TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

D2.1 v1

People Behaviour understanding in single and multiple camera settings

Video Processing and Understanding Lab

Escuela Politécnica Superior

Universidad Autónoma de Madrid

Supported by

AUTHORS LIST

Álvaro García alvaro.garcia@uam.es

HISTORY

Version	Date	Editor	Description
0.9	7 December 2015	Álvaro García	Final Working Draft
1.0	11 December 2015	José M. Martínez	Editorial checking

CONTENTS:

1. Introduction

1.1. Motivation

This work package 2 (WP2) aims at developing the required tools, models and control signals in order to enable the development of adaptive and collaborative approaches for video-based understanding of people behaviour. In particular, the goal of this task is to provide the required video analysis tools to fulfill the objectives of the project. It considers both comprehensive related work studies and the development of novel approaches for long-term analysis.

This deliverable describes the work related with the task T.2.1 Analysis Tools for human behaviour understanding. The people behaviour understanding in this project has been already designed as a sequential combination of object segmentation, people detection, object tracking and behaviour recognition. In particular, during the first part of the project there have been a focus on developing different approaches for segmentation and people detection.

1.2. Document structure

This document contains the following chapters:

- Chapter 1: Introduction to this document
- Chapter 2: Object segmentation analysis tools
- Chapter 3: People detection analysis tools
- Chapter 3: Conclusions

2. Object segmentation analysis tools

2.1. Long-Term Stationary Object Detection Based on Spatio-Temporal Change Detection

Stationary Object Detection (SOD) has recently experienced extensive research [1] due to its contribution to prevent terrorist attacks by detecting abandoned objects [2] and illegal parked vehicles [3]. SOD aims to detect the objects in the scene that remain stationary after previous motion. Typically, a Background Subtraction (BS) algorithm extracts the objects and SOD decides whether they are stationary or not [4]. However, current BS algorithms present many shortcomings to label foreground and background regions in real situations [5], thus highly determining the SOD accuracy. In this paper [6] we propose a block-wise approach to detect stationary objects based on spatio-temporal change detection without using BS (see Figure 1).

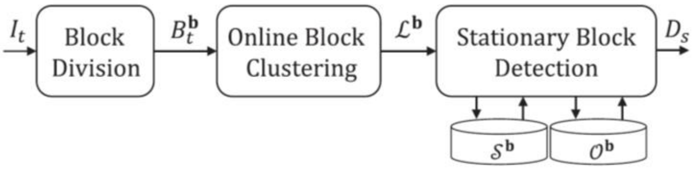

Figure 1. Block diagram of the proposed approach.

Firstly, a Block Division stage decomposes each frame into non-overlapping $N \times N$ blocks $B_t^{\mathbf{b}}$ at each instant t, where $\mathbf{b}$ denotes the block location. Secondly, an Online Block Clustering stage models each location $\mathbf{b}$ over time, updating a cluster partition $\mathcal{L}^{\mathbf{b}}$. This stage handles the temporal adaptation to scene changes, by assigning each incoming block to one cluster of the partition or creating a new one. Only stationary blocks (i.e., without motion with respect to $B_{t-1}^{\mathbf{b}}$) are analyzed at this stage. This clustering provides robustness against illumination changes by considering pixel ratios at block level which groups blocks even if their illumination has changed. Finally, a Stationary Block Detection stage outputs a result image D_s with stationary objects (see Figure 2), where s defines the sampling instant each k frames. Data associated to the last stable cluster $\mathcal{S}^{\mathbf{b}}$, old stable clusters $\mathcal{O}^{\mathbf{b}}$ and the alarm time T is used to respectively detect the spatio-temporal stability changes, discard those changes caused by previously visualized clusters (i.e. empty scene or previous detections) and detect stationarity for changes longer than the alarm time.

Figure 2. Examples of D_s image in different datasets. Detections are marked in red.

The last stage improves the state-of-the-art by reducing false alarms due to intermittent object motion and allowing to detect stationarity for objects not fully visible during T. Figure 3 presents an example of the scene analysis.

