evaluation studies in information visualisation. In *Proc. 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization* (2006), ACM, pp. 15–21. doi:10.1145/1168149.1168152.

[18] ENDERT, A., ANDREWS, C., FINK, G. A., AND NORTH, C. Professional analysts using a large, high-resolution display. In *IEEE Symposium on Visual Analytics Science and Technology, 2009. VAST'09* (2009), IEEE Computer Society, pp. 273–274. doi:10.1109/VAST.2009.5332485.

[19] FERGUSON, P., GURRIN, C., LEE, H., SAV, S., SMEATON, A., O'CONNOR, N., CHOI, Y.-H., AND PARK, H. Enhancing the functionality of interactive TV with content-based multimedia analysis. In *Multimedia, 2009. ISM '09. 11th IEEE International Symposium on* (2009), pp. 495 –500. doi:10.1109/ISM.2009.70.

[20] FISHER, R. The PETS04 surveillance ground-truth data sets. In *Proc. 6th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance* (2004), IEEE Computer Society, pp. 1–5.

[21] GÜTING, R., AND SCHNEIDER, M. *Moving objects databases*. Morgan Kaufmann, 2005.

[22] HÄGERSTRAND, T. What about people in regional science? *Papers in Regional Science 24*, 1 (1970), 6–21. doi:10.1007/BF01936872.

[23] HÖFERLIN, B., HÖFERLIN, M., WEISKOPF, D., AND HEIDEMANN, G. Information-based adaptive fast-forward for visual surveillance. *Multimedia Tools and Applications* (2010), in press. doi:10.1007/s11042-010-0606-z.

[24] HÖFERLIN, M., GRUNDY, E., BORGO, R., WEISKOPF, D., CHEN, M., GRIFFITHS, I. W., AND GRIFFITHS, W. Video visualization for snooker skill training. *Computer Graphics Forum 29*, 3 (2010), 1053–1062. doi:10.1111/j.1467-8659.2009.01670.x.

[25] HÖFERLIN, M., HÖFERLIN, B., AND WEISKOPF, D. Video visual analytics of tracked moving objects. In *Proc. 3rd Workshop on Behaviour Monitoring and Interpretation* (2009), vol. 541, Ghent University, Belgium, CEUR Workshop Proceedings, pp. 59–64.

[26] HONOVICH, J. Security manager's guide to video surveillance. V3. IPVideoMarket.info. [online, ebook] Available: http://ipvideomarket.info/book, 2009.

[27] HORNSBY, K., AND EGENHOFER, M. Modeling moving objects over multiple granularities. *Annals of Mathematics and Artificial Intelligence 36*, 1 (2002), 177–194. doi:10.1023/A:1015812206586.

[28] HUANG, C.-H., SHIH, M.-Y., WU, Y.-T., AND KAO, J.-H. Loitering detection using Bayesian appearance tracker and list of visitors. In *Advances in Multimedia Information Processing - PCM 2008*, vol. 5353/2008 of *Lecture Notes in Computer Science*. Springer, 2008, pp. 906–910. doi:10.1007/978-3-540-89796-5_111.

[29] JERIAN, M., PAOLINO, S., CERVELLI, F., CARRATO, S., MATTEI, A., AND GAROFANO, L. A forensic image processing environment for investigation of surveillance video. *Forensic Science International 167*, 2-3 (2007), 207–212. doi:10.1016/j.forsciint.2006.06.048.

[30] KEIM, D., MANSMANN, F., SCHNEIDEWIND, J., THOMAS, J., AND ZIEGLER, H. Visual analytics: Scope and challenges. *Lecture Notes In Computer Science 4404* (2008), 76–90. doi:10.1007/978-3-540-71080-6_6.

[31] KEIM, D., MANSMANN, F., SCHNEIDEWIND, J., AND ZIEGLER, H. Challenges in visual data analysis. In *Proc. 10th International Conference on Information Visualization* (2006), IEEE Computer Society, pp. 9–16. doi:10.1109/IV.2006.31.

[32] KRAAK, M. The space-time cube revisited from a geovisualization perspective. In *Proc. 21st International Cartographic Conference* (2003), pp. 1988–1996.

