HAVi Newsletters

#01 & #02 & #3-

December 2015

TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

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

Welcome!

The provisional announcement of project approval arrived April 2015 and the final proposal of approval arrived September 2015 … whilst waiting for the notification of definitive approval, we have been working in the lines described in the proposal. Please, refer to section "First year progress report" for details about the rescheduling of the project.

Introducing HAVideo

The objective of this project is to tackle the problem of high availability video analysis, that is, the long-term analysis of video sequences. This project will focus on video-based understanding of people behaviour and it will investigate strategies to characterize the monitored scene, detect the existing people and identify their behaviour (e.g., movements and interactions) over long periods of time, either in single or multiple camera settings. The outcome of the project will provide a people behaviour description that can benefit a wide range of potential applications, nowadays mainly related to security (i.e., video surveillance) but also to people care (e.g., independent living) or monitoring (e.g., commercial areas monitoring) where long-term video analysis is a key issue.

This project assumes, based on our experience on the topic, that most of the recent approaches for people behaviour understanding perform well for short periods of time, for which they are designed, but they fail if applied for long periods of time (i.e., for hours or for a continuous operation). This limitation is due to the fact that the scene-changes over time are not correctly captured by the models employed for analysis and, therefore, they become outdated: models of scene background, which should be robust to multimodal situations; models of the targets or objects of interest, which should be

robust to scale and appearance changes; behaviour models, which should cope with point of views and occlusions; or context models, which for instance adapt alarms to the status of a changing traffic light. These scene-changes may affect all stages of the video analysis pipeline (e.g., object segmentation, feature extraction, target tracking, behaviour recognition), thus dramatically underperforming decreasing overall performance in long-time operation video sequences systems, and limiting their use in real and practical applications.

These situations require the design of new approaches for adapting the visual analysis tools to the scene dynamics. This project investigates approaches, either in single or multiple camera settings, to perform such adaptation by exploiting the scene context, a continuous self-performance analysis and the use of multiple and possibly also complementary sensors. The project will contribute with innovative approaches that incorporate using an online analysis of the results quality and contextual data to drive online adaptation of models and algorithms (e.g., configuration). Moreover, collaborative approaches will also be considered to provide feedback-based interaction between processing stages in single cameras or to coordinate multiple cameras operation to efficiently exploit for taking advantage of the multiple viewpoints and camera capabilities. Both adaptation and coordination will improve the system performance will allow the maximization of results performance and the optimal usage of resources for long-term analysis, thus enabling the high availability of video analysis tools.

The proposed research hypotheses are based on previous experience and on initial achievements of the proponent research Group, resulting from work in previous and ongoing research and development projects. A sustained effort in the proposal of methodological evaluation frameworks, including the generation of high quality video sequences with ground-truth data, will be the basis for a rigorous validation of the developed algorithms.

First year progress report

The HAVideo project started officially January 2015, nevertheless the official grant notification arrived in September. Nevertheless, the VPULab has been working during 2015 in the research lines proposed in the project. Therefore, the project started to produce results before the official announcement of the project start.
Whilst the workplan has been mainly followed, due to the delay in notifications, the publication dates of all the deliverables of the first year of the project have been rescheduled to December 2015. This decision implied a delay, not affecting the project progress, of deliverables D1.1, D1.2, D4.1 and D4.3v1 (edited together with v2). Also the HAVi Newsletters of year 2015 were published together December 2015 (numbers 1-3): this issue!

WP1 "Video Analysis Framework"

WP.1 activities were focused on the development of a virtual camera network simulator, the improvement of an existing camera network simulator to include embedded hardware modelling, sensor models and real communication protocols, and the improvement of an existing manual annotation tool, to provide semiautomatic tools that speed up the annotation process. With respect to the originally planned acquisitions, they have been postponed to the beginning of the second year in order to adjust both the notification delay and budget reduction. Deliverables D1.1v1 "System Infrastructure", D1.2v1 "DiVA documentation" and D1.3v1 "Simulator documentation" have been released December 2015.

