



HAVi Newsletters

#05 - August 2016

TEC2014-53176-R HAVideo (2015-2017)

High Availability Video Analysis for People Behaviour Understanding

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

Fifth four-months period progress report

Again, the workplan has been slightly modified; in this case instead of the delays announced in HAVi Newsletter #4, in June 2016 we have decided to extend Task 2.2 till month 30 (June 2017) in order to further work on context in multi-camera setups.

Besides this, the project has been working on plan (below, the main achievements from each workpackage are enumerated) and all the deliverables due June 2016 have been published on schedule: D2.2 "Contextual modelling and extraction for people behaviour understanding" version 1, D2.3 "Online quality analysis of people behaviour understanding" version 1, and the semestral D4.3 "Results Report" version 3.

WP1 "Video Analysis Framework"

For workpackage WP.1, besides some equipment acquisitions, the work has been focused on improving the virtual camera network simulator using the Unity tool. This work is expected to be completed in the fourth semester.

WP2 "Video analysis tools, models and performance indicators"

For workpackage WP.2, the main achievements are:

- Development of a measure to estimate tracking quality based on reverse trajectory analysis employed for combining multiple algorithms
- Development of a new Visual Attention Model enhancing discrimination among objects and its application to tracking
- Enhancements to a long-term teacher-tracking system for the real-time distribution of classroom activities
- Development of a robust to appearance changes tracker based on the use of a state-of-the-art point-of-interest detector
- Study to include context information to people detection and combine information obtained from different cameras

-
- Improvement of an existing tool and of its associated user interface to propagate the annotation of static and dynamic contextual objects in video sequences
 - Development of an online panoramic background initialization algorithm for PTZ cameras
 - Study of online people detection algorithms quality analysis using correlation metrics
 - Development of a Background Estimation algorithm for video sequences robust to stationary objects
 - Development of a long-term abandoned object detector robust against sudden illumination changes and stationary pedestrians

WP3 “Self-configurable approaches for long-term analysis”

As scheduled, WP3 is still idle.

WP4 “Evaluation framework, demonstrators and dissemination”

For workpackage WP.4, the main achievements are:

- Development of an evaluation framework and a comparative analysis of state of the art local features detection and description algorithms
- Generation of a dataset containing 4 categories or challenges with 10 video sequences for the task of Background Estimation
- International tracking challenge (VOT challenge) evaluation (results will be available in the next months)

The 2016 Developers Workshop was held May 2016 at the Escuela Politécnica Superior of the Universidad Autónoma de Madrid. The Workshop consisted of two parts:

- Dissemination Day (May 23rd 2016): Video Analysis Technologies and applications for Computer Vision.
- Short course: (May 24th–26th 2016): Introduction to Computer Vision applications programming with OpenCV

There were 16 participants.

Fifth four-months period results

Datasets

Background Estimation dataset – BEds

The Background Estimation dataset (BEds) proposes a corpus of video sequences

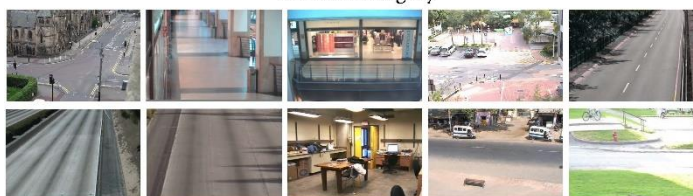
generated from publicly available video-surveillance datasets to cover several Background Estimation challenges. This dataset has been created to allow a complete evaluation of Background Estimation algorithms by providing a set of sequences that cover a variety of challenges.

In particular, BEds contains four categories or challenges: Baseline, that contains simple sequences with few background occlusions; Clutter, that contains sequences with continuous background occlusions; Low framerate, that contains simple sequences recorded with low framerate; and Static objects, that contains sequences with stationary objects remaining stopped for more than the 50% of the sequence duration.

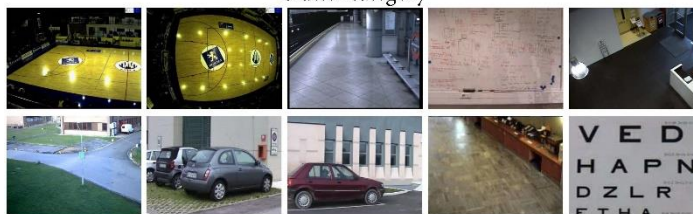
The ground-truth of this dataset consist of background images that algorithms are expected to build. If possible, these images have been created by selecting instants where the background is entirely visible. Otherwise, several background areas from different temporal instants have been combined to conform the background image.

This dataset is freely available for research purposes and can be downloaded following the Content Sets link at the HAVideo web site.

