Master Thesis Proposal

Official Master EPS: i2-ICT
Title: Online contextual updating in multi-camera scenarios
Student: LÓPEZ CIFUENTES, Alejandro
Director: ESCUDERO VIÑOLO, Marcos
Supervisor: BESCÓS CANO, Jesús

Course: 2016-17
Date: 27/03/2017

1. Introduction

Nowadays, we live surrounded by electronic devices which objective is to ensure the safety and security of the global population or to ease our lives on everyday tasks. These range from biometric systems [1] to all kind of different electrical sensors, including video surveillance cameras. These cameras are the ones that are of real interest when developing Image Processing and Computer Vision algorithms in the video surveillance scope [2].

The combination of these veins of research could lead to the automation of high-level human semantic tasks such as people detection [3], object detection and recognition [4, 5, 6] and extraction of contextual information [7]. The automation of these processes permits end-users build on these information sources to define the latest stages of video surveillance systems. These are usually the critical ones, e.g. alarm raising when some predefined event occurs.

Among Computer Vision applications running on a multi-camera scenario, a pivotal field of research is the analysis of public spaces. These are often crowd-populated scenarios which analysis requires the combination of the data obtained by all recording cameras. It is of real interest to analyse people behaviour patterns [8, 9, 10] and temporal usage of a given area in large-scale scenarios such as shopping malls, universities and, generally, in any public-use building, either to extract statistical measures of behaviour or to detect anomalous unexpected events [11]. This analysis will come from a combination of complementary algorithms such as semantic area classification, people detection and crowd behaviour analysis.

2. Goals and Planning

The main objective of this master thesis is to extract a contextual description about a populated large-scale scenario while extracting temporal statistical usage data from relevant scene spaces using various cameras. This should be supported with a GUI application.

To fulfil this objective, this work will embrace two different blocks of objectives that will complement each other. The first one will have to do with the design of a graphical user interface (GUI) whereas the second block will deal with algorithm and research-related objectives.

Graphical User Interface

The main user interface should be able to visualise and dynamically arrange usage statistics from different areas of interest in a public space—either pre-generated or generated under a real-time constraint, under a user-friendly environment.

Algorithm

The algorithm related objectives are:

1. To integrate a semantic segmentation algorithm to perform contextual element analysis in video sequences. The objective is to detect and classify and determine the spatial extend on each frame of the video of relevant elements such as doors, desks, corridors and floor areas. We aim to:
    a) Identify the current state of all these elements in each of the processed cameras. The state should distinguish between visible and occluded.
    b) Identify the usage rate of some important elements of the scene measured by number of people per time interval.
2. To integrate state-of-the-art people detection algorithms results per view. To this aim, we need to:
    a) Create a fusion mechanism to take advantage of the multi-camera scenario. The results from all the cameras are projected on a common space and combined such that—by result's refinement—the individual detector performances are increased.
    b) Analyse people and crowd motion patterns to—by combining them with contextual information—define usage statistical measures on predefined spaces where specific activities take place.

1. 3. Methodology and work schedule

The work methodology for the consecution of the objectives requires:

- The study of a general perspective of the available methods in the different scopes that will be analysed: People detection, semantic classification: 80 hours.
- Research and integration/development of the selected State Of the Art methods that will be included in the project: 350 hours.
- Development of the graphical user interface: 150 hours.
- Quantitative and qualitative evaluation of the integrated and developed methods: 170 hours.

The total number of scheduled hours is 750, which will be distributed on a 37,5 per week basis during 20 weeks.

1. 4. Initial bibliography

[1] A. K. Jain, L. Hong, and Y. Kulkarni, "A multimodal biometric system using fingerprint, face and speech," in 2nd Int'l Conf. AVBPA, vol. 10, 1999.

[2] X. Wang, "Intelligent multi-camera video surveillance: A review," Pattern recognition letters, vol. 34, no. 1, pp. 3–19, 2013.

[3] P. Dollar, C. Wojek, B. Schiele, and P. Perona, “Pedestrian detection: An evaluation of the state of the art,” IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 4, pp. 743–761, 2012.
[4] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, “Object detection with discriminatively trained part-based models,” IEEE transactions on pattern analysis and machine intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.
[5] S. Ren, K. He, R. Girshick, and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” in Advances in neural information processing systems, pp. 91–99, 2015.
[6] J. Hosang, R. Benenson, P. Dollár, and B. Schiele, “What makes for effective detection proposals?,” IEEE transactions on pattern analysis and machine intelligence, vol. 38, no. 4, pp. 814–830, 2016.
[7] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” arXiv preprint arXiv:1612.01105, 2016.
[8] R. Mazzon and A. Cavallaro, “Multi-camera tracking using a multi-goal social force model,” Neurocomputing, vol. 100, pp. 41–50, 2013.
[9] A. Turner and A. Penn, “Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment,” Environment and planning B: Planning and Design, vol. 29, no. 4, pp. 473–490, 2002.
[10] P. Scovanner and M. F. Tappen, “Learning pedestrian dynamics from the real world,” in Computer Vision, 2009 IEEE 12th International Conference on, pp. 381–388, IEEE, 2009.
[11] F. Jiang, J. Yuan, S. A. Tsaftaris, and A. K. Katsaggelos, “Anomalous video event detection using spatiotemporal context,” Computer Vision and Image Understanding, vol. 115, no. 3, pp. 323–333, 2011.

5. Proposal of the evaluating committee

José María Martínez Sánchez

Álvaro García Martín.

Marcos Escudero Viñolo (director)