WP2 "Video analysis tools, models and performance indicators"

The main WP2 achievements related to T2.1 "Analysis tools for human bahaviour understanding", T2.2 "Contextual modelling and extraction", and T2.3 "Quality Analysis" are:

- Development of a long-term stationary object detection based on spatio-temporal change detection.
- Development of a people density estimation in crowded scenarios and a people detection in groups.
- Development of a context-aware people detector which uses static contextual annotation to integrate and weight partial evidences from a part-based people detector algorithm.
- Development of a camera calibration algorithm to detect the relative position of a camera respect to another in wide-baseline scenarios.
- Development of a vehicle detection algorithm that uses on-line generated context to separate between object classes.
- Development of a tool and of its associated user interface to annotate static contextual objects in video sequences.

Additionally, several "Exploration and viability studies" (T2.4) were finished, dealing with heart rate detection using video, privacy preserving techniques in video surveillance and material identification through the Kinect technology.

Deliverables D2.1v1 "People Behaviour understanding in single and multiple camera settings" and D2.4v1 "Exploration and viability studies for people behaviour understanding" have been released December 2015.

WP3 "Self-configurable approaches for long-term analysis"

As scheduled, WP3 "was idle during the first year of the project.

WP4 "Evaluation framework, demonstrators and dissemination"

With respect to WP.4 "Evaluation framework, demonstrators and dissemination", within T4.1 "Evaluation framework", the project participated in the International VOT tracking challenge, developed a new dataset for wheelchairs users and standing people detection in a real in-door senior residence environment (the WUds http://www-vpu.eps.uam.es/DS/WUds/), and released D4.1v1 "Evaluation methodology and datasets" December 2015.
"Dissemination" (T4.3) officially started with the opening of the project web page in September 2015. Nevertheless, the first workshop was held May 2015 (due to lack of notification of evaluation, the HAVideo has not officially acknowledged) at the Escuela Politécnica Superior of the Universidad Autónoma de Madrid. The Workshop consisted of two parts: Dissemination Day (May 25th 2015): Video Analysis Technologies and applications for Computer Vision; and Short course (May 26th-28th 2015): Introduction to Computer Vision applications programming with OpenCV. D4.3v1-2 "Project Results" was released December 2015.

