



*HAVi Newsletters*

*#04 - April 2016*

---

TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

<http://www-vpu.eps.uam.es/HAVideo/>

---

### Fourth four-month period progress report

After analyzing the progress during the first year, we have decided in the first days of January to slightly reschedule the project: Task 3.1 "Adaptation approaches" start has been delayed till June 2016 and T4.2 "Use cases and demonstrators" start has been delayed till January 2017; the corresponding milestones and deliverables have been rescheduled accordingly.

After the workplan adjustment, these months the project has been working on plan. As scheduled, there were no deliverables due. We have adjusted the acquisition plan to the budget and launched the main purchases.

Following the successful workshop held May 2015, the second edition of the Developers Workshop is being organized for May 2016.

### Fourth four-month results

#### Journals

ObaidUllah Khalid, Juan C. SanMiguel, Andrea Cavallaro, "Multi-tracker Partition Fusion", IEEE Transactions on Circuits and Systems for Video Technology, Jan. 2016 (accepted), IEEE, ISSN 1051-8215 (DOI [10.1109/TCSVT.2016.2542699](https://doi.org/10.1109/TCSVT.2016.2542699))

**Abstract:** We propose a decision-level approach to fuse the output of multiple trackers based on their estimated individual performance. The proposed approach is divided into three steps. First, we group trackers into clusters based on the spatiotemporal pair-wise correlation of their short-term trajectories. Then, we evaluate performance based on reverse-time analysis with an adaptive reference frame and define the cluster with trackers that appear to be successfully following the target as the on-target cluster. Finally, the state estimations produced by trackers in the on-target cluster are fused to obtain the target state. The proposed fusion approach uses standard tracker outputs and can therefore combine various types of trackers. We tested the proposed

approach with several combinations of state-of-the-art trackers, and also compared it with individual trackers and other fusion approaches. The results show that the proposed approach improves the state estimation accuracy under multiple tracking challenges.