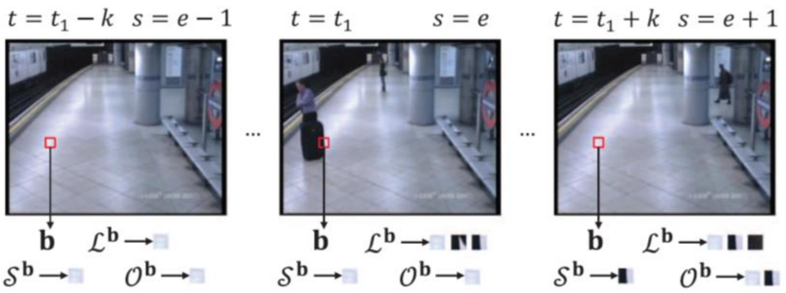

Figure 3. Example of the temporal analysis for a block location where the stability is modified changing from the empty scene to a suitcase.

3. People detection analysis tools

3.1. People density estimation in crowded environments

Currently, the use of artificial vision systems has acquired great relevance due to the advancement of digital image and video processing technologies and the cheapening of capture tools. The crowd density estimation as part of artificial vision systems has an important niche market in video-surveillance. The crowd density estimation is an important tool to detect abnormal situations in public places such as fights, disturbances, violent protests, panic or congestion. Density information could be also helpful for creating a business strategy according to the distribution of people in public places or shopping centers and the distribution of people over time.

So far, a large number of crowd density estimators has been implemented. A significant part of these estimators use background subtraction and extract features from foreground pixels [11] [12]. All of them use foreground-background segmentation getting good results for the studied scenarios. This work [8] introduces the use of people-background [13] segmentation for crowd density estimation. With the goal of comparing both types of segmentation, one algorithm of the state of the art for density estimation has been implemented and then, results for both types of segmentation, foreground-background and people-background segmentation, has been compared in several scenarios (see Figure 4 and Figure 5).

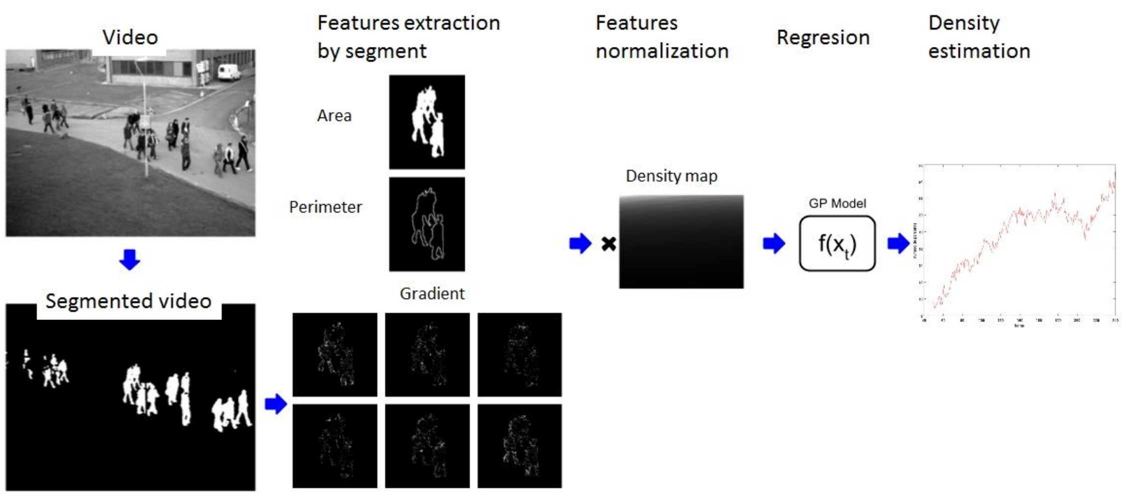

Figure 4. People density estimation system based on foreground-background segmentation.