[33] KRISTENSSON, P., DAHLBÄCK, N., ANUNDI, D., BJÖRNSTAD, M., GILLBERG, H., HARALDSSON, J., MÅRTENSSON, I., NORDVALL, M., AND STÅHL, J. An evaluation of space time cube representation of spatiotemporal patterns. *IEEE Transactions on Visualization and Computer Graphics 15*, 4 (2009), 696–702. doi:10.1109/TVCG.2008.194.

[34] KUIJPERS, B., AND OTHMAN, W. Trajectory databases: Data models, uncertainty and complete query languages. In *Database Theory, ICDT 2007*, vol. 4353/2006 of *Lecture Notes in Computer Science*. Springer, 2009, pp. 224–238. doi:10.1016/j.jcss.2009.10.002.

[35] LAUBE, P., DENNIS, T., FORER, P., AND WALKER, M. Movement beyond the snapshot–dynamic analysis of geospatial lifelines. *Computers, Environment and Urban Systems 31*, 5 (2007), 481–501. doi:10.1016/j.compenvurbsys.2007.08.002.

[36] LEO, M., D ORAZIO, T., CAROPPO, A., MARTIRIGGIANO, T., AND SPAGNOLO, P. Automatic monitoring of forbidden areas to prevent illegal accesses. In *Pattern Recognition and Image Analysis*, vol. 3687/2005 of *Lecture Notes in Computer Science*. Springer, 2005, pp. 635–643. doi:10.1007/11552499_70.

[37] LEYK, S., BOESCH, R., AND WEIBEL, R. A conceptual framework for uncertainty investigation in map-based land cover change modelling. *Transactions in GIS 9*, 3 (2005), 291–322. doi:10.1111/j.1467-9671.2005.00220.x.

[38] LIU, C.-B., AND AHUJA, N. Vision based fire detection. In *Proc. 17th International Conference on Pattern Recognition (ICPR'04) Volume 4* (Washington, DC, USA, 2004), IEEE Computer Society, pp. 134–137. doi:10.1109/ICPR.2004.979.

[39] MACEACHREN, A., ROBINSON, A., HOPPER, S., GARDNER, S., MURRAY, R., GAHEGAN, M., AND HETZLER, E. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science 32*, 3 (2005), 139–161. doi:10.1559/1523040054738936.

[40] MARK, D., AND EGENHOFER, M. Geospatial lifelines. In *Integrating spatial and temporal databases. Dagstuhl Seminar Report* (1998), vol. 228.

[41] MEYER, S. *Data analysis for scientists and engineers*. John Wiley & Sons, Inc., 1975. doi:10.1021/ed054pA300.3.

[42] MILLER, H. A measurement theory for time geography. *Geographical Analysis 37*, 1 (2005), 17–45. doi:10.1111/j.1538-4632.2005.00575.x.

[43] MITAIM, S., AND KOSKO, B. What is the best shape for a fuzzy set in function approximation? In *IEEE International Conference on Fuzzy Systems* (1996), IEEE Computer Society, pp. 1237–1243. doi:10.1109/FUZZY.1996.552354.

[44] NAISBITT, J. *Megatrends: Ten new directions transforming our lives*. Warner Books New York, 1982.

[45] NILSSON, F. *Intelligent network video: Understanding modern video surveillance systems*. CRC Press. Taylor & Francis Group, 2009.

[46] PANG, A., WITTENBRINK, C., AND LODHA, S. Approaches to uncertainty visualization. *The Visual Computer 13*, 8 (1997), 370–390. doi:10.1007/s003710050111.

[47] PEKER, K., AND DIVAKARAN, A. Adaptive fast playback-based video skimming using a compressed-domain visual complexity measure. In *2004 IEEE International Conference on Multimedia and Expo, 2004. ICME'04* (2004), vol. 3, pp. 2055–2058.

[48] PETROVIC, N., JOJIC, N., AND HUANG, T. Adaptive video fast forward. *Multimedia Tools and Applications 26*, 3 (2005), 327–344. doi:10.1007/s11042-005-0895-9.

