A Lagrangian Framework for Video Analytics

Alexander Kuhn #1, Tobias Senst *2, Ivo Keller *, Thomas Sikora *, Holger Theisel #

Visual Computing Group, University of Magdeburg
Germany
1 akuhn@isg.cs.uni-magdeburg.de

* *Communication Systems Group, Technical University Berlin*
Germany
2 senst@nue.tu-berlin.de

***Abstract*—The extraction of motion patterns from image sequences based on the optical flow methodology is an important and timely topic among visual multi media applications. In this work we will present a novel framework that combines the optical flow methodology from image processing with methods developed for the Lagrangian analysis of time-dependent vector fields. The Lagrangian approach has been proven to be a valuable and powerful tool to capture the complex dynamic motion behavior within unsteady vector fields. To come up with a compact and applicable framework, this paper will provide the concepts on how to compute trajectory-based Lagrangian measures in series of optical flow fields, a set of basic measures to capture the essence of the motion behavior within the image and a compact hierarchical, feature-based description of the resulting motion features. The resulting toolbox will bee shown to be suitable for an automated image analysis as well as compact visual analysis of image sequences in its spatio-temporal context. We show its applicability for the task of motion feature description and extraction on different temporal scales, crowd motion analysis, and automated detection of abnormal events within video sequences.**

I. Introduction

One of the visual key features in video data is the representation of the underlying physical motion process. Optical flow methods, allow to capture this underlying motion behavior in terms of vector fields providing the transport information of image data between individual frames of the video data. On the one side, in computer vision there has been a large body of work on how to evaluate *local* measures based on optical flow and trajectory-based data [1], to capture and trace spatial motion information within subsequent image frames. On the other side, a physically oriented trajectory-based analysis of 2D time-dependent vector fields recently has been shown to provide valuable tool to capture the *global* dynamic structure within image motion on different temporal scales. So called particle-based or Lagrangian methods allow to describe the creation and evolution of more complex motion patterns in their spatial context together with their temporal development. In fluid flow applications especially the notion of Lagrangian Coherent Structures (LCS) [2] and their extraction using Finite Time Lyapunov Exponents (FTLE) [3] has gained much attention within the last decade. While FTLE provides a excellent starting point to gain insight into the global motion behavior within image sequences, special care has to be taken for the design and application of Lagrangian measures towards image data, as there exist fundamental differences between image-based optical flow fields and real physical flows. Hence, the goal of this paper is to provide a consistent framework to compute and evaluate Lagrangian features within image sequences, and to provide global, feature driven segmentation on variable temporal scales to be used in a variety of computer vision related tasks.

MMSP'12, September 17-19, 2012, Banff, Canada.
Germany, June 25 ©2012 IEEE.

To achieve this goal this paper provides the following contributions:

1) Generalization and application of the Lagrangian methodology towards the analysis of optical flow sequences extracted from video data
2) Introduction of a compact set of Lagrangian measures (namely trajectory arc length, direction, and separation) capturing core aspects of the geometric motion behavior within the time-dependent optical flow fields
3) Feature based segmentation based on a persistence-based topological approach to provide an hierarchical and compact representation of the Lagrangian measure results
4) Application of the presented concepts in a selected set of practical examples and computer vision tasks, including crowd movement segmentation and abnormal behavior detection

In the remainder the paper is organized as follows: Section II will provide an overview and formal definition about existing trajectory-based Lagrangian approaches, optical flow methodologies and scalar field topology using persistence. Following that, Section III will introduce the core components of the Lagrangian framework for computer vision tasks, and outline its fundamental properties and application scenarios. This includes the description of the relevant Lagrangian measures in Section III-A and the description of the post-processing in III-B. Using the previous definition Section IV will show our results on examining a selected set of video sequences under different analysis aspects. Finally, we will conclude our remarks with a discussion on limitations and future prospects of the presented framework.