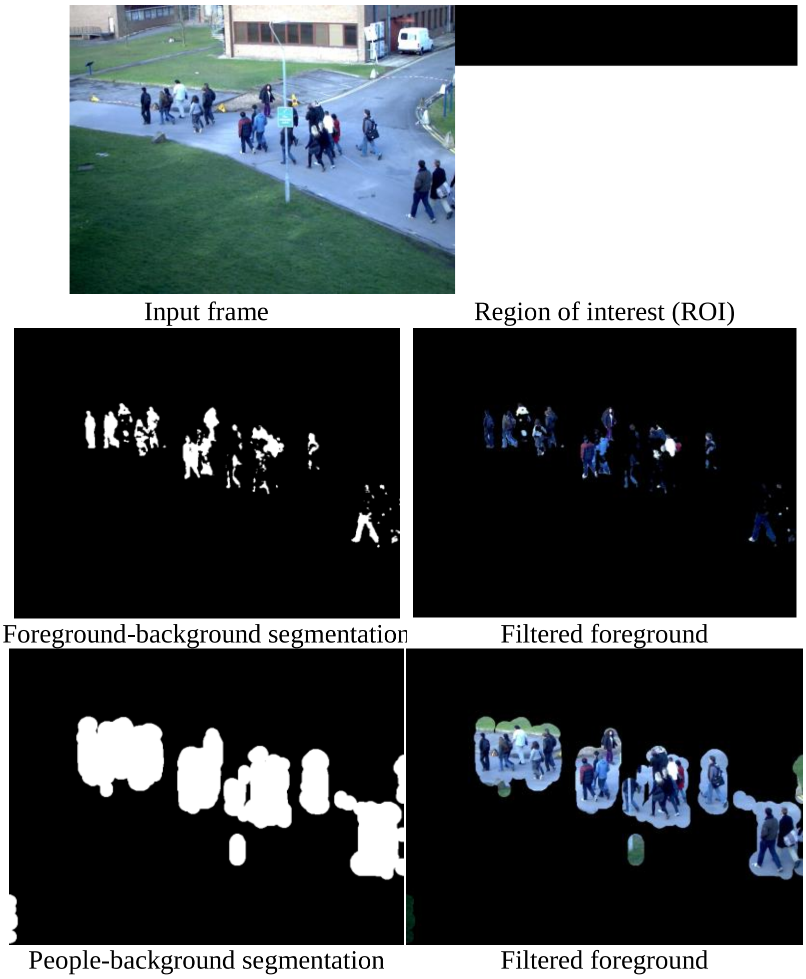

Figure 5. Examples of foreground-background segmentation and people-background segmentation.

3.2. People detection in groups

In video signal processing, people detection is one of the most difficult tasks, even though there are actually some algorithms that give good results. However, when the detection takes part in complex environments the quality of the performance decreases, that is why the aim of this work [9] is the implementation of a people detection in groups algorithm in C++, as well as the integration on the proprietary video analysis platform called Distributed Video Analysis (DiVA). This platform allows us the execution with online and off-line videos, with the objective of verifying the functionality and efficiency in crowded environments.

The algorithm implemented, known as Multi-configuration Part-based Person Detector [14], is an adaptation of the algorithm known as DTDP (Discriminatively Trained Part Based Models) [15] to improve the detection in these types of environments. It is founded in an exhaustive search of person models, which are formed by the mixture of different body parts. Thanks to this search it obtains good results despite of the processing time, therefore the analysis of this algorithm is not possible in real time.

In order to complete our goal, we have developed a user interface so that any user can interact with the algorithm and prove easily the functionality of the algorithm in different environments as well as test the efficiency of the different models defined (see Figure 6 and Figure 7).