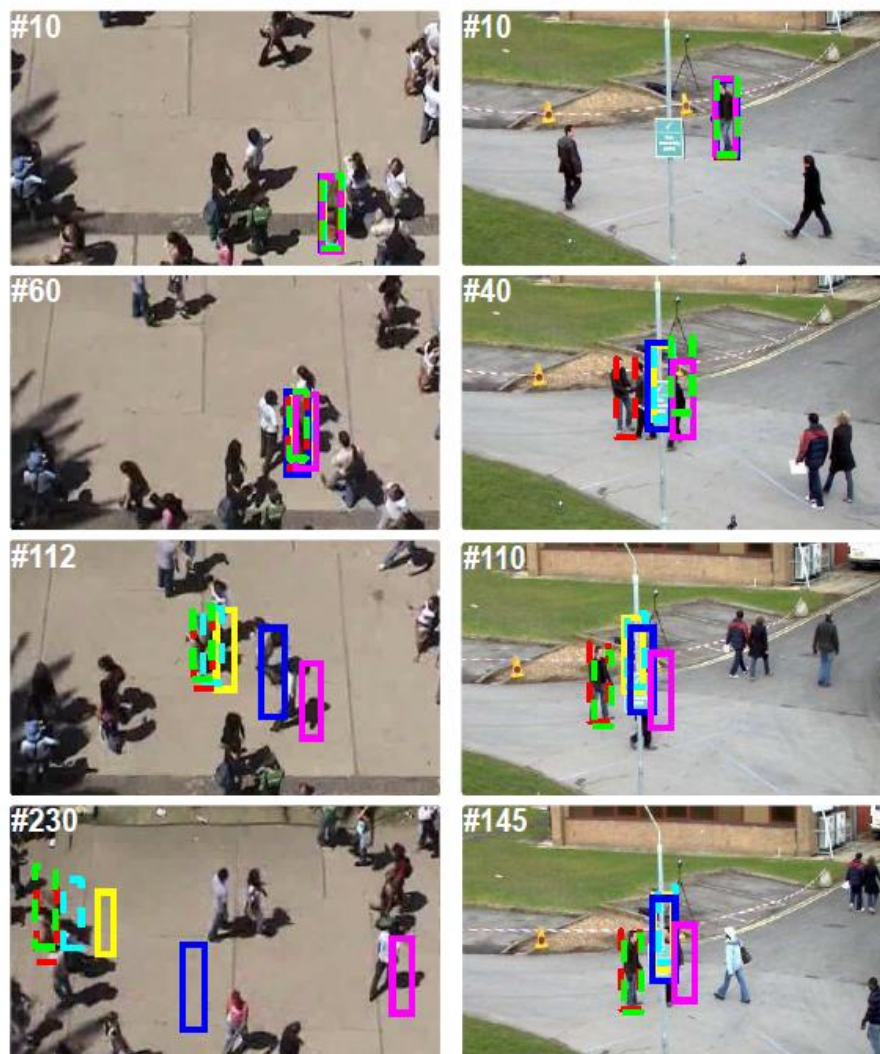
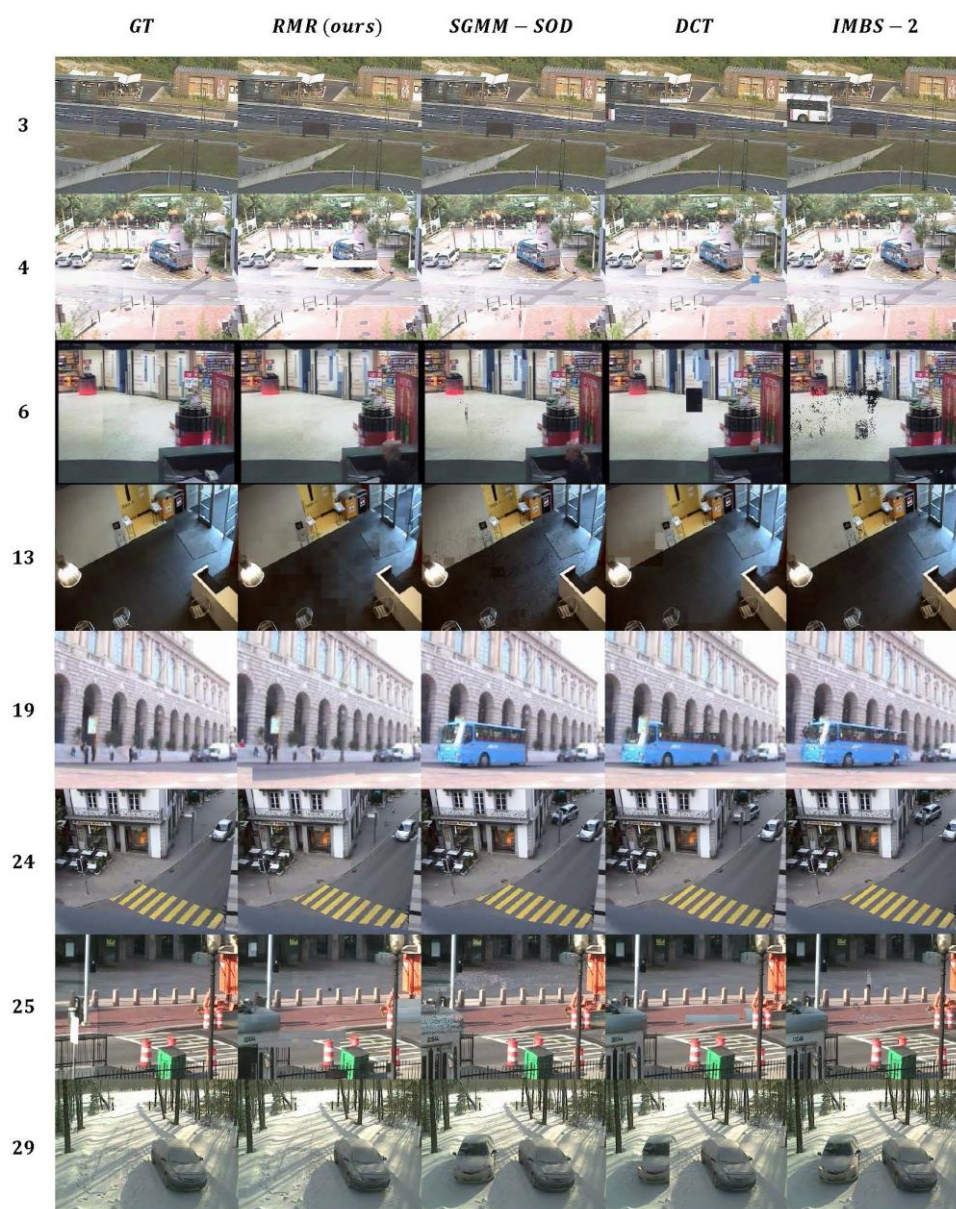