II. Related Work

The Lagrangian methods we will use to define our framework have their origins in the theory of dynamical systems and have been formalized in the context of time-dependent vector field analysis, i.e. to describe fluid flow or magnetic flux fields. Most of this analysis is focused on the notion of transport barriers or so called Lagrangian Coherent Structures (LCS). LCS represent a powerful way to capture the temporal dynamics of a motion process by separating areas of coherent motion within the vector field. An basic overview about some applications has been provided by Peacock et al. [2]. The current standard to describe LCS is based on height-ridges in the Finite Time Lyapunov Exponents (FTLE) field introduced by Haller et al. [3], [4]. It measures the rate of neighboring particle trajectories within the flow field. FTLE has already been successfully deployed for the description and segmentation of crowd video footage by Ali et al. [5] and Umair et al.[6]. Despite capturing large scale motion events it has also been used to evaluate motion patterns in microscopic images of cilia organelles by Lukens et al. [7]. In close relation to this, Brox et al. [8] presented a particle-based segmentation based on geometric trajectory clustering. In general Shi et al. [9] showed, that path line clustering can be improved using multiple geometric descriptors in terms of trajectory attributes. Recently, Pobitzer et al. [10] presented a comparison on a set of basic Lagrangian measures, capturing different geometric aspects of trajectory such as helicity, arc length, and vortex measures. It is stated that a small subset attributes is sufficient to capture the most important aspects of the motion dynamics using clustering.

For most video analysis tasks segmentation of the motion information plays a central role. Ali et al. [5] used ridge information obtained from the FTLE field to provide a rough segmentation of the motion information. Since FTLE describes only motion boundaries, this has to be accomplished by artificially closing the resulting ridges using enclosed regions as segments. In addition, combining forward and backward regions [11], [12] allows to directly obtain coherent motion sets as enclosed regions. However, FTLE ridges in optical flow fields are created by a variety of different effects, and are prone to over-segmentation of complex motion fields.

Furthermore, when applying the classic definition of FTLE directly to motion flow sequences, special considerations have to be made: First, FTLE is especially designed to extract boundaries *between* areas of similar flow behavior, while for most tasks in video analysis we are interested in those similar regions itself. This can be resolved using additional path line properties (such as arc length), but requires a feature oriented description of the scalar field. One such stable, feature oriented description of scalar fields is presented by Weinkauf et al. [13]. Using the concept of scalar field topology and persistence further allows to abstract the structure of the underlying field and handle noise in a consistent, feature preserving manner. Second, most of the introduced Lagrangian descriptors have explicitly designed for the analysis of physical flows, and require fundamental properties such as area preservation of the flow field [3] or continuous flow fields. This is not necessarily the case for optical flow fields. Despite, there are still strong correspondences between motion features and the notion of FTLE. Finally, many physically inspired Lagrangian measures are tuned towards specific motion features (i.e. vortices) while some of those aspects are less interesting for video analysis. In contrast, especially changes in velocities, direction and correspondence with special spatial regions are of high importance.

III. Concept

One core aspect of Lagrangian methods hereby is the computation of the so called *flow map* defined as $\phi^{\tau}(\mathbf{x}, t_0) = \phi(\mathbf{x}, t_0, \tau)$. The flow map defines a mapping of an initial point to its advected position after a predefined integration time τ starting at t_0. Combining all position for one specific point over the interval $[t_0, t_0+\tau]$ creates a polynomial curve denoted as *path line* that describes a particle trajectory over time. For a series of 2D flow fields, this trajectory has two spatial components by means of the image position, and a temporal component that defines the transition between subsequent images. Considering both components simultaneously leads to the definition of the space-time domain, that is already known for the description of local image features [14]. In our context, local means that a certain measure considers only information from fixed region in space-time domain, i.e. a kernel with a fixed filter size in spatial or temporal direction. In contrast, global Lagrangian measures allows to compactly describe the geometric properties of path lines over an arbitrary temporal interval τ, while the spatial behavior is dictated by the optical flow fields. Formally, given a vector field $\mathbf{v}(\mathbf{x}, t)$, at every specified space-time point $(\mathbf{x}_0, t_0) \in D$ we can start a path line. This can be formulated in terms of an initial value problem:

$$\frac{d}{dt}\begin{pmatrix}\mathbf{x}\\ t\end{pmatrix} = \begin{pmatrix}\mathbf{v}(\mathbf{x}(t), t)\\ 1\end{pmatrix}, \quad \begin{pmatrix}\mathbf{x}\\ t\end{pmatrix}(0) = \begin{pmatrix}\mathbf{x}_0\\ t_0\end{pmatrix}$$

A particle trajectory or path line, that describes the transport of image information through the sequence can now be defined as:

$$\mathbf{p}(\mathbf{x}, t) = \begin{pmatrix}\mathbf{v}(\mathbf{x}, t)\\ 1\end{pmatrix} \tag{1}$$

One crucial aspect is the choice of the time interval parameter τ, that defines the temporal scale of features we are interested in. The parameter τ determines the length and complexity of Lagrangian features and how many image frames are considered. All Lagrangian measures defined on trajectories $\mathbf{p}(\mathbf{x}, t)$ can further be computed in forward and backward direction.

A. Lagrangian Measures for Computer Vision Tasks

For this framework we considered three types of Lagrangian measures $\Lambda_T(\mathbf{x}, t_0)$, i.e. path line attributes:

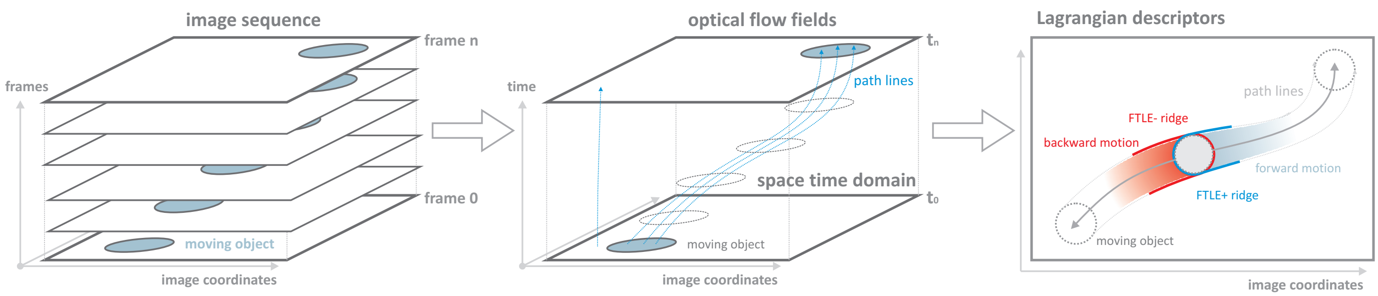

Fig. 1. Concept of Lagrangian descriptors obtained from an given image sequence using a series of optical flow fields.

- **arc length:** Accumulating all velocities along one specific trajectory allows to define regions of similar flow speed offer the interval τ.

$$\Lambda_{arcL}(\mathbf{x}, t_0) = \int ||\mathbf{v}(\phi(\mathbf{x}, t_0, \tau))||_2 \partial\tau \tag{2}$$

- **direction:** The motion direction can be obtained by accumulating average differences to a given reference direction.

$$\Lambda_{arcX/Y}(\mathbf{x}, t_0) = \zeta \cdot \int v_{x/y}(\phi(\mathbf{x}, t_0, \tau))\partial\tau \tag{3}$$

  with ζ being a normalization term

$$\zeta = \frac{min(1, ||\mathbf{v}(\phi(\mathbf{x}, t_0, \tau))||_2)}{||\mathbf{v}(\phi(\mathbf{x}, t_0, \tau))||_2 + \epsilon} \tag{4}$$

  and ϵ a small constant to avoid ζ getting singular.
- **separation:** The separation between neighboring particle trajectories is encoded in the FTLE measure [3]. The FTLE value is obtained by considering the time normalized logarithm of the spatial flow map derivatives.