First year results

Datasets

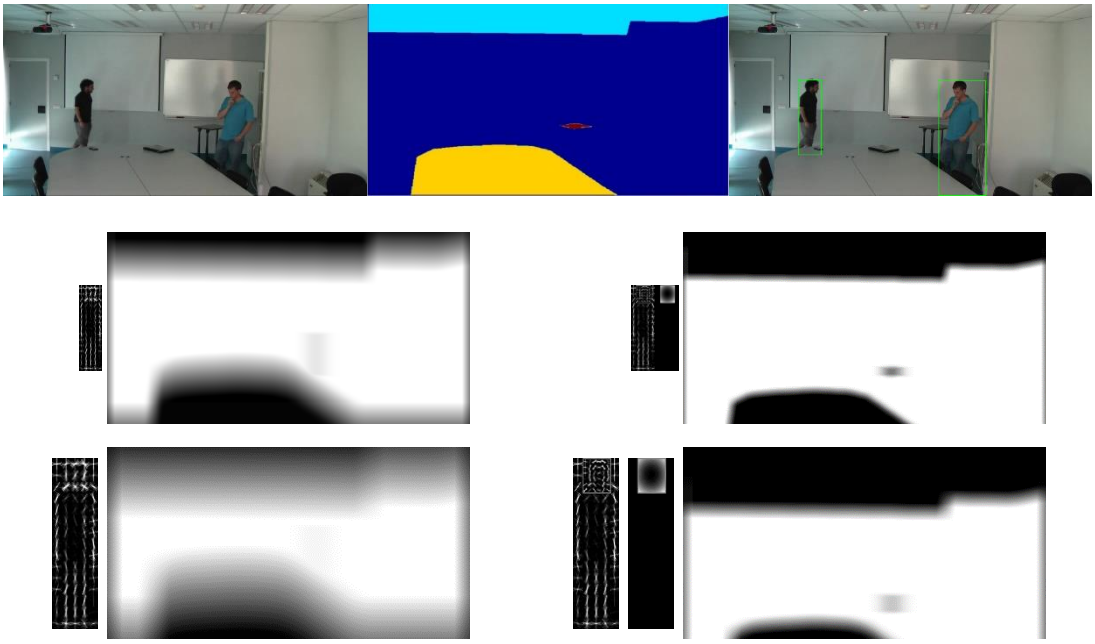
Wheelchair Users dataset – WUds

The WUds was recorded due to the lack of public wheelchair datasets. The sequences were recorded in a real environment of a senior residence, in order to work with an environment as realistic as possible (due to privacy issues, real recording with actual residents was not possible).
The dataset consists of 11 sequences (S1 to S11), each of them recorded from two points of views (V1 and V2), resulting in a total of 22 sequences. All sequences were recorded in the same room, using two GoPro cameras, HERO3 White edition. The fisheye effect was corrected using the GoPro Studio software tool.
The ground truth of this dataset was manually annotated for each frame of each sequence. The annotated ground truth considers the wheelchair users and the standing people present in every frame, even if they are highly occluded.
This dataset is freely available for research purposes and can be downloaded following the Content Sets link at the HAVideo web site.

Journals

Álvaro García-Martín, Juan C. SanMiguel, **Context-aware part-based people detection for video monitoring**, Electronic Letters, 51(23):1865:1867, Nov. 2015, IET, ISSN 0013-5194 (DOI 10.1049/el.2015.3099)

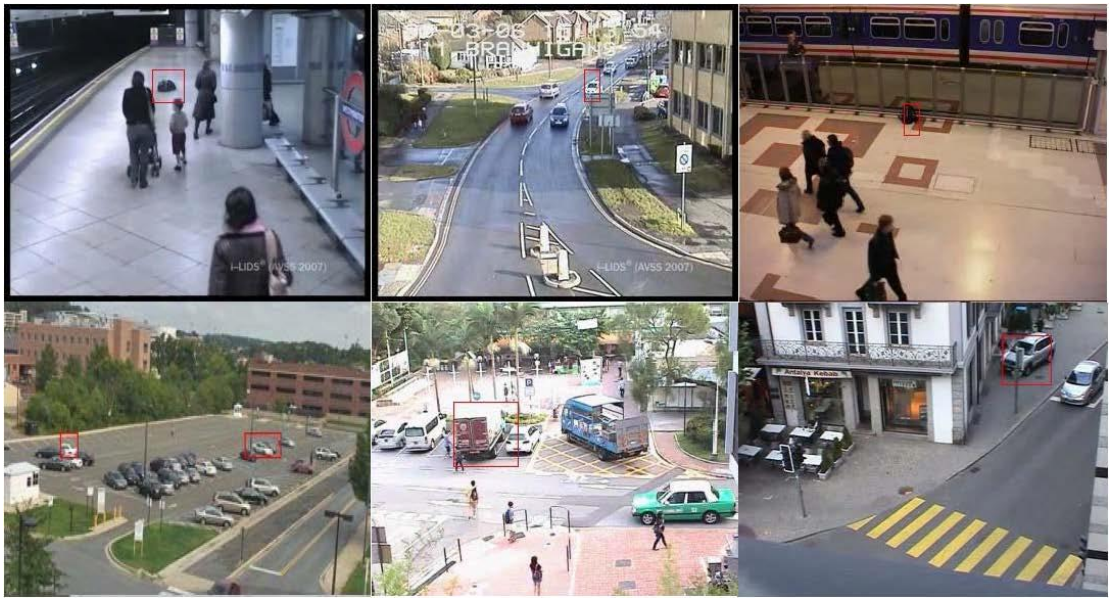
Abstract: A novel approach for part-based people detection in images that uses contextual information is proposed. Two sources of context are distinguished regarding the local (neighbour) information and the relative importance of the parts in the model. Local context determines part visibility which is derived from the spatial location of static objects in the scene and from the relation between scales of analysis and detection window sizes. Experimental results over various datasets show that the proposed use of context outperforms the related state-of-the-art.