Fig. 11. Sample tracking results for Students-P1 (left column) and PETS-P10 (right column). - - -: TPF; - - -: STRUCK; —: VTS; - - -: SymT; —: AvgF; —: KCF.

Diego Ortego, Juan C. SanMiguel, José M. Martínez, “**Rejection based multipath reconstruction for background estimation in video sequences with stationary objects**”, *Computer Vision and Image Understanding*, Mar. 2016 (accepted), Elsevier, ISSN 1077-3142 (DOI [10.1016/j.cviu.2016.03.012](https://doi.org/10.1016/j.cviu.2016.03.012))

**Abstract:** Background estimation in video consists in extracting a foreground-free image from a set of training frames. Moving and stationary objects may affect the background visibility, thus invalidating the assumption of many related literature where background is the temporal dominant data. In this paper, we

present a temporal-spatial block-level approach for background estimation in video to cope with moving and stationary objects. First, a Temporal Analysis module obtains a compact representation of the training data by motion filtering and dimensionality reduction. Then, a threshold-free hierarchical clustering determines a set of candidates to represent the background for each spatial location (block). Second, a Spatial Analysis module iteratively reconstructs the background using these candidates. For each spatial location, multiple reconstruction hypotheses (paths) are explored to obtain its neighboring locations by enforcing inter-block similarities and intra-block homogeneity constraints in terms of color discontinuity, color dissimilarity and variability. The experimental results show that the proposed approach outperforms the related state-of-the-art over challenging video sequences in presence of moving and stationary objects.



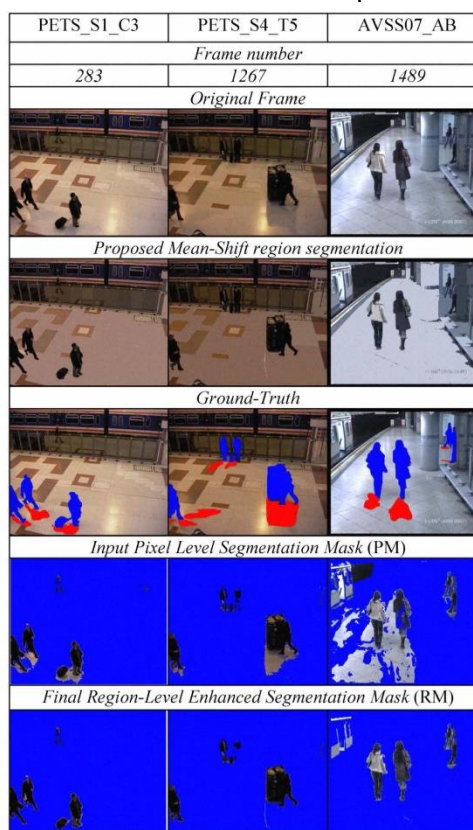
## PhD thesis

**Contributions to region-based image and video analysis: feature aggregation, background subtraction and description constraining**, Marcos Escudero Viñolo (advisor: Jesús Bescós Cano), Tesis Doctoral (PhD Thesis), Escuela Politécnica Superior, Univ. Autónoma de Madrid, Jan. 2016.