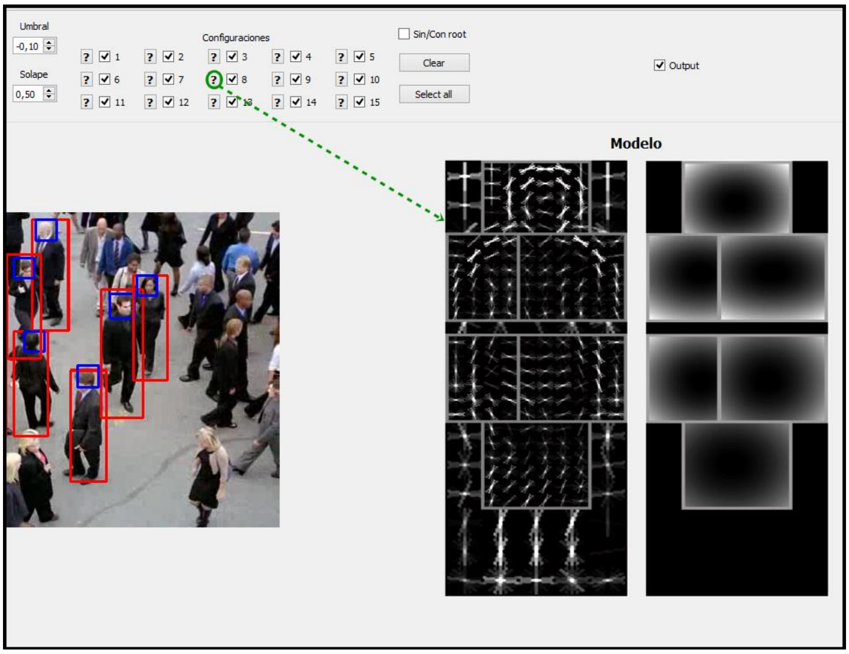

Figure 6. Visual example of the body parts configuration including six body parts, named configuration 8, following [16].

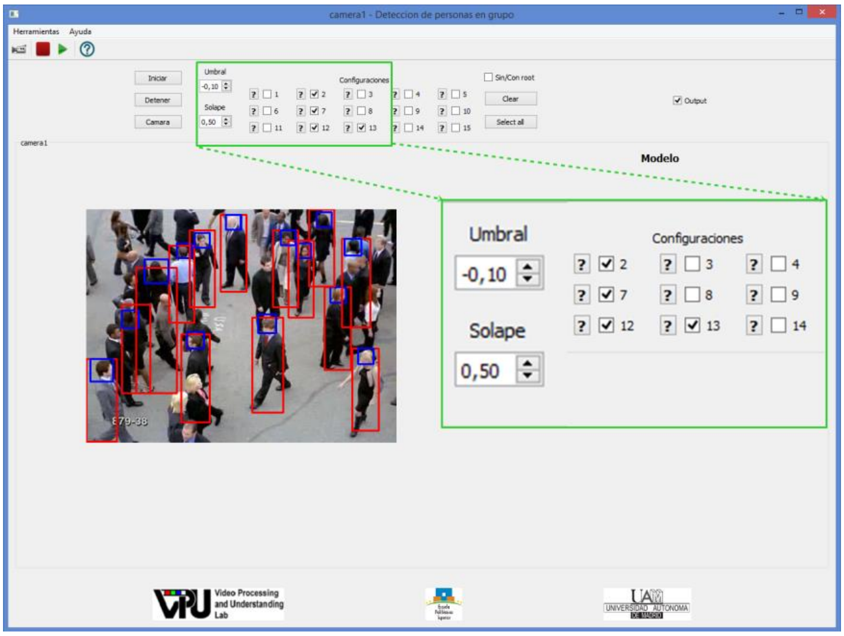

Figure 7. User interface example with selected configurations 2,7,12 and 13, following [16].

3.3. People detection in presence of groups

In this work [9] we address one of the most typical problems of people detection in presence of groups of people: in this kind of scenarios, traditional people detectors have difficulties dealing with several occlusions. In order to deal with this problem, we propose the use of two different hierarchies. The first one consists of a hierarchy of people, i.e., the use of the detections of different people belonging to a group in order to refine the individual's detections. The second one consists of a hierarchy of parts [14], i.e., the use of different combinations of body parts in order to refine the final detection.

Our main aim is to be robust to different groups configurations, camera point of view, scene constraints, etc. Therefore, we propose to update this hierarchies structures frame by frame, so we can adapt the detection system to specific scene variations.

