EVENT DETECTION IN CROWDS OF PEOPLE BY INTEGRATING CHAOS AND LAGRANGIAN PARTICLE DYNAMICS

C. SPAMPINATO, A. FARO, S. PALAZZO
Department Of Electrical, Electronics And Computer Engineering
University of Catania - Viale Andrea Doria, 6 - 95127 Catania, Italy
Email: {cspampin, afaro, palazzosim}@dieei.unict.it

ABSTRACT
This paper proposes a system for automatic video analysis to detect events in video sequences with crowds of people. In detail, the proposed system consists of three subsystems: 1) the first identifies the motion areas, resorting to chaos theory using joint histogram between consecutive frames, 2) the second one creates a flow motion map that describes the behavior of motion pixels by using Lagrangian Particle Dynamics Theory and, 3) the last one uses self organizing maps (SOM) for segmenting the flow motion map in order to detect events.
The proposed method was tested on a set of 30 videos, describing crowds in different scenarios, collected from the BBC Motion Gallery, achieving an average accuracy of about 87% in detecting events such as people stopping, people laying on the ground, group of people fighting, obstacles on the road and long queues.

KEY WORDS
Intelligent Video Analysis, Optical Flow, Object Detection and Tracking

1 INTRODUCTION

Huge amounts of video streams are daily recorded for surveillance purposes. There is a great interest in automatic video analysis for event detection since manual mining of video streams is a very tedious and time consuming task.
However, the problem is still far from being completely solved and current solutions show many drawbacks, especially if compared with the human vision system, which has a remarkable ability to detect interesting events in very changing and complex environments such as crowds of people. Several approaches for event detection have been recently proposed in the literature (e.g. [1]), which can be classified as based on: tracking, flow models, spatio-temporal analysis of shapes and interesting points analysis. The tracking based methods, such as the ones proposed in [2], [3], process the entire video sequence and extract objects by comparing the current frame and a background that represents the scene without objects of interest. Flow based models, instead, process directly the spatio-temporal sequence (e.g. optical flow) without performing any segmentation. In detail, events of interest are detected, for instance, by correlating flow maps with actions, as proposed by Efros et al. in [4] where the optical flow map is treated not as a pixel displacements map, but as a spatial pattern related to human actions. Spatio-temporal analysis of shapes methods, such as the one proposed in [5], usually, analyze the spatio-temporal volume of a videosequence considering it as a 3D object, i.e. they use a model (called

motion history volumes) that fuses action cues seen from different viewpoints and in fixed time periods into a three dimensional representation. Ke et al. in [6] proposed a hybrid flow-shape approach by considering an event template not as an atomic entity, but composed by different parts both in space and in time. Recently, methods based on analysis of interesting points have gained an increasing interest. The idea behind these methods is to identify interesting events by processing space-time interest points. For example, in [7] a video sequence is represented as a collection of space-time interest points extracted from space-time shapes and features such as local saliency [8]. Then, the algorithm calculates the probability distributions of these spatial-temporal points and associates them to human actions. All these approaches lack when dealing with complex and crowded scenes, e.g. sports events with thousands of people, since they are not able to build models (e.g. interesting points, trajectories or motion history volumes) that describe the behavior of each person in the scene. Moreover, the above approaches cannot be applied in most cases when computational resources are scarce, since the motion detection phase is relatively heavy in terms of processing time. In this paper we propose a novel system for event detection in crowds of people that overcomes these limitations where motion pixels are identified by exploiting a chaos based approach [9] and the people behavior in the crowd is analyzed by the flow segmentation method proposed in [10].

The remainder of the paper is as follows: Section 2 gives an overview of the proposed system; in Section 3 the model for motion map estimation based on chaos theory is presented. Sect. 4 shows the system for people flow segmentation and suspicious event detection by integrating Lagrangian particle dynamics theory with Self Organizing Maps. Finally, experimental results and concluding remarks are, respectively, reported in Section 5 and 6.

2 THE PROPOSED SYSTEM

In this work, we propose an unsupervised approach (whose flowchart is shown in fig. 1) for event detection using chaos-like behavior of motion pixels for understanding which part of the scene is affected by motion, and identifying, among the motion pixels, those ones that behave differently from the rest by computing Finite-Time Lyapunov Exponent (FTLE) map [11].