Diego Ortego, Juan C. SanMiguel and José M. Martínez, **Long-term Stationary Object Detection based on spatio-temporal change detection**, IEEE Signal Processing Letters, 22(12):2368-2372, Dec. 2015. (DOI 10.1109/LSP.2015.2482598)

Abstract: We present a block-wise approach to detect stationary objects based on spatio-temporal change detection. First, block candidates are extracted by filtering out consecutive blocks containing moving objects. Then, an online clustering approach groups similar blocks at each spatial location over time via statistical variation of pixel ratios. The stability changes are identified by analyzing the relationships between the most repeated clusters at regular sampling instants. Finally, stationary objects are detected as those stability changes that exceed an alarm time and have not been visualized before. Unlike previous approaches making use of Background Subtraction, the proposed approach does not require foreground segmentation and provides robustness to illumination changes, crowds and

intermittent object motion. The experiments over an heterogeneous dataset demonstrate the ability of the proposed approach for short- and long-term operation while overcoming challenging issues.



Conferences

Matej Kristan et al., **The Visual Object Tracking VOT2015 challenge results**, Proc. of 3rd Visual Object Tracking Challenge Workshop at International Conference on Computer Vision, Santiago, Chile, December 2015, pp.564-586. (DOI 10.1109/ICCVW.2015.79)

Abstract: The Visual Object Tracking challenge 2015, VOT2015, aims at comparing short-term single-object visual trackers that do not apply pre-learned models of object appearance. Results of 62 trackers are presented. The number of tested trackers makes VOT 2015 the largest benchmark on short-term tracking to date. For each participating tracker, a short description is provided in the appendix. Features of the VOT2015 challenge that go beyond its VOT2014 predecessor are: (i) a new VOT2015 dataset twice as large as in VOT2014 with full annotation of targets by rotated bounding boxes and per-frame attribute, (ii) extensions of the VOT2014 evaluation methodology by introduction of a new performance measure. The dataset, the evaluation kit as well as the results are publicly available at the challenge website.

Michael Felsberg et al., **The Visual Object Tracking VOT-TIR2015 challenge results**, Proc. of 3rd Visual Object Tracking Challenge Workshop at International Conference on Computer Vision, Santiago, Chile, December 2015, pp.639-651. (DOI 10.1109/ICCVW.2015.86)

Abstract: The Thermal Infrared Visual Object Tracking challenge 2015, VOT-TIR2015, aims at comparing short-term single-object visual trackers that work on thermal infrared (TIR) sequences and do not apply pre-learned models of object appearance. VOT-TIR2015 is the first benchmark on short-term tracking in TIR sequences. Results of 24 trackers are presented. For each participating tracker, a short description is provided in the appendix. The VOT-TIR2015 challenge is based on the VOT2013 challenge, but introduces the following novelties: (i) the newly collected LTIR (Linköping TIR) dataset is used, (ii) the VOT2013 attributes are adapted to TIR data, (iii) the evaluation is performed using insights gained during VOT2013 and VOT2014 and is similar to VOT2015.

Master thesis

Estimación de la densidad de personas en entornos densamente poblados (**Estimation of people density in crowded enviroments**), Rosely Sánchez (advisor: Álvaro García-Martín, supervisior: José M. Martínez), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, May 2015.

Abstract: Currently, the use of computer vision systems has acquired great relevance thanks to the advance of digital image and video processing technologies. Crowd density estimation is an important tool to detect abnormal situations in public places such as fights, disturbances, violent protests, panic situations 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 its evolution over time. So far, many crowd density estimators have been implemented, most of them using foreground-background segmentation and extracting features from foreground pixels, getting good results for the studied scenarios. This project introduces the use of people-background 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 will be implemented and results for both types of segmentation, foreground-background and people-background segmentation, are compared in several scenarios.

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, Univ. Autónoma de Madrid, Sep. 2015.

Abstract: Computer vision is a field in computer science, which tries to provide machines with that appreciable sense that is sight. This coupled with artificial intelligence, where you want a computer to have consciousness, can make automatically and autonomously that computers, for example, monitor people safety. Furthermore, video surveillance can process different types of inputs to generate alerts that a user can use to avoid an unsatisfactory event This project involves, therefore, a junction of these issues. Our final aim is to detect people effectively in crowd scenes, where detection is even more complex detection, since people are highly occluded with other persons. Specific software is developed as a prototype that has all the desired features.

Preservación de privacidad de personas en vídeo-seguridad (**People privacy preservation in video-surveillance**), Jaime Mateo Herrero (advisor: José M. Martínez), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Universidad Autónoma de Madrid, Escuela Politécnica Superior, Oct. 2015.