Besides this predefined measures, any scalar descriptor available in the spatio-temporal context (e.g. regional- or edge information) of trajectories can be defined in an integral fashion.

B. Hierarchical Topologies on Lagrangian Measures

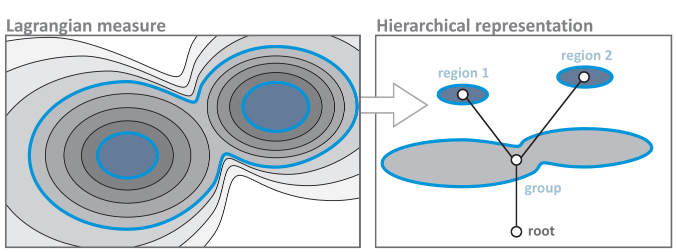

Fig. 2. Concept of contour-based hierarchical representation of scalar fields [15]

The output of all previously described Lagrangian measures is a scalar field, usually in the same resolution as the input image sequence. Lagrangian features in this scalar fields are now defined as locally extremal areas, hence minima- or maxima regions. To extract those features, we propose the use a topology oriented approach based on the extraction of iso-level contours on those fields. For this, we determine iso-levels $\widehat{\Lambda}^l_T$ within the interval $[0, max(\Lambda_T)]$ of similar motion behavior and perform a connected component analysis on those levels. As a result we obtain a contour tree similar to [15] as hierarchical representation of the Lagrangian measure. The contour tree of the scalar field is a graph $G(E, V)$ containing a set of nodes V and edges E. Each node $v \in V$ is associated to a binary blob mask $b_v(\mathbf{x})$ that describes a connected component of an iso-level contour,

$$b_v(\mathbf{x}, t_0) = \begin{cases} 1, & \text{if } \Lambda_T(\mathbf{x}, t_0) \leq \widehat{\Lambda}^l_T \\ 0, & \text{else} \end{cases} \tag{5}$$

and the iso-level value itself $\widehat{\Lambda}^l_{T,v}$.

Note, that this can be modified to any given input dimension. In order to avoid over-segmentation or inconsistent feature introduced by noise, we apply a persistence-based simplification of the resulting hierarchy [13], that usually significantly reduces the number of salient features.

Having obtained the Lagrangian scalar fields and hierarchical descriptors, offers a variety of post processing options to be tuned towards the respective analysis tasks. In a general analysis setting we can differentiate two major way to deal with the resulting output:

First, we can directly use the output for a direct and interactive visual analysis. This allows for a compressed representation of motion features in space time domain, and even to identify specific motion features within longer video sequences without considering the complete sequence.

Second, we can construct automated analysis approaches based on the resulting representations. This way we can efficiently describe repeating patterns of motion or detect irregularities in the corresponding representation.

IV. Results and Applications

In order to demonstrate the practical usability of our framework we applied the presented concepts to set of basic test data sets.

A. Crowd Segmentation

As described in the previous Section the parameter τ allows to describe motion events in the video on different temporal scales. Figure 3 illustrates the effect of increasing τ on a traffic sequence. While small tau values allow to capture individual cars, higher values capture the notion of lanes. Further the direction measure directly delivers clearly separated clusters in terms of groups moving in opposite directions. The Integration of multiple frames allows for a temporal smoothing of salient motion information over time, as single optical flow fields tend to contain significantly more noise. In our experiments we found, that using smaller values for τ compactly describes motion features over multiple frames while keeping small but salient features without additional smoothing. In addition to this, larger temporal scales allow to group together objects of similar motion characteristics, as they blend together in the respective Lagrangian field as illustrated in Figure 3, last image.