In detail, the motion detection system relies on the assumption that the trajectories of motion pixels in phase space show a chaos-like behavior [9]; therefore, by exploring pixels' trajectories in consecutive frames, it is possible to extract a motion map. At the same time, the system is provided with an algorithm that estimates pixels' behavior in the videosequence by analyzing the ridges in a Finite-Time Lyapunov Exponent (FTLE) map. This FTLE map is then passed to a self-organizing map (SOM) to segment the input frame according to the different detected pixels' behaviors.

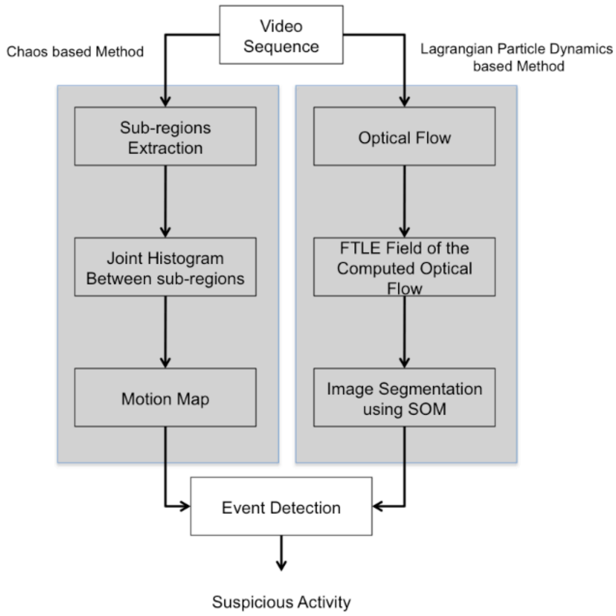

Fig. 1. Flow diagram of the proposed event detection system.

Finally, events of interest are detected by integrated the motion map and the aforementioned segmented image. In the following each one of the above subsystems is described in detail.

3 CHAOS THEORY FOR MOTION MAP ESTIMATION

The first subsystem aims at identifying motion areas by means of the chaos theory. Unlike standard motion detection approaches [12], [13], [14], [15], our method estimates differences between pixel amplitudes of two images I_{t0} and I_{t1} by assessing the mutual information $H(I_{t0}; I_{t1})$ between the images, defined as:

$$H(I_{t0}, I_{t1}) = \sum_{i \in I_{t0}, j \in I_{t1},} p(i,j) \cdot \frac{p(i,j)}{p(i) \cdot p(j)}$$

where $p(i)$ and $p(j)$ are, respectively, the distributions of images I_{t0} and I_{t1} and $p(i,j)$ is the joint distribution of the two images. The joint distribution can be approximated using the joint histogram as well as the distributions of images can be approximated using their grey level histograms. It has been proved in [16] that the result of motion in image sequences is a complex trajectory of joint histograms (see fig. 2), whereas the result of illumination changes is a linear trajectory as shown in Fig. 3. Therefore, motion in video sequences results in a chaotic behavior in joint histograms due to a non-linear multiplicative process in the reflectance [16] and we exploit this behavior to build our motion detection system, robust against illumination changes. In detail, our approach divides the

input frame in subregions, and for each sub-region the joint histogram with the corresponding sub-region of the previous frame is computed.
The sub-images that show a chaotic behavior are then considered as regions containing moving people, otherwise not. Finally, a motion map is computed by combining all the subregions that are affected by motion, as shown in Fig. 4.
The motion map indicates the areas with motion and represents the input for the next subsystem that aims at analyzing the flow of people, in order to understand their behavior and to detect anomalous events.

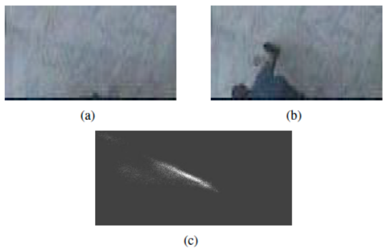

Fig. 2. Chaotic behavior due to a moving object. a) Image at time $\mathbf{t_0}$, b) Image at time $\mathbf{t_1}$ and c) Joint histogram between image **t0** and image $\mathbf{t_1}$.

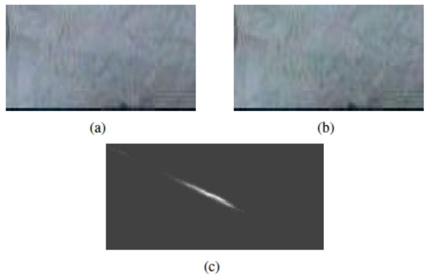

Fig. 3. Linear behavior due to illumination changes. a) Image at time $\mathbf{t_0}$, b) Image at time $\mathbf{t_1}$ and c) Joint histogram between image **t0** and image $\mathbf{t_1}$.