[49] PFOSER, D., AND JENSEN, C. Capturing the uncertainty of moving-object representations. In *Advances in Spatial Databases*, vol. 1651/1999 of *Lecture Notes in Computer Science*. Springer, 1999, pp. 111–131. doi:10.1007/3-540-48482-5_9.

[50] PRITCH, Y., RATOVITCH, S., HENDEL, A., AND PELEG, S. Clustered synopsis of surveillance video. In *Proc. 6th IEEE International Conference on Advanced Video and Signal Based Surveillance* (2009), IEEE Computer Society, pp. 195–200. doi:10.1109/AVSS.2009.53.

[51] SCHOEFFMANN, K., AND BOESZOERMENYI, L. Video browsing using interactive navigation summaries. *International Workshop on Content-Based Multimedia Indexing 7* (2009), 243–248. doi:10.1109/CBMI.2009.40.

[52] SHI, J., AND TOMASI, C. Good features to track. In *Proc. Computer Vision and Pattern Recognition (CVPR '94)* (Jun 1994), IEEE Computer Society, pp. 593–600. doi:10.1109/CVPR.1994.323794.

[53] SHNEIDERMAN, B. The eyes have it: A task by data type taxonomy for information visualizations. In *Proc. IEEE Symposium on Visual Languages* (1996), pp. 336–343. doi:/10.1109/VL.1996.545307.

[54] SHNEIDERMAN, B., AND PLAISANT, C. Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In *Proc. 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization* (2006), ACM, pp. 1–7. doi:10.1145/1168149.1168158.

[55] SILER, W., AND BUCKLEY, J. *Fuzzy expert systems and fuzzy reasoning*. Wiley-Blackwell, 2005.

[56] TAYLOR, B., AND KUYATT, C. Guidelines for evaluating and expressing the uncertainty of nist measurement results. NIST Technical Note 1297, United States Department of Commerce Technology Administration, National Institute of Standards and Technology, 1994.

[57] THIRDE, D., LI, L., AND FERRYMAN, F. Overview of the PETS2006 challenge. In *Proc. 9th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS 2006)* (2006), IEEE Computer Society, pp. 47–50.

[58] THOMAS, J., AND COOK, K. *Illuminating the Path: The Research and Development Agenda for Visual Analytics*. IEEE Computer Society, 2005.

[59] TRAJCEVSKI, G., WOLFSON, O., HINRICHS, K., AND CHAMBERLAIN, S. Managing uncertainty in moving objects databases. *ACM Transactions on Database Systems 29*, 3 (2004), 463–507. doi:10.1145/1016028.1016030.

[60] VAST CHALLENGE COMMITTEE. IEEE VAST Challenge 2009. IEEE Symposium on Visual Analytics Science and Technologies (VAST) 2009. [online] Available: http://hcil.cs.umd.edu/localphp/hcil/vast/index.php.

[61] WELCH, G., AND BISHOP, G. An introduction to the Kalman filter. Tech. Rep. TR 95-041, Department of Computer Science, University of North Carolina at Chapel Hill, 2004.

[62] WOLFSON, O. Moving objects information management: The database challenge (vision paper). In *Next Generation Information Technologies and Systems: Proc. 5th International Workshop*, vol. 2382/2002 of *Lecture Notes in Computer Science*. Springer, 2002, pp. 15–26. doi:10.1007/3-540-45431-4_7.

[63] WREN, C., AZARBAYEJANI, A., DARRELL, T., AND PENTLAND, A. Pfinder: Real-time tracking of the human body. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 7 (1997), 781. doi:10.1109/AFGR.1996.557243.

[64] ZADEH, L. Fuzzy sets. *Information and Control 8*, 3 (1965), 338–353. doi:10.1016/S0019-9958(65)90241-X.