In order to update both the hierarchy of people and the hierarchy of parts, we study the detection results and determine which have been the most typical or representative configurations over time (scales and body parts configurations). Using only the last most representative configuration, we are able to reduce the computational cost and false positive detection and therefore the global detection results.

4. Conclusions

In relation with segmentation, a long-term stationary object detection based on spatio-temporal change detection has been implemented and evaluated [7], we propose a block-wise approach to detect stationary objects based on spatio-temporal change detection without using background subtraction.

In relation with people detection several approaches have been implemented and tested: people density estimation in crowded scenarios [8] and people detection in groups [9], [9]. A significant part of people density estimation approaches from the state of the art use background subtraction and extract features from foreground pixels. All of them use foreground-background segmentation getting good results for the studied scenarios. The proposed work [8], introduces the use of people-background segmentation for crowd density estimation. On the other side, [9] and [9] propose two different approaches in order to deal with people detection in crowded scenarios or in presence of groups of people.

5. References

[1] D. Ortego, J. C.s San Miguel, J. M. Martínez, "Long-Term Stationary Object Detection Based on Spatio-Temporal Change Detection", IEEE Signal Processing Letters, 22(12), 2368-2372, Dec. 2015.
[2] Q. Fan, P. Gabbur, and S. Pankanti, "Relative attributes for large-scale abandoned object detection", in Proc. IEEE Int. Conf. Computer Vision (ICCV), Dec. 2013, pp. 2736–2743.
[3] K. Lin, S. Chen, C. Chen, D. Lin, and Y. Hung, "Abandoned object detection via temporal consistency modeling and back-tracing verification for visual surveillance", IEEE Trans. Inf. Forensics Secur., 2015.
[4] A. Albiol, L. Sanchis, A. Albiol, and J. Mossi, "Detection of parked vehicles using spatiotemporal maps", IEEE Trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1277–1291, 2011.
[5] A. Bayona, J. SanMiguel, and J. Martinez, "Comparative evaluation of stationary foreground object detection algorithms based on background subtraction techniques", in Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS), 2009, pp. 25–30.
[6] T. Bouwmans, "Traditional and recent approaches in background modelling for foreground detection: An overview", Comput. Sci. Rev., vol. 11–12, pp. 31–66, 2014.
[7] D. Ortego, J.C. SanMiguel, J.M. Martinez, "Long-Term Stationary Object Detection Based on Spatio-Temporal Change Detection", IEEE Signal Processing Letters, vol. 22, no. 12, pp. 2368-2372, 2015.
[8] Estimación de la densidad de personas en entornos densamente poblados (People density estimation in crowded environments), Rosely Sánchez (advisor: Álvaro García-Martín), Proyecto Fin de Carrera (Master Thesis), Ingeniería de

Telecomunicación, Escuela Politécnica Superior, Univ. Autónoma de Madrid, May 2015.

[9] Detección de Personas en Grupos (People detection in groups), Marta Villanueva Torres (advisor: Álvaro García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jul. 2015.

[10] Detección de personas en presencia de grupos (People detection in presence of groups), Sergio Merino Martínez, (advisor: Álvaro García Martín), Trabajo Fin de Máster (Master Thesis), Master en Investigación e Innovación en TIC, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Sep. 2015.

[11] A.N. Marana, L.F. Costa, R.A. Lotufo, and S.A. Velastin. On the e-cacy of texture analysis for crowd monitoring. In Computer Graphics, Image Processing, and Vision, 1998. Proceedings. SIBGRAPI '98. International Symposium on, pages 354-361, 1998.

[12] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin, and L. Xu. Crowd analysis: a survey. Machine Vision and Applications, 19(5-6):345-357, 2008.

[13] A. Garcia-Martin, A. Cavallaro, and J.M. Martinez. People-background segmentation with unequal error cost. In Image Processing (ICIP), 2012 19th IEEE In-ternational Conference on, pages 157-160, 2012.

[14] A. Garcia-Martin, R. Heras Evangelio and T. Sikora. A Multi-configuration Part-based Person Detector. In Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications (ICETE 2014).

[15] 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(9), pp. 1627-1645, 2010.