Abstract: Privacy is a main open issue in video surveillance systems. The unstoppable growth of these systems in many scenarios of our daily life makes necessary the development of techniques that permit hiding personal features of those individuals involved in video-surveillance sequences, providing that the original sequence can be restored for forensic use with the corresponding judicial authorization. Firstly, an exhaustive study of state of the art is performed. Existing privacy preserving techniques, those reversible and those that are not, are detailed. After this, a technique that respects privacy, reversibility and coding losses' boundaries is selected. Later on, an algorithm that performs the selected technique is programmed and evaluated with several person detectors in order to study the behavior, introducing alternatives to improve its performance as well.

Graduate thesis

Detección de ritmo cardíaco mediante vídeo (**Heart rate detection using video**), Erik Velasco Salido (advisor: José M. Martínez), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

Abstract: The objective of this Graduate Thesis is to design and develop an algorithm that allows us to detect the heart rate by analyzing natural color, segmentation masks and depth video sequences. Using these last kind of sequences allows the preservation of the privacy of the monitored person and working in conditions of low or no lighting. A new approach has been designed and implemented, starting from a base algorithm based on previous work in which we rely. After implementing and validating the base algorithm on color natural sequences, the new approach has been designed and developed: an algorithm working on color sequences, as the base algorithm, but also on masks segmentation and depth sequences provided by the Kinect camera. We have analyzed the performance of different modalities depending on the distance of the camera to the person in order to assess the feasibility of a possible combined system using the various modalities supported by the proposed algorithm. To validate the algorithm, a dataset has been recorded, composed of natural color sequences, segmentation masks sequences and depth sequences, in addition to recording of the heart rate measured by a heart rate monitor. Using this dataset a complete set of results in the different situations under study has been obtained.

Identificación automática de materiales usando el sensor Kinect (**Material identification through the Kinect technology**), Alejandro López Cifuentes (advisor: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

Abstract: This project has as main objectives the design and capture an image database and the development of a system capable to train a model and classify a set of previously captured images. The aim of this process is to develop a prototype that performs an approximation to automatic material recognition through the Kinect sensor. In order to achieve these objectives several stages need to be first fulfilled. A study about recent versions of the Kinect sensor is first required. Then, existing methods for material characterization are reviewed, mainly understood as versions of the Bidirectional Reflectance Distribution Function (BRDF); in particular Histogram of Oriented Gradients (HoG) can be seen as a two-dimensional approximation to BRDF. The generation of knowledge models via Support Vector Machines (SVM) is also analyzed. Datasets currently

available for research have been analyzed and a new dataset has been recorded using Microsoft Kinect, capturing a materials selection from different capture and illumination incidence angles. In order to include additional information to colour images, depth and infrared images are also included in the dataset. Each material image has an associated patch which aim is to isolate the material from its background. Once the dataset has been recorded a system has been designed, using GDF-HOG and EigHess-HOG descriptors in order to extract patch characterization. SVM knowledge models have been generated, which allow to predict and classify untrained input instances. Finally, system performance has been evaluated achieving good results in supervised situations and promising results in real scenes.

Detección de la posición relativa de una cámara en situaciones de grandes desplazamientos (**Detection of camera relative position in wide-baseline scenarios**), María Narváez Encinal (advisor: Marcos Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

Abstract: This project aims to design and develop a set of algorithms able to cope with the problem of modelling the epipolar geometry between two cameras capturing the same scene. The ultimate goal of this process is the automatic estimation of the relative position between both cameras. Whereas the epipolar geometry is fully explained through the fundamental matrix, the relative position may be extracted, up to a scene homography, from the fundamental matrix. To fulfill the faced task, a study of the image capture and camera calibration processes is first required. Automatic calibration strategies usually rely in a previous set of inter-image correspondences, hence, the analysis of this research topic is also required. Specifically, Hessian-based detectors on the Scale- Space are explored for the feature detection stage, SIFT-DSIFT and DAISY descriptors are assessed to describe the extracted features, whereas the NRDC algorithm is evaluated as an alternative for feature matching. The use of these feature-matching pipeline leads in conjunction with the RANSAC estimator to an initial estimation of the fundamental matrix. Afterwards, such matrix is refined by a set of strategies designed and developed during this project. Moreover, a Graphical User Interface (GUI) which eases the manually introduction of reliable correspondences is also designed and developed. The aim of this GUI is to face scenarios where the feature-matching pipeline is not able to produce an initial estimation of enough quality. Finally, both the initial estimates of the fundamental matrix and its subsequent refinements are evaluated by means of different experiments and error measurements.