(a) (b) (c)

Fig. 4. a) Subregions of frame at time t_0, b) Subregions of frame at time t_1 and c) Motion map between frame t_0 and frame t_1.

4 IMAGE SEGMENTATION USING LAGRANGE PARTICLE DYNAMICS THEORY

This subsystem relies on the notion of coherent structure (CS) [17], which is generally used both in turbulence theory and in 2D or 3D fluid mechanics, and basically, it segments the input frames by exploring how clouds of particles mix together and how they are transported under the action of a flow field (computed by optical flow) generated by the crowd motion. The rationale is that the trajectories generated from the advection of particles through a flow are representative of cases such as people stopping, locations of the barriers, anomalous movements, etc. In this work, the study of particles' trajectory under the action of a flow is performed by the identification of the coherent structures, which are separatrices that split the flow of particles into dynamically distinct

regions where all particles show a similar (coherent) behavior.
One of the most common method to calculate the coherence structure is the Finite-Time Lyapunov Exponent (FTLE) [18] that describes different behaviors of the pixels (in terms of flow in a scene) that cannot be seen using the velocity vector field provided by the optical flow. The description on how to compute the FTLE is behind the aim of the paper and it can be found in [10]. In order to identify variations in the behavior of crowds of people (i.e. identifying LCS), the proposed system, initially, estimates the optical flow according to [19], as a functional that integrates three motion features: the brightness variation, the gradient constant variation and a discontinuity-preserving spatial-temporal smoothness constraint. Afterwards, it identifies the FTLE field of the computed optical flow, then it divides the flow into regions with different dynamics by applying a self-organizing map (SOM) [20] to the FTLE field, segmenting the input frame according to the motion pixels' behavior. Each identified region (color-coded in Fig. 5-e) in the segmented image is a coherence structure.
An example on how the algorithm performs is shown in Fig. 5, where an interesting event is depicted in red; this event is detected by comparing the motion map with the segmented image and it represents the only coherence structure that is entirely included in the area outside the motion map.
More in general, we detect anomalous events by estimating the number of coherence structures entirely included either within or outside the motion map. For example, a coherence structure inside the motion map indicates a different behavior of some particles in an area with people moving, e.g. people lying on the ground; whereas a coherence structure outside the motion map may indicate events when most of the people are stationary, but some of them suddenly move, e.g. people fighting.

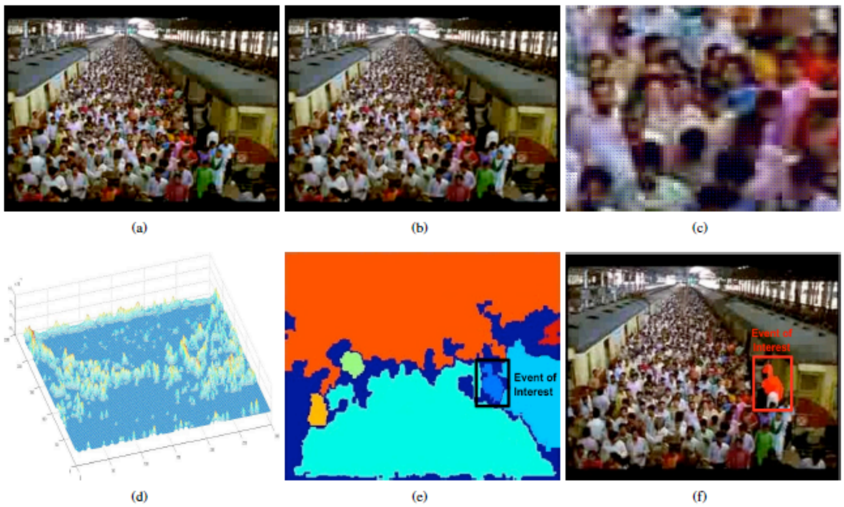

Fig. 5. FTLE estimation. a) Image at time t0, b) Image at time t1, c) Optical flow superimposed on the Image at time t1, d) FTLE field, e) Image segmented by the SOM according to the particles' behavior shown in the FTLE field and f) Interesting event detected

