Electronic Letters on Computer Vision and Image Analysis 13(2):15-16, 2014

Detecting and indexing moving objects for Behavior Analysis by Video and Audio Interpretation

Alessia Saggese [*+]

[*] *Dept. of Information Eng., Electrical Eng. and Applied Mathematics, University of Salerno, Salerno, Italy*

[+] *GREYC UMR CNRS 6072 ENSICAEN, Universite' de Caen Basse-Normandie, Caen, France*

Advisors: Mario Vento[*] and Luc Brun[+]

24 February 2014, University of Salerno

Received 29 Jan 2014; accepted 25th May 2014

1 Abstract

In the last decades we have assisted to a growing need for security in many public environments. This consideration has lead the proliferation of cameras and microphones, which represent a suitable solution for their relative low cost of maintenance, the possibility of installing them virtually everywhere and, finally, the capability of analyzing more complex events. However, the main limitation of traditional *audio-video surveillance* systems lies in the so called *psychological overcharge issue* of the human operators responsible for security, that causes a decrease in their capabilities to analyze raw data flows from multiple sources of multimedia information. For the above mentioned reasons, it would be really useful to design an *intelligent* surveillance system, able to provide images and video with a semantic interpretation, for trying to bridge the gap between their low-level representation in terms of pixels, and the high-level, natural language description that a human would give about them. The aim of this thesis [11] is to face the above mentioned issues, as fascinating as challenging. The proposed system analyzes videos and by extracting trajectories of objects populating the scene (*tracking*): trajectory is a very discriminant feature, since the movement of objects in a scene is not random, but instead have an underlying structure which can be exploited to build some models. The main novelties of the proposed tracking algorithm lie in the following aspects: first, the entire history of each object populating the scene is analyzed by means of a Finite State Automaton; second, the update of information related to each object is performed by a graph-based approach. Finally, occlusions are properly managed by tracking into a different way single objects and groups of objects [10][7][8]. The proposed tracking algorithm has been evaluated during an international competition (PETS 2013), ranking in the first places for all the considered scores over an high number of participants (more than thirty). Once extracted, this large amount of trajectories needs to be indexed and properly stored in order to improve the overall performance of the system during the retrieving step [9]: the main novelty of this module pertains the enhancement of off-the-shelf solutions, namely PostGis (the spatial extension of the traditional PostGres database) in order to deal with trajectories, which are very complex elements to manage because of their spatio-temporal nature. In general, the main advantage of the proposed approach lies in the fact that a human operator can interact with the system in different ways: first of all, he is informed by the system as soon as an *abnormal* behavior occurs. It is evident that the system has to be robust enough to deal with errors typically occurring during the tracking phase, related for instance to broken

Correspondence to: <asaggese@unisa.it>

Recommended for acceptance by <Alicia Fornes and Volkmar Frinken>
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autonoma de Barcelona, Barcelona, Spain

trajectories. Furthermore, an high level of generalization is required in order to avoid wrongly classifying as abnormal those normal trajectories which only rarely occur. In order to cope with the above mentioned issues, the similarity between trajectories is evaluated by means of a novel metric based on string kernel, especially defined for this purpose [2]. Whereas the information extracted from the videos are not sufficient or not sufficiently reliable, the proposed system is enriched by a module based on a bag of *aural* words approach, in charge of recognizing audio events such as shoots, screams or broken glasses [4][5]. In the proposed framework a human operator can also extract typical paths occurring inside a scene without any knowledge about the particular scenario (*clustering*): each path can represent normal behaviors, related to both people moving into a train station and vehicles crossing an highway [1]. The proposed clustering algorithm, based on a tree structure, exploits the kernel-based similarity metric defined above in order to perform its task. On the other hand, the human operator can ask different typologies of queries to the proposed framework by personalizing the parameters only at query time, that is in the moment the query is thought: for instance, the objects crossing a given area in a given time interval (*dynamic spatio temporal query*) [6] or the objects following a particular trajectory, similar to the one hand drawn by the user (*query by sketch*) [3]. Each proposed module has been tested both over standard datasets and in real environments; the promising obtained results confirm the insight with respect to the state of the art, as well as the applicability of the proposed method in real scenarios.

References

[1] Brun, L., Saggese, A., Vento, M.: A clustering algorithm of trajectories for behaviour understanding based on string kernels. In: IEEE SITIS. pp. 267–274 (2012), isbn 978-1-4673-5152-2

[2] Brun, L., Saggese, A., Vento, M.: Learning and classification of car trajectories in road video by string kernels. In: VISAPP. pp. 709–714 (2013), isbn 978-989-8565-47-1

[3] Brun, L., Saggese, A., Vento, M.: Dynamic scene understanding for behavior analysis based on string kernels. IEEE Trans. on Circuits and Systems for Video Technology PP(99), 1–1 (2014), isbn 1051-8215

[4] Carletti, V., Foggia, P., Percannella, G., Saggese, A., Strisciuglio, N., Vento, M.: Audio surveillance using a bag of aural words classifier. In: IEEE AVSS. pp. 81–86 (2013), isbn 10.1109/AVSS.2013.6636620

[5] Conte, D., Foggia, P., Percannella, G., Saggese, A., Vento, M.: An ensemble of rejecting classifiers for anomaly detection of audio events. In: IEEE AVSS. pp. 76–81 (2012)

[6] d'Acierno, A., Leone, M., Saggese, A., Vento, M.: A system for storing and retrieving huge amount of trajectory data, allowing spatio-temporal dynamic queries. In: IEEE ITSC. p. 989994 (2012), issn 2153-0009

[7] Di Lascio, R., Foggia, P., Percannella, G., Saggese, A., Vento, M.: A real time algorithm for people tracking using contextual reasoning. Computer Vision and Image Understanding 117(8), 892–908 (2013), issn 1077-3142

[8] Di Lascio, R., Foggia, P., Saggese, A., Vento, M.: Tracking interacting objects in complex situations by using contextual reasoning. In: VISAPP. pp. 104–113 (2012), issn 978-989-8565-04-4

[9] dAcierno, A., Leone, M., Saggese, A., Vento, M.: An efficient strategy for spatio-temporal data indexing and retrieval. In: KDIR. p. 227232. SciTePress (2012), isbn 978-989-8565-29-7

[10] Foggia, P., Percannella, G., Saggese, A., Vento, M.: Real-time tracking of single people and groups simultaneously by contextual graph-based reasoning dealing complex occlusions. In: IEEE PETS. pp. 29 – 36 (2013), issn 2157-491X

[11] Saggese, A.: Behavior Analysis. LAP LAMBERT Academic Publishing (2014), isbn 978-3659529634