Figure 4 shows the result for visualizing the resulting Lagrangian scalar fields on $\tau = 10$ for the direction measure on a pedestrian crossing sequence. Looking at the resulting field in space time domain gives a clear notion on the temporal relation of single motion events and group motion over the respective video sequence. In addition to this, the presentation in space time allows to visually compress the motion information of the video sequence into one image.

An example for combining different Lagrangian measures to perform task based segmentations is presented in Figure 5. Using the arc length measure allows to identify the fastest person in the marathon sequence based on their image location in 5 a) and temporal occurrence in the sequence in 5 d). Figure 5 b) shows the image blended together with color coded directions, which emphasizes portions of the image moving against the major flow direction. Again, the illustration in space time in 5 e) allows to clarify, at which time the associated events have occurred.

B. Abnormal Event Detection

Besides segmentation and visual analysis, our frame work can be used to automate the detection of salient features in image sequences. The contour tree that contains the topology of the time-normalized arc length $\Lambda_{arcL}(\mathbf{x}, t_0)$ contains the structured information about the undirected motion components of a short video sequence. We used the mean of the surface integrals:

$$\mu_{arcL}(G) = \frac{1}{|V|}\sum_{V}\left(\sum_{\mathbf{x}}\widehat{\Lambda}^{l}_{arcL,v} \cdot b_v(\mathbf{x}, t_0)\right) \qquad (6)$$

This setting is tested on the UMN dataset to detect abnormal events. The dataset contains 11 different scenarios of an escape event in 3 different indoor and outdoor scenes.

Figure 6 shows the results of the experiments, where the response $\mu_{arcL}(G)$ of is used in combination with a adaptive threshold to detected frames containing abnormal motion behavior. A Gaussian model is used to learn the initial part of normal behavior and is updated online. Abnormal frames correspond to outliers that are detected by the Mahalanobis distance.

The results are produced using integration interval $\tau = 10$, 15 levels and a persistence measure of 2. The overall results shows that the proposed framework is capable to detect each abnormal event annotated in the database without having a