5 EXPERIMENTAL RESULTS

The proposed event detection system was tested on a set of 30 videos collected from the BBC Motion Gallery. The videos had a spatial resolution of 320x240 with 30 fps. The main target was crowds of people in different locations such as train station, airport, stadium and subway station. The ground truth was hand labeled by using ViPER (the Video Performance Evaluation Resource [21]) and consisted of 859 events, categorized as: 1) People Stopping (PS), 2) People lying on the ground (PL), 3) Group of People Fighting (PF), 4) Obstacles (O) on the road (road working or cleaning), 5) Long Queues (Q). Table I shows the obtained results in terms of correct detections over the test set. It has to be noticed that the proposed system does not recognize the events but it is only able to detect them; this means that it was tested on each category of event and the number of detected coherence structures is computed.

Tab.1 Experimental results

Type of Event	Total Number	Detected Number	Percentage of Success
PS	296	263	88.85 %
PL	92	76	82.60%
PF	73	64	87.60%
O	157	136	86.62%
Q	241	217	90.04%

As a further example, fig. 6 shows the results of processing a scene recorded at La Mecca where there are people walking (most of them) in an ordered flow and some people praying and stopping, located more or less at the middle of the image.

In this particular scenario, the proposed system is able to detect the different behavior of praying people (red area in fig. 6-b) with respect the other walking people.

(a) (b)

Fig. 6. Event Detection. a) Input Image, b) Event detected: people praying

6 CONCLUDING REMARKS

In this paper we have proposed an unsupervised system, applicable to real-world settings, for identifying events in videos with crowds of people by using joint histograms between consecutive frames, derived from the chaos theory, to detect motion areas and then using Lagrangian particle dynamics theory combined with self organizing maps for image segmentation according to people flow behavior. Unlike existing methods, our system processes very quickly the input frames in order to estimate the motion areas, without requiring background/objects separation. Our system shows promising performance on real-world videos (about 87%) in detecting events in complex scenes with highly cluttered dynamic backgrounds. As future work, we plan to provide our system with a denoising module [22], [23] to remove the noise affecting the grabbed frames, since the motion detection has been proved to be robust against noise, whereas the performance of the flow map estimation module might be affected by noise. Moreover novel clustering approaches such as the ones proposed in [24] for investigating improved performance in the segmentation of the FTLE map and also we will use GRID based approaches [25] to increase the efficiency in the FTLE computing. We also aim at adding a further layer for automatic event recognition and at integrating this system in a more complex framework able to track [26] and recognize people [27] (more in general objects and their behavior [28]) involved in the detected events. We are also aiming at integrating the proposed system in a distributed environment where software agents [29] detect and integrate the events, thus supporting event detection on multiple cameras.

REFERENCES

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," Computer Vision and Image Understanding, vol. 73, pp. 428–440, 1999.

[2] S. S. Intille, A. F, and Bobick, "Recognizing planned, multiperson action," Computer Vision and Image Understanding, vol. 81, pp. 414–445, 2001.

[3] L. M. Fuentes and S. A. Velastin, "People tracking in surveillance applications," Image and Vision Computing, vol. 24, no. 11, pp. 1165 –1171, 2006, performance Evaluation of Tracking and Surveillance.

[4] A. A. Efros, A. C. Berg, G. Mori, and J. Malik, "Recognizing action at a distance," in ICCV '03: Proceedings of the Ninth IEEE International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 2003, p. 726.

[5] D. Weinland, R. Ronfard, and E. Boyer, "Free viewpoint action recognition using motion history volumes," Comput. Vis. Image Underst., vol. 104, no. 2, pp. 249–257, 2006.

[6] Y. Ke, R. Sukthankar, and M. Hebert, "Event detection in crowded videos," in IEEE International Conference on Computer Vision, October 2007.

[7] J. C. Niebles, H. Wang, and L. Fei-Fei, "Unsupervised learning of human action categories using spatial-temporal words," Int. J. Comput. Vision, vol. 79, no. 3, pp. 299–318, 2008.

[8] M. Blank, L. Gorelick, E. Shechtman, M. Irani, and R. Basri, "Actions as space-time shapes," Computer Vision, IEEE International Conference on, vol. 2, pp. 1395–1402, 2005.

[9] M. E. Farmer, "A chaos theoretic analysis of motion and illumination in video sequences," Journal of Multimedia, vol. 2, no. 2, pp. 53–64, 2007.