Baseline category



Clutter category



Low framerate category



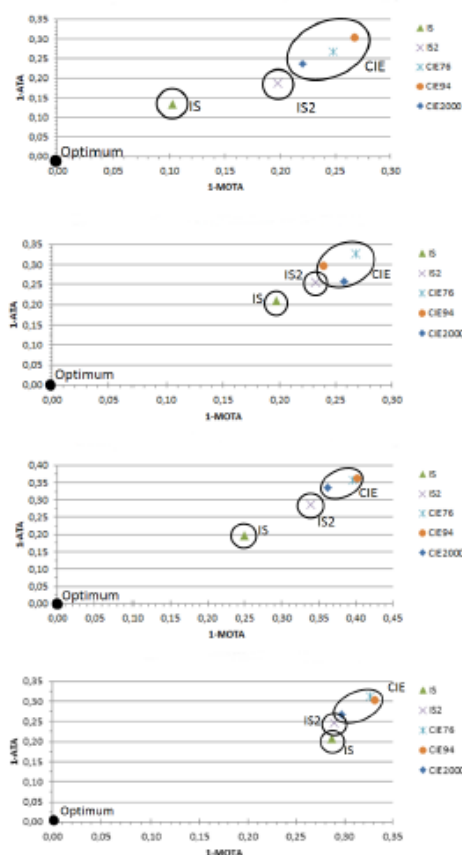
Static objects category



Journals

Víctor Fernández-Carbajales, Miguel Ángel García y José M. Martínez, “**Visual Attention Based on a Joint Perceptual Space of Color and Brightness for Improved Video Tracking**”, Pattern Recognition, 60:571–584, Dec. 2016 (online June 2016), Elsevier, ISSN 0031–3203 (DOI [10.1016/j.patcog.2016.06.007](https://doi.org/10.1016/j.patcog.2016.06.007))

Abstract: This paper proposes a new visual attention model based on a joint perceptual space of both color and brightness, and shows that this model is able to extract more discriminant visual features, especially when dealing with objects that are very similar visually. That joint color and brightness space is based on a biologically inspired theoretical perceptual model originally proposed by Izmailov and Sokolov in the scope of psychophysics. The present paper proposes a computational model that allows the application of Izmailov and Sokolov's theoretical model to digital images, since the original model can only be applied to perceptual data directly drawn from psychophysical experiments. Experimental results with real video sequences show that the proposed visual attention model yields significantly more accurate results in the particular application scope of video tracking than well-known visual attention models that process color and brightness separately.



Master thesis

Estimación de fondo de escena en secuencias de vídeo (**Background estimation in video sequences**), Jaime Gallo de Cal (advisor: Diego Ortego), Proyecto fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: In this work, we propose a block-wise and online Background Initialization algorithm for static camera video sequences based in a spatio-temporal scene analysis. First, the state-of-the-art is studied to understand the Background Initialization task and the challenges that it poses. In particular, previous techniques previously used in the laboratory VPULab where specially

considered. Then, the work focuses in incorporating a temporal online analysis proposed in the literature into the proposed algorithm. Furthermore, the initial stage of the spatial background reconstruction, the seed selection or partial background initialization stage, has been improved leading to more initial information before the entire background reconstruction. Additionally, a new measure of spatial continuity is proposed to perform the complete background reconstruction by using both information of the block borders and its interior. Finally, experiments are carried out to validate each stage of the developed algorithm and the performance achieved against related literature, thus understanding the algorithm advantages and weaknesses see definition of weakness.

Evaluación comparativa de técnicas de detección y descripción de puntos de interés en imágenes (**Comparative evaluation of Points of Interest techniques for detection and description in images**), Miguel Martín Redondo (advisor: Fulgencio Navarro), Proyecto fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: One of the main objectives of this project has been to develop a framework for the evaluation of keypoint detection and description algorithms, in order to update the references of the state of the art in this topic. Based on this framework, the other main objective has been to perform a comparative evaluation of the state of the art algorithms in local features field. To this aim, three main stages have been faced. The first one has been to propose a new dataset, together with a evaluation methodology, based on an analysis of the strengths and weaknesses of previous frameworks in the state of the art. This proposal will set a new evaluation framework for the following stages. The second one, an exhaustive study of the state of the art allowed to select the main techniques and categorize them according to its properties. Finally, those selected techniques were tested on the proposed evaluation framework. Summarizing, at the end of the project, both objectives, the evaluation framework proposal, and the comparative evaluation, have been satisfied.

Automatización de funciones en el seguimiento del profesor para la emisión de clases presenciales, Alberto Palero Almazán (**Automatization of functions for teacher tracking in lectures broadcasting**), (advisor: Jesús Bescós), Trabajo Fin de Máster (Master Thesis), Master en Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: The main motivation behind this Project has been to automate the process of tracking a teacher in a classroom so as to be able to broadcast classes via the internet to students who, for various reasons, can't physically attend the class that is being taught. This project is a continuation of several previous

works, as a result of which there is an algorithm that uses images from a video sequence taken by a fixed camera to track, in real time, the position of a teacher in a classroom, and guides in that direction a mobile camera whose video signal is the image the student will see. Furthermore, there is a web application that allows users to view these classes. The integration of an HOG people detector in the original algorithm is detailed in this project. This detector is useful so as to be able to automate the initialization phase, and it allows a faster recovery of the target once the algorithm determines that it has lost it. Once the changes that have been made to the tracking algorithm have been explained, two new proposals that improve the mobile camera movements are introduced. With these proposed solutions better movement control and zoom levels are obtained, making seem as if a person is controlling the mobile camera. Of these approximations, one is based on the original rule scheme, while the other attempts to predict the target's future position using a Kalman filter. Finally, the initialization of the algorithm, the recovery of the target and the mobile camera's movement control have been evaluated, and the appropriate conclusions have been drawn.

Reconocimiento de actividades utilizando información de color y profundidad (**Activity recognition using color and depth information**), Borja Olmo Esteban (advisor: Juan C. SanMiguel), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: Providing