UNIVERSIDAD AUTÓNOMA DE MADRID
ESCUELA POLITÉCNICA SUPERIOR

ONLINE CONTEXTUAL UPDATING IN MULTI-CAMERA SCENARIOS

Alejandro López Cifuentes
Director: Marcos Escudero Viñolo
Supervisor: Jesús Bescós Cano

-MASTER THESIS-

Departamento de Tecnología Electrónica y de las Comunicaciones
Escuela Politécnica Superior
Universidad Autónoma de Madrid
July 2017

ONLINE CONTEXTUAL UPDATING IN MULTI-CAMERA SCENARIOS

Alejandro López Cifuentes

Director: Marcos Escudero Viñolo
Supervisor: Jesús Bescós Cano

Video Processing and Understanding Lab
Departamento de Ingeniería Informática
Escuela Politécnica Superior
Universidad Autónoma de Madrid
July 2017

Abstract

The

Keywords

Keywords.

Acknowledgements

Alejandro López Cifuentes.
2017.

Contents

Abstract **v**

Acknowledgements **vii**

1 Introduction **1**
1.1 Motivation 1
1.2 Objectives 2
1.3 Thesis Structure 2

2 State of the Art **5**
2.1 Pedestrian Detection. 5
2.2 Contextual Information. 5
2.3 Required Analysis. 5
2.3.1 Multi camera scenarios. 5
2.3.2 Semantic segmentation. 5
2.3.3 Spatiotemporal constraining. 5

3 Results **7**
3.1 Hardware 7
3.1.1 Camera Specifications 7
3.1.2 Camera User Interface (GUI) 8

4 Conclusions and Future Work **9**
4.1 Conclusions 9
4.2 Future Work 9

Bibliography **10**

List of Figures

3.1 Sony SNC-RZ50P Pan/Tilt Range 7

List of Tables

3.1 Sony SNC-RZ50P Specifications 8

Chapter 1

Introduction

1.1 Motivation

Nowadays, we live completely surrounded of electronic devices whose objective is to ensure the safety and security of the global population. From biometric systems [1] to all kind of different electrical sensors, not forgetting the wide range of, for example, video surveillance cameras. Those last are the ones that could be of real interest when working along with Image Processing and Computer Vision algorithms in the video surveillance scope [2].

The combination of these two systems could lead to the automation of high-level tasks such as people detection [3], object detection and classification [4, 5, 6], contextual information extraction [7] ... etc. The automation of all these processes permits people working with them to focus on the latests stages of video surveillance systems which should be the most critical ones, such as alarm raise when some event has occurred, letting the heavy and tedious computational part to computers.

One of the automation task that could appear when dealing with Computer Vision and multi video surveillance cameras is the analysis of public spaces which are often populated with crowds and the combination of the data coming from all of the camera instances. It could be from real interest to analyze people patrons [8, 9, 10] and temporal area/space usage data in large spaces such as public shopping centers, universities, common building areas... either to extract statistical measures or anomalous events [11]. This analysis will come from a combination of algorithms of distinct scopes such as semantic area classification, people detection or crowd patron analysis.

1.2 Objectives

This master thesis will embrace two different blocks of objectives that will complement each other. The first one will have to do with the performance of an user interface application while the second block will deal with algorithm and investigation related objectives.

User Interface

The main user interface should be able to visualize temporal statistical usage data, in a user-friendly environment, either pre-generate or generated in a real time constraint from different areas of a common space.

Algorithm

The algorithm related objectives will be:

1. Integrate an algorithm to perform contextual elements analysis in video sequences. The objective is detect and classify labels and its position over the important frame elements such as doors, desks, corridors, floor...
   (a) Identify the current state of all these elements in each of the processed cameras. The state should distinguish between visible or occluded.
   (b) Identify usage rate of some important elements of the scene measured by people per second.
2. Integrate people detection over the cameras frames to :
   (a) Create a fusion taking advantage of a multi-camera scenario so that the information is transformed to common planes and the detector performance is increased.
   (b) Analyze people and crowd patrons to, combined with contextual information, create statistical measures about important areas of influence where particular activities often occur.

1.3 Thesis Structure

The master thesis is divided into the following chapters:

- Chapter 1. Introduction.

- Chapter 2. State of the Art.
- Chapter 3. Results.
- Chapter 4. Conclusions and Future Work.
- Annex.
- References.

Chapter 2

State of the Art

As explained during Chapter 1 the process of analysis of a common crowed space will join many algorithms from different Computer Vision disciplines. Throughout this Chapter we will summarize the actual and most used algorithms in the different categories that we will use during the development of the project.

2.1 Pedestrian Detection.

2.2 Contextual Information.

2.3 Required Analysis.

2.3.1 Multi camera scenarios.

2.3.2 Semantic segmentation.

2.3.3 Spatiotemporal constraining.

Chapter 3

Results

3.1 Hardware

The project has been developed in the Escuela Politécnica Superior (Universidad Autónoma de Madrid). Due to this fact, the testing environment has been the hall of the mentioned engineering school which has a set up of three different Internet Protocol Cameras (IP Cameras). This type of cameras can send and receive data via a computer network and the Internet which allows the user to set the configuration and receive frames from the cameras.

3.1.1 Camera Specifications

Specifically, the camera model used along the project has been the Sony SNC-RZ50P PTZ Camera . This is a PTZ camera which means that it will be able to Pan, Tilt and Zoom all over the scene are. Precisely this camera will have a pan range of 340 degrees and a tilt range of 115 degrees, enabling users to monitor a wide area over the scene if the camera is moved (Figure 3.1) .

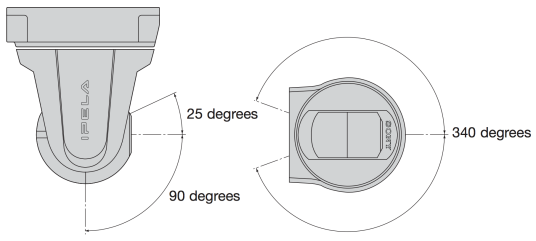

Figure 3.1: Sony SNC-RZ50P Pan/Tilt Range

The complete and relevant specifications are detailed in Table 3.1:

Camera	
Horizontal viewing angle	1.7 to 42.0 degrees
Focal length	f=3.5 to 91.0 mm
F-number	F1.6 (wide), F3.8 (tele)
Minimum object distance	320 mm (wide),1,500 mm (tele)
Pan angle	-170 to +170 degrees
Pan speed	300 degrees/s (max.)
Tilt angle	-90 to +25 degrees
Tilt speed	300 degrees/s (max.)

Image		
Image size (H x V)	640 x 480, 320 x 240, 160 x 120	
Compression format	JPEG, MPEG-4, H.264	
Maximum frame rate	JPEG/MPEG-4	25 fps (640 x 480)
	H.264	8 fps (640 x 480)

Table 3.1: Sony SNC-RZ50P Specifications

3.1.2 Camera User Interface (GUI)

Chapter 4

Conclusions and Future Work

4.1 Conclusions

4.2 Future Work

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.

[12] "Sony snc-rz50 ptz camera specifications."