**Abstract:** The use of regions for image and video analysis has been traditionally motivated by their ability to diminish the number of processed units and hence, the number of required decisions. However, as we explore in this thesis, this is just one of the potential advantages that regions may provide. When dealing with regions, two description spaces may be differentiated: the decision space, on which regions are shaped—region segmentation—, and the feature space, on which regions are used for analysis—region-based applications—. These two spaces are highly related. The solutions taken on the decision space severely affect their performance in the feature space. Accordingly, in this thesis we propose contributions on both spaces.

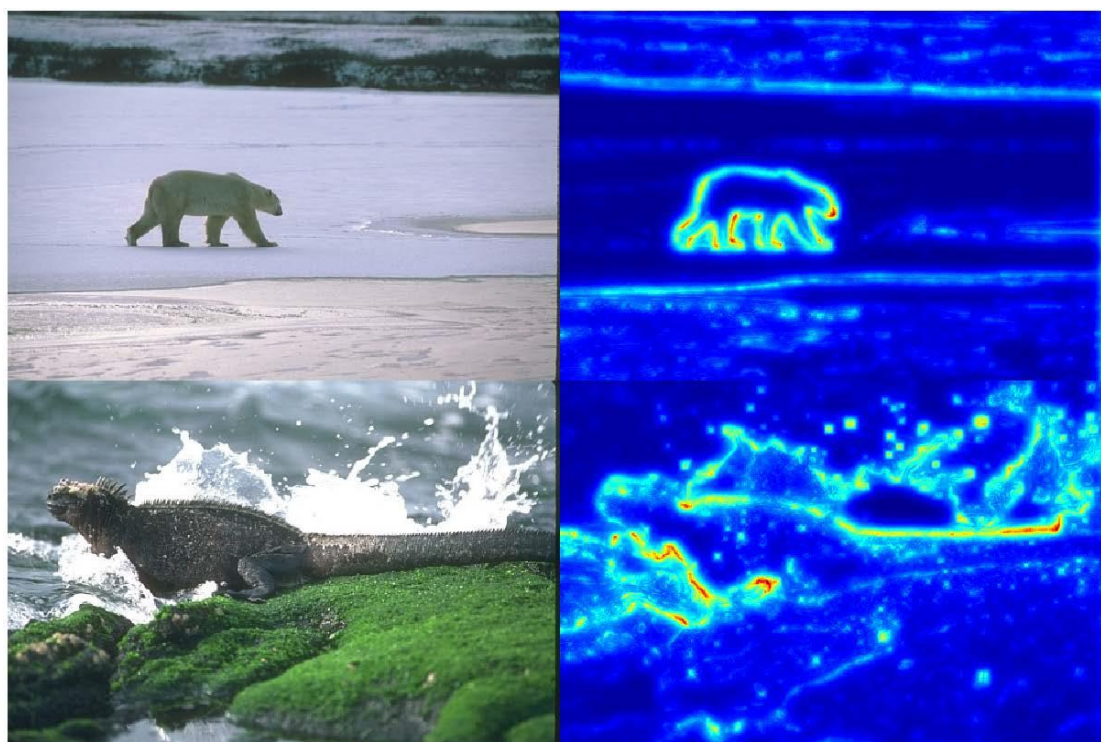
Regarding the contributions to region segmentation, these are two-fold. Firstly, we give a twist to a classical region segmentation technique, the Mean-Shift, by exploring new solutions to automatically set the spectral kernel bandwidth. Secondly, we propose a method to describe the micro-texture of a pixel neighbourhood by using an easily customisable filter-bank methodology—which is based on the discrete cosine transform (DCT)—.