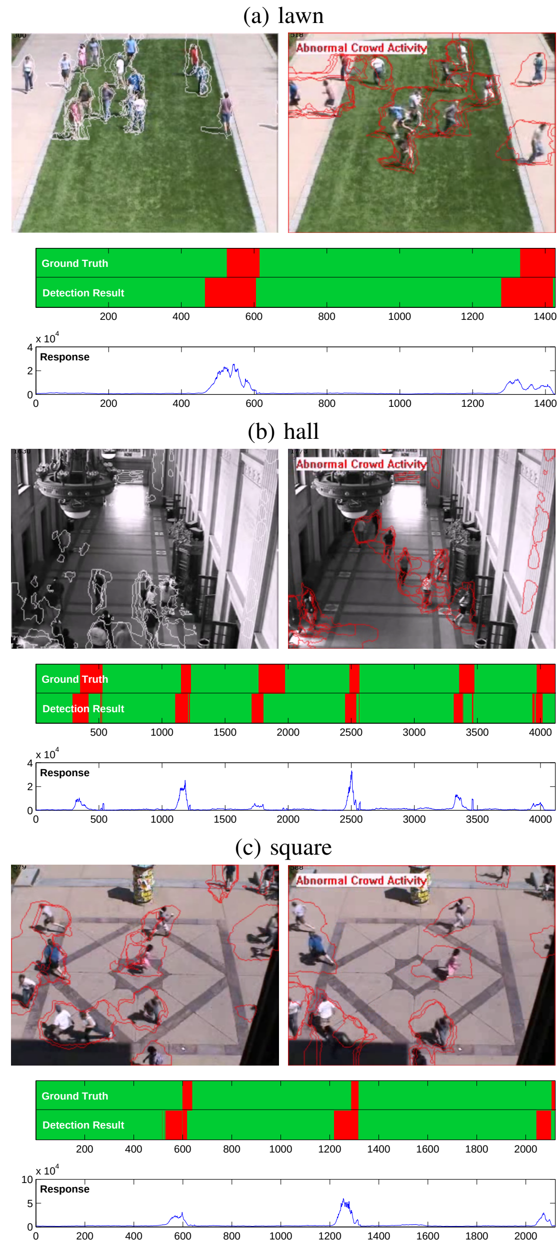

Fig. 6. Illustration of the results for abnormal event detection for the three scenarios of the UMN dataset. The ground truth bar and detection bar shows the labels of each frame, green represents normal frames and red abnormal frames and is presented in conjunction with the response $\mu_{arcL}(G)$ of the arc lenght multi layer tree. The contours of the tree nodes are shown for normal and abnormal behavior.

hugh amount of false positives. Generally the detections of the proposed systems are longer and begin at an earlier time than the annotated ground truth. On reason is found by the path line integration were motion information of future frames are taken into account to estimate the current scalar field. Another reason is found in the annotation itself as shown in Figure 6(c) the abnormal activity in many cases labeled very late, after the person starting to run which explains the delay as the proposed system reacts immediately to motion changes.

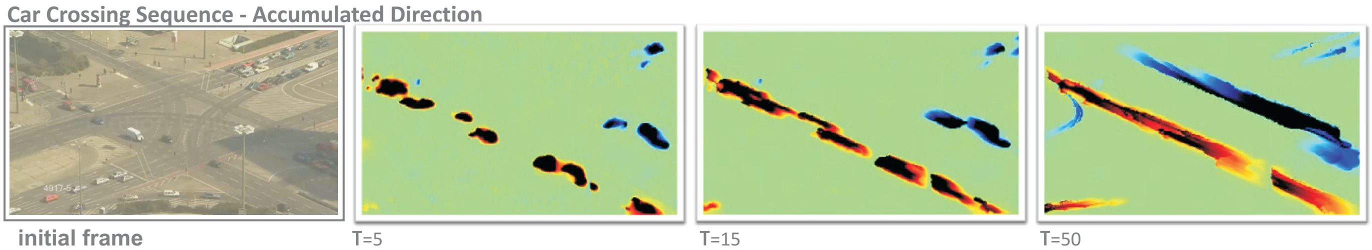

Fig. 3. Direction measure applied to a car crossing sequence. Increasing τ allows to describe motion features on different temporal scales. While individual cars can be captured on small τ scales, large values capture the notions of lanes, that become clearly segmented by the direction measure (red denotes large positive motion in x-direction, blue negative movement in x-direction).

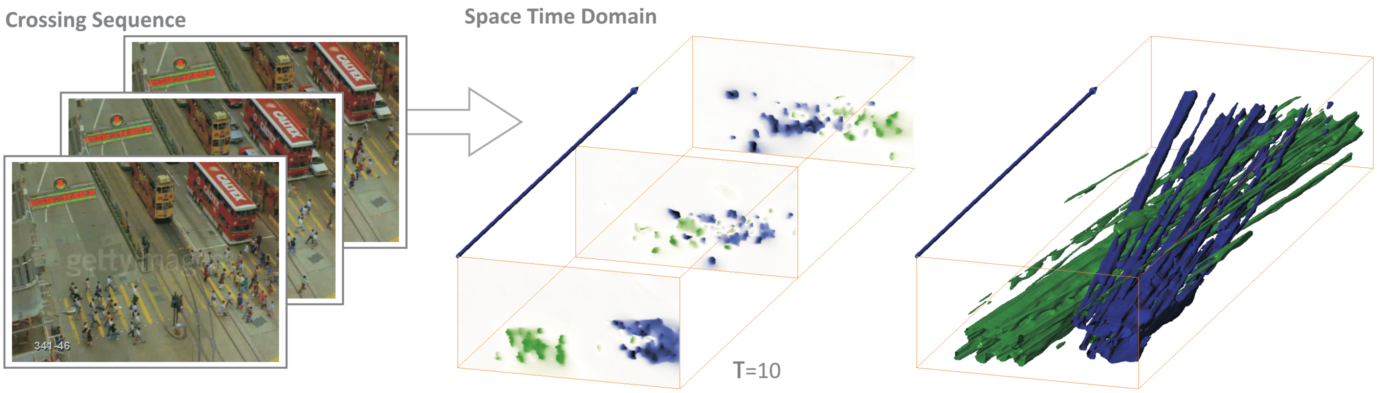

Fig. 4. Using the direction measure and a space time view to capture time-dependent crowd dynamics in a junction sequence for a time interval of $\tau = 10$.

