Research Activities @MIVIA Lab

Vincenzo Carletti, Luca Del Pizzo, Rosario Di Lascio, Pasquale Foggia,
Gennaro Percannella, Alessia Saggese, Nicola Strisciuglio, Mario Vento

Dept. of Computer Eng. and Electrical Eng., University of Salerno
via Giovanni Paolo II, 84084 Fisciano (SA) - Italy
mivia@unisa.it

Abstract. The actual research activity at MIVIA (Machine Intelligence for recognition of Video, Images and Audio) lab involves the study of innovative methods for behavioral analysis in surveillance videos and for events detection in audio streams, the development of techniques for biomedical images analysis and algorithms for graph matching. Moreover, in the last years, part of the research activity of the MIVIA lab is dedicated to the implementation of those approaches on embedded systems.

1 Graph Matching

Graph matching encompasses different methodologies for graph analysis applied in many fields, such as video analysis, bioinformatics, social network analysis and search engines, as discussed in [19, 18, 12, 5]. Our research group has worked on these problems for more than 10 years and it has provided many important contributions, such as the well-known *VF* algorithm [7], a reference in the state of the art methods for exact graph matching.

Among the above mentioned application fields, two of the most promising are bioinformatics and chemoinformatics, due to the extensively use of graph based representations and the huge quantity of data collected in the last decades. For this reason, in the last years our research activities focused on graph matching applications for these two application fields. Recently, we presented a preliminary benchmark of exact graph matching algorithms on biological graph databases [4]. Furthermore, we are involved in the organization of an international contest on pattern search on proteins and molecules databases, hosted by the next ICPR 2014 (http://biograph2014.unisa.it/).

2 Behavior Analysis by Video Interpretation

The need for security in many public environments has lead to 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 as well as the capability of analyzing very complex events. For this reasons, in the last years we deeply investigated toward the possibility of designing an *intelligent* surveillance system, able to provide images and video with a semantic interpretation, 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 [17]. In particular, our activities focus on two main typologies of behaviors: *actions* and *activities*. On one side, for the problem of recognizing human actions, our contributions pertain the introduction of high level representations to describe the actions, respectively based on bag of words [16], string kernels [1] and deep architectures [14].

On the other side, activity recognition is based on the analysis of trajectories of moving objects: it is due to the fact that trajectory is a very discriminant feature, since the movement of an object in a scene is not random, but instead has an underlying structure that can be exploited to build some models. For this reason, we proposed a tracking algorithm based on a finite state automaton able to extract moving objects trajectories [9]. The proposed tracking algorithm has been evaluated during an international competition (PETS 2013) and ranked at the first position for all the considered scores competing against a high number of participants [15]. Once extracted, this large amount of trajectories needs to be indexed and properly stored in order to improve the overall performance of the framework during the retrieving step [8]. Furthermore, the proposed system allows the human operator to interact with the system in order to identify an abnormal behavior as soon as it occurs [2]: this last feature is achieved by representing each trajectory as a string and by evaluating the similarity between trajectories by a kernel based approach.

3 Audio Event Recognition

Currently used surveillance systems are mainly based on the automatic analysis of video streams. However, in some cases the video-based surveillance in not reliable or suffers from a number of problems, both logistic and about privacy. As an example, poorly illuminated zones during the night or huge parking areas are difficult and very expensive to monitor by means of cameras. In other situations, instead, like in public toilets, it is not possible to install surveillance cameras due to privacy issues. Thus, in these kinds of situation, the analysis of the audio stream could be useful to detect abnormal events and consequently fire an alarm to a human operator. There are many kinds of event that are difficult to be recognized in video streams but can be effectively detected by using audio sensors such as gunshots, human shouting or crying, glass breaking and so on.

Thus, the interest of MIVIA Lab on automatic systems for abnormal audio events detection increased in the last years, leading to the publication of a number of papers and to a more intensive research activity. A method based on two classifiers that work at different time scales [6] has been proposed in order to deal with both impulsive and sustained sounds (for instance gunshots and screams). The classifiers employ a reject option to lower the false alarm rate. A different approach, based on the *bag of words* paradigm, has been proposed in [3]. The main idea is to describe the audio stream with a two layers representation based on the histogram of the occurrences of different small units of sound, which we call *aural words*. Such representation and the use of Support Vector Machines to learn a model of the data is able to highlight the distinctive aural words of the sounds of interest making it possible to distinguish their occurrence in different environments. The most recent work, accepted for publication at AVSS Conference 2014 [13], is based on a biological-inspired image representation of the audio signal, called Gammatonegram, that corresponds to the way the cochlea membrane in the human auditory system responds to stimulation. The detection is performed by employing a pool of cascade classifiers, each of them trained to detect a particular class of sounds. The proposed approach is able to detect co-occurring events and is more robust to background noise with respect to the previous proposed methods.