[10] S. Ali and M. Shah, "A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 0, pp. 1–6, 2007.

[11] F. Shadden, S.C.and Lekien and J. E. Marsden, "Definition and properties of Lagrangian coherent structures from finite-time Lyapunov exponents in two-dimensional aperiodic flows," Phys. D, vol. 212, no. 3-4, pp. 271–304, 2005.

[12] A. Faro, D. Giordano, and C. Spampinato, "Evaluation of the Traffic Parameters in a Metropolitan Area by Fusing Visual Perceptions and CNN Processing of Webcam Images," IEEE Transactions on Neural Networks, vol. 19, no. 6, pp. 1108–1129, 2008.

[13] C. Spampinato, Y. H. Chen-Burger, G. Nadarajan, and R. B. Fisher, "Detecting, Tracking and Counting Fish in Low Quality Unconstrained Underwater Videos," in 3rd International Conference on Computer Vision Theory and Applications (VISAPP'08), 2008, pp. 514–519.

[14] A. Faro, D. Giordano, and C.Spampinato, "Soft-computing agents processing webcam images to optimize metropolitan traffic systems," Computational Imaging and Vision Computational Imaging and Vision - Computer Vision and Graphics, vol. 32, pp. 968–974, 2006.

[15] A. Faro, D. Giordano, and C. Spampinato, "Integrating location tracking, traffic monitoring and semantics in a layered its architecture," IET Intelligent Transport Systems, vol. 5, no. 3, pp. 197–206, 2011.

[16] Y. Xu and A. Roy Chowdhury, "The joint illumination and motion space of video sequences," in ICIP05, 2005, pp. II: 1310–1313.

[17] G. Haller and G. Yuan, "Lagrangian coherent structures and mixing in two-dimensional turbulence," Phys. D, vol. 147, no. 3-4, pp. 352–370, 2000.

[18] X. Z. Tang and A. H. Boozer, "Finite time Lyapunov exponent and advection-diffusion equation," Phys. D, vol. 95, no. 3-4, pp. 283–305, 1996.

[19] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in ECCV, vol. 4, 2004, pp. 25–36.

[20] A. Ultsch, "Clustering with SOM: U*C," in Proc. Workshop on Self-Organizing Maps, 2005, pp. 75–82.

[21] D. Doermann and D. Mihalcik, "Tools and techniques for video performance evaluation," in Proceedings of the 15th International Conference on Pattern Recognition, vol. 4, 2000, pp. 167–170 vol.4.

[22] A. Faro, D. Giordano, G. Scarciofalo, and C. Spampinato, "Bayesian networks for edge preserving salt and pepper image denoising," in IPTA08, 2008, pp. 1–5.

[23] F. Cannavo, G. Nunnari, D. Giordano, and C. Spampinato, "Variational method for image denoising by distributed genetic algorithms on GRID environment," in Proceedings of the 15th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises. Washington, DC, USA: IEEE Computer Society, 2006, pp. 227–232.

[24] A. Faro, D. Giordano, and F. Maiorana, "Mining massive datasets by an unsupervised parallel clustering on a grid: Novel algorithms and case study," Future Gener. Comput. Syst., vol. 27, pp. 711–724, June 2011.

[25] A. Crisafi, D. Giordano, and C. Spampinato, "Griplab 1.0: Grid image processing laboratory for distributed machine vision applications," in WETICE, 2008, pp. 188–191.

[26] C. Spampinato, "Adaptive objects tracking by using statistical features shape modeling and histogram analysis," in Seventh International Conference on Advances in Pattern Recognition, 2009. ICAPR '09., Feb. 2009, pp. 270–273.

[27] A. Faro, D. Giordano, and C. Spampinato, "An automated tool for face recognition using visual attention and active shape models analysis," Conf Proc IEEE Eng Med Biol Soc, vol. 1, pp. 4848–4852, 2006.

[28] C. Spampinato, D. Giordano, R. Di Salvo, Y.-H. J. Chen-Burger, R. B. Fisher, and G. Nadarajan, "Automatic fish classification for underwater species behavior understanding," in Proceedings of ARTEMIS '10, ACM, 2010, pp. 45–50.

[29] A. Faro, D. Giordano, and A. Musarra, "Ontology based intelligent mobility systems," in Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, vol. 5, 2003, pp. 4288+.4848–4852, 2006.