Appendix: Error propagation

Given the sample position $T^{(i)}_{\text{pos}}$ of a trajectory at index i with the covariance matrix $C^{(i)}_{\text{pos}}$, we can calculate the covariance matrix $C^{(i,j)}$ of the vector $v^{(i,j)} = T^{(j)}_{\text{pos}} - T^{(i)}_{\text{pos}}$ going from at sample position $T^{(i)}_{\text{pos}}$ to $T^{(j)}_{\text{pos}}$ by

$$C^{(i,j)} = \begin{pmatrix} var_x^{(i,j)} & cov_{x,y}^{(i,j)} \\ cov_{y,x}^{(i,j)} & var_y^{(i,j)} \end{pmatrix} = C^{(i)}_{\text{pos}} + C^{(j)}_{\text{pos}} = \begin{pmatrix} var_x^{(i)} + var_x^{(j)} & cov_{x,y}^{(i)} + cov_{x,y}^{(j)} \\ cov_{y,x}^{(i)} + cov_{y,x}^{(j)} & var_y^{(i)} + var_y^{(j)} \end{pmatrix} \quad (5)$$

with the variance $var_x^{(i)}$ of $x^{(i)}$ and the covariance term $cov_{x,y}^{(i)}$ of $x^{(i)}$ and $y^{(i)}$. The length of vector $v^{(i,j)} = (x^{(i,j)}\ y^{(i,j)})^T$ is given by $\|v\| = \sqrt{(x^{(i,j)})^2 + (y^{(i,j)})^2}$. Using linear error propagation its variance $var_{\|v\|}^{(i,j)}$ is calculated by

$$var_{\|v\|}^{(i,j)} = \frac{(x^{(i,j)})^2\ var_x^{(i,j)} + (y^{(i,j)})^2\ var_y^{(i,j)} + 2x^{(i,j)}y^{(i,j)}\ cov_{x,y}^{(i,j)}}{(x^{(i,j)})^2 + (y^{(i,j)})^2} \tag{6}$$

Here, we assume only small variance of the data. Hence, we can use the linear Taylor approximation to estimate the error propagation of the quadratic equation. This approximation is also used to derive the variance and covariance terms of the squared sum of x and y that is used to calculate $var_{(x^2+y^2)}^{(i,j)} = var_{x^2}^{(i,j)} + var_{y^2}^{(i,j)} + 2\ cov_{x^2,y^2}^{(i,j)}$

Hence, the uncertainty (standard deviation) $\sigma_{\text{speed}}^{(1,N)}$ of a trajectory's speed (cf. (1)) can be expressed as

$$\sigma_{\text{speed}}^{(1,N)} = \frac{N}{fps}\sqrt{\sum_{1}^{N-1} var_{\|v\|}^{(i,i+1)}} \tag{7}$$

To derive the uncertainty $\sigma_{\text{azimuth}}^{(1,N)}$ of a trajectory's azimuth (see (2)), we first consider the normalized projected $\tilde{x}^{(1,N)}$-coordinate of vector $v^{(1,N)}$ and make use of (6):

$$\tilde{x}^{(1,N)} = \frac{v^{(1,N)}}{\left\|v^{(1,N)}\right\|}\begin{pmatrix}1\\0\end{pmatrix}$$

$$var_{\tilde{x}}^{(1,N)} = \frac{var_x^{(1)} + var_x^{(N)}}{\left\|v^{(1,N)}\right\|^2} + \frac{var_{\|v\|}^{(1,N)}(x^{(1,N)})^2}{\left\|v^{(1,N)}\right\|^4} \tag{8}$$

For error propagation of $\cos^{-1}$ we also use the linear Taylor approximation

$$\sigma_{\cos^{-1}(a)} = \left|\frac{\partial \cos^{-1}}{\partial a}\right|\sigma_a = \frac{\partial \sin^{-1}}{\partial a}\sigma_a = \frac{\sigma_a}{\sqrt{1-a^2}}$$

Hence, we can write the uncertainty $\sigma_{\text{azimuth}}^{(1,N)}$ as

$$\sigma_{\text{azimuth}}^{(1,N)} = \frac{180}{\pi}\sqrt{\frac{var_{\tilde{x}}^{(1,N)}}{1 - \frac{(x^{(1,N)})^2}{(x^{(1,N)})^2+(y^{(1,N)})^2}}} \tag{9}$$

and further rewrite it as (4).