Detección de vehículos mediante el análisis de imágenes (**Image-based vehicle detection**), Juan Ignacio Bravo (advisor: Luis Salgado), Trabajo Fin de Grado

(Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jun. 2015.

Abstract:

The goal of this Graduate Thesis is the implementation and analysis of a system able to segment a video flow generated by a forward looking camera placed in the frontal of a vehicle, with the aim of detecting vehicles, pavement and lane marks. The system handles prior information of the scene to generate statistical models of the regions of interest, which will be optimized through the Expectation-Maximization algorithm to classify the images using the Bayesian decision theory. All the stages of the system are studied and implemented, contributing also with proposals to improve the performance of some of them. Subsequently, to evaluate the performance of the algorithm and the validity of the improvements, specific tests for each class are designed, where three different versions of the systems are compared against a ground-truth created specifically for this purpose. Finally all results are discussed and some guidelines for future work are given.

Simulador virtual para sistemas multi-cámara distribuidos (**Distributed multicamera systems virtual simualtor**), Luis Pérez Llorente (advisor: Juan C. San Miguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jul. 2015.

Abstract: The main objective of this Graduate Thesis is the development of a system that is able to simulate the management of multiple cameras taking visual information from 3D virtual environments. This system enables to define test environments in order to research on computer vision algorithms, for example people tracking or detection algorithms recorded in a video, which allows to repeat the situation as often as necessary. The development of the system is based on the Virtual Video software, increasing the number of features that the tool has in order to allow the user to create, cancel and modify the location and the visual field of different cameras simultaneously. In addition, the system shows a real-time viewing of the multiple cameras that were installed. The virtual environment in which the cameras are be installed simulate in a very realistic way different kinds of scenarios, which can be generated with 3D map designing tools (v.gr. Hammer). Taking into account the required resources, a performance analysis of the developed system is done.

Desarrollo de herramienta para la anotación manual de secuencias de vídeo (**Developement of a video sequences manual annotation tool**), Yoel Witmaar (advisor: Juan C. San Miguel), Trabajo Fin de Grado (Graduate Thesis), Grado en

Ingeniería Informática, Escuela Politécnica Superior, Universidad Autónoma de Madrid, Jul. 2015.

Abstract: Annotating images and videos is something important for those who work in image processing, but it is a hard task to do it only manually. That is why it is necessary to develop annotation tools which make this work easier, so that the program uses implemented algorithms which help to do the image and video annotation more efficient. The first step of this work is to investigate actual tools and evaluate their features. With this information you can get good ideas about which features are going to help the users more for annotating with more effectiveness or annotating any image or video format it is needed to. The next step is choosing a tool for a start, one which performs manual annotations at least, so more functionalities can be added, like semiautomatic and automatic annotations. It is also an aim making the tool support more image and video formats. The most effective thing for improving a tool's annotation efficiency is to add it automatic annotations which use implemented object detection algorithms. It's also worth to add it annotation propagation mechanisms, which can use implemented algorithms too. An annotation tool developed by the VPULab, which is capable of performing manual annotations and doing automatic pre-segmentation, has been improved including new functionalities allowing the tool to read videos frame by frame for annotating them, on any format; to do automatic annotations using object detectors, so that they detect likely objects and the user can choose which are correct for annotating them, and using a frame by frame manual annotation propagation mechanism, which consists in moving an annotation while the frames are changing.

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.

Abstract: Video signal processing has undergone a constant evolution in the past few years. It may have happened because of its relevance in such a growing field as video security, therefore it has become so interesting the research in this area so that we innovate with new techniques or improve the existing ones. 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 project is the implementation of a people detection in groups algorithm in C++, as well as the integration on a proprietary video analysis platform called DiVA. This display lets the execution with 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, is an adaptation of the algorithm known as DTDP (Discriminatively Trained Part Based Models) 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.

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