The rest of the thesis is devoted to describe region-based approaches to several highly topical issues in computer vision; two broad tasks are explored: background subtraction (BS) and local descriptors (LD). Concerning BS, regions are here used as complementary cues to refine pixel-based BS algorithms: by providing robust to illumination cues and by storing the background dynamics in a region-driven background modelling. Relating to LD, the region is here used to reshape the description area usually fixed for local descriptors. Region-masked versions of classical two-dimensional and three-dimensional local descriptions are designed. So-built descriptions are proposed for the task of object identification, under a novel neural-oriented strategy. Furthermore, a local description scheme based on a fuzzy use of the region membership is derived. This



characterisation scheme has been geometrically adapted to account for projective deformations, providing a suitable tool for finding corresponding points in wide-baseline scenarios.

Experiments have been conducted for every contribution, discussing the potential benefits and the limitations of the proposed schemes. In overall, obtained results suggest that the region—conditioned by successful aggregation processes—is a reliable and useful tool to extrapolate pixel-level results, diminish semantic noise, isolate significant object cues and constrain local descriptions. The methods and approaches described along this thesis present alternative or complementary solutions to pixel-based image processing.



### Master thesis

Detección de caídas mediante vídeo-monitorización (**Fall detection using video**), David Dean Pulido (advisor: José M. Martínez), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Universidad Autónoma de Madrid, Escuela Politécnica Superior, Mar. 2016.

**Abstract:** This Master Thesis Project consists on studying the development of a fall detection video-based system, primarily intended for implementation in home environments to promote independent living for the elderly. Firstly, we conducted a comprehensive study of the art of existing methods in the detection of falls. After analyzing different algorithms, we chose the one that best fits the needs of our project and provides reasonable results in the detection of falls. Then, we implement an algorithm that characterizes the technique chosen. The

---

algorithm is evaluated by a different set of videos with different features and various approaches that allow us to draw conclusions and future research lines.

---

Análisis de actividad en vídeos deportivos multicámara (**Activity analysis in multicamera sports videos**), Ángel Mora Sánchez (advisor: Rafael Martín), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Universidad Autónoma de Madrid, Escuela Politécnica Superior, Mar. 2016.

**Abstract:** Sport video-content analysis systems are on the rise both from the commercial and the researching viewpoints. The previously existing VPULab system for sports analytics performs the detection and tracking of players in sport videos, and provides statistical information about their behavior, obtaining good quantitative and qualitative results, but presenting some deficiencies which motivate this mater thesis. These deficiencies were mainly related to aspects as usability, system interaction, results visualization and fine-tuning. This project is focused in providing a solution for those problems by working on three main tasks. Firstly, the project develops an unified prototype fully programmed in C++, full working and portable. Secondly, efforts are aimed to improve the usability, interactions and results visualization of the prototype. Two applications are developed and adapted to guarantee specific sports support, football and tennis. They allow a non-expert user to fully control the prototype and visually obtain its results, via a Graphical User Interface (GIU). Finally, keeping in mind that this tools use to work under supervision in commercial applications, the prototype and the interface have been equipped with tools to allow the online interaction with its results. This improvement allows a supervisor to control the application and correct its results when necessary, obtaining more reliable results than any other automatic system.

---

## Graduate thesis

Detección de personas en entornos multicámara utilizando información contextual (**People detection in multicamera environments using contextual information**), Alejandro Miguélez Sierra (advisor: Rafael Martín), Trabajo Fin de Grado, Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jan. 2016.

**Abstract:** Nowadays, people detection in video surveillance environments is an issue that has become one of the main tasks in computer vision. There are many object detection algorithms in the state of the art, but all of them present errors when used in scenarios where situations that hinder their operation occur: occlusions, illumination changes, perspective change, etc. Therefore, the main object of this Graduate Thesis is to include context information to people detection and combine information obtained from different cameras, to check if an improvement in results is obtained. All the stages of the system are studied

---

---

and implemented using a test dataset, obtaining a satisfactory approximation that improves the results obtained from each camera. To evaluate the performance of the algorithm, real detections from one of the algorithms of the state of the art are used.

---

The HAVideo (TEC2014-53176-R) project is supported by