4 Embedded Systems

Nowadays, embedded systems are everywhere: cameras, phones, electrical appliances, clothes, etc. The diffusion of such systems is due to the hardware miniaturization that makes the manufacturers able to produce systems, for general or specific purposes, with reduced space and power requirements. Thus, new interesting challenges arises to employ embedded systems in many different fields such as robotics, automotive, medicine and so on. One of the most promising research sector is Embedded Vision that aims to combine computer vision algorithms and embedded systems. The strong interest in Embedded Vision is due to the wide number of useful applications that could be realized by developing low cost video analytic systems embedded on cameras. Part of our recent research activities have been focused on re-engineering computer vision algorithms for smart camera embedded platforms in order to realize analytic systems for vehicles and people counting or intrusion detection. Moreover, the implementation of audio analysis algorithms on such systems is also involving part of our research. With regards to this, a big effort has been done to design an algorithm for audio event recognition suitable to low profile embedded systems in order to realize cheap systems for events detection based on the *bag of words* paradigm.

5 HEp2 Cells Classification

The Anti-Nuclear Antibodies (ANA) test is commonly used to diagnose Connective Tissue Diseases. The gold standard for performing this test is the Indirect Immunofluorescence (IIF) protocol using Human Epithelial type 2 (HEp-2) cells, which detects a wide variety of human auto-antibodies. Nevertheless, the protocol is time consuming and labour intensive. One way to address these issues is by employing Computer Aided Diagnosis (CAD) systems, which provide a more objective analysis that could be incorporated to the overall test results. For this reason, in the last years this topic has been attractive for several scientists working in the fields of computer vision and pattern recognition.

Our lab has been very active in this field in the last years [10, 11]: indeed, we organized three different international competitions, aiming to compare the state of the art approaches: *HEp-2 cells classification contest*, hosted by ICPR 2012; *Competition on Cells Classification by Fluorescent Image Analysis*, hosted by ICIP2013 and *Performance Evaluation of Indirect Immunofluorescence Image Analysis Systems*, hosted by the next ICPR 2014. In particular, the next ICPR will also host the I3A workshop (*International Workshop on Pattern Recognition Techniques for Indirect Immunofluorescence Images Analysis*, http://i3a2014.unisa.it/), organized as the last two contests in collaboration with the University of Queensland (Australia) and the Sullivan Nicolaides Pathology (Australia) institute.

References

1. Brun, L., Percannella, G., Saggese, A., Vento, M.: Hack: A system for the recognition of human actions by kernels of visual strings. In: Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on. pp. 142–147 (Aug 2014)
2. 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
3. Carletti, V., Foggia, P., Percannella, G., Saggese, A., Strisciuglio, N., Vento, M.: Audio surveillance using a bag of aural words classifier. In: IEEE Int. Conf. on Advanced Video and Signal Based Surveillance. pp. 81–86 (Aug 2013)
4. Carletti, V., Foggia, P., Vento, M.: Performance comparison of five exact graph matching algorithms on biological databases. In: ICIAP Workshops 2013. vol. 8158, pp. 409–417. Springer Berlin Heidelberg (2013)
5. Conte, D., Foggia, P., Sansone, C., Vento, M.: Thirty years of graph matching in pattern recognition. International Journal Of Pattern Recognition And Artificial Intelligence 18(3), 265–298 (2004)
6. 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)
7. Cordella, L., Foggia, P., Sansone, C., Vento, M.: A (sub)graph isomorphism algorithm for matching large graphs. IEEE Transactions on Pattern Analysis And Machine Intelligence 26(10), 1367–1372 (2004)
8. 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. 98917994 (2012), issn 2153-0009
9. 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
10. Foggia, P., Percannella, G., Saggese, A., Vento, M.: Pattern recognition in stained hep-2 cells: Where are we now? Pattern Recognition (2014)
11. Foggia, P., Percannella, G., Soda, P., Vento, M.: Benchmarking hep-2 cells classification methods. Medical Imaging, IEEE Transactions on 32(10), 1878–1889 (Oct 2013)

12. Foggia, P., Percannella, G., Vento, M.: Graph matching and learning in pattern recognition in the last 10 years. International Journal of Pattern Recognition and Artificial Intelligence (2013), http://www.worldscientific.com/doi/abs/10.1142/S0218001414500013
13. Foggia, P., Saggese, A., Strisciuglio, N., Vento, M.: Cascade classifiers trained on gammatonegrams for reliably detecting audio events. In: Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on. pp. 50–55 (Aug 2014)
14. Foggia, P., Saggese, A., Strisciuglio, N., Vento, M.: Exploiting the deep learning paradigm for recognizing human actions. In: Advanced Video and Signal Based Surveillance (AVSS), 2014 11th IEEE International Conference on. pp. 93–98 (Aug 2014)
15. 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
16. Foggia, P., Percannella, G., Saggese, A., Vento, M.: Recognizing human actions by a bag of visual words. In: IEEE SMC. pp. 2910–2915 (2013)
17. Saggese, A.: Behavior Analysis. LAP LAMBERT Academic Publishing (2014), isbn 978-3659529634
18. Vento, M.: A one hour trip in the world of graphs, looking at the papers of the last ten years. In: Kropatsch, W., Artner, N., Haxhimusa, Y., Jiang, X. (eds.) Graph-Based Representations in Pattern Recognition, Lecture Notes in Computer Science, vol. 7877, pp. 1–10. Springer Berlin Heidelberg (2013), http://dx.doi.org/10.1007/978-3-642-38221-5_1
19. Vento, M.: A long trip in the charming world of graphs for pattern recognition. Pattern Recognition (2014), http://www.sciencedirect.com/science/article/pii/S0031320314000053