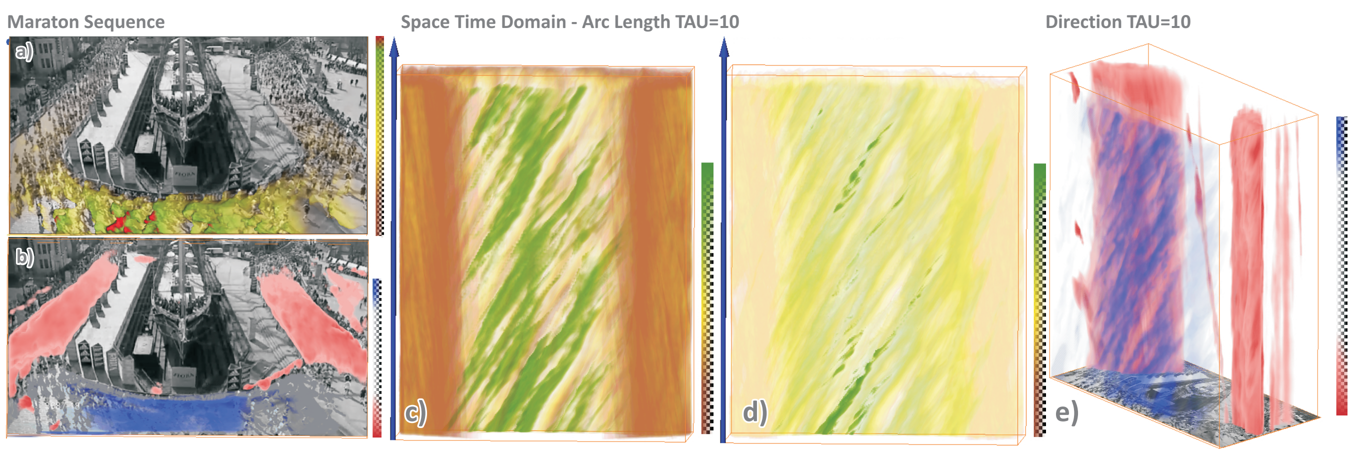

Fig. 5. Blending of Lagrangian fields with initial frame to highlight important regions. Figure a) highlights the fastest group and person in the sequence, while Figure b) emphasizes all objects moving against the major flow direction. Further, the space time views in c) to e) reveal the temporal relation of those events (the blue arrow denotes temporal direction).

V. Conclusion

In this work we presented the concept of a Lagrangian methodology framework for computer vision tasks. We proposed the conceptual basics in order to apply Lagrangian methods to a series of optical flow fields, and proposed basic measures such as arc length, direction and separation to analyze time-dependent motion behavior of image sequences. For post-processing and evaluation of motion features we used an hierarchical contour-based representation, that allows an feature related abstract evaluation and combination. We applied the resulting framework for crowd segmentation and abnormal behavior detection in video sequences.

One limitation to obtain reliable results of the Lagrangian analysis is the quality of the underlying optical flow fields sequence. Although Lagrangian features can be extracted using faster, less accurate optical flow methods, artifacts will also be present in the resulting Lagrangian description. Further, Lagrangian methods are only suited to describe low-level motion features, while e.g. occlusion cannot be directly expressed.

Acknowledgment

The research leading to these results has received funding from the European Communitys FP7 under grant agreement number 261743 (NoE VideoSense). The project SemSeg acknowledges the financial support of the Future and Emerging Technologies (FET) programme within the Seventh Framework Programme for Research of the European Commission, under FET-Open grant number 226042.

References

[1] G. Brostow and R. Cipolla, "Unsupervised bayesian detection of independent motion in crowds," in *Computer Vision and Pattern Recognition (CVPR 2006)*, 2006, pp. 594–601.

[2] T. Peacock and J. Dabiri, "Introduction to focus issue: Lagrangian coherent structures," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 1, p. 017501, 2010. [Online]. Available: http://link.aip.org/link/?CHA/20/017501/1

[3] G. Haller, "Lagrangian structures and the rate of strain in a partition of two-dimensional turbulence," *Physics of Fluids*, vol. 13, no. 11, 2001. [Online]. Available: http://link.aip.org/link/?PHFLE6/13/3365/1

[4] ——, "A variational theory of hyperbolic Lagrangian Coherent Structures," *Physica D*, vol. 240, pp. 574–598, 2010. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167278910003143

[5] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *Computer Vision and Pattern Recognition (CVPR 2007)*, june 2007, pp. 1 –6.

[6] S. Umair and A. M. Rizal, "Underwater crowd flow detection using lagrangian dynamics," *Indian Journal of Geo-Marine Sciences (IJMS 2009)*, vol. 38, no. 3, pp. 359–364, September 2009.

[7] S. Lukens, X. Yang, and L. Fauci, "Using lagrangian coherent structures to analyze fluid mixing by cilia," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 20, no. 1, p. 017511, 2010. [Online]. Available: http://link.aip.org/link/?CHA/20/017511/1

[8] T. Brox and J. Malik, "Object segmentation by long term analysis of point trajectories," in *Proceedings of the 11th European conference on Computer vision: Part V*, ser. ECCV'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 282–295. [Online]. Available: http://dl.acm.org/citation.cfm?id=1888150.1888173

[9] K. Shi, H. Theisel, H. Hauser, T. Weinkauf, K. M. H.-C. Hege, and H.-P. Seidel, "Path line attributes - an information visualization approach to analyzing the dynamic behavior of 3d time-dependent flow fields," in *Proc. Topo-In-Vis 2007*, Kloster Nimbschen bei Leipzig, 04.-06. Mrz 2007 2007.

[10] A. Pobitzer, A. Lez, K. Matkovic, and H. Hauser, "A statistics-based dimension reduction of the space of path lineattributes for interactive visual flow analysis," in *Proceedings of the IEEE Pacific Visualization Symposium (PacificVis 2012)*, March 2012, pp. 113–120.

[11] C. Garth, G. Li, X. Tricoche, C. Hansen, and H. Hagen, "Visualization of coherent structures in transient 2d flows," *Topology-Based Methods in Visualization II*, pp. 1–13, 2009. [Online]. Available: http://www.springerlink.com/index/w3l12813212g7374.pdf

[12] S. Bachthaler, F. Sadlo, C. Dachsbacher, and D. Weiskopf, "Space-time visualization of dynamics in lagrangian coherent structures of time-dependent 2d vector fields," *International Conference on Information Visualization Theory and Applications*, pp. 573–583, 2012.

[13] T. Weinkauf, Y. Gingold, and O. Sorkine, "Topology-based smoothing of 2D scalar fields with c1-continuity," *Computer Graphics Forum (Proc. EuroVis)*, vol. 29, no. 3, pp. 1221–1230, June 2010. [Online]. Available: http://tinoweinkauf.net/

[14] R. Poppe, "A survey on vision-based human action recognition," *Image and Vision Computing*, vol. 28, no. 6, pp. 976 – 990, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885609002704

[15] H. Carr, J. Snoeyink, and U. Axen, "Computing contour trees in all dimensions," in *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, ser. SODA '00. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2000, pp. 918–926. [Online]. Available: http://dl.acm.org/citation.cfm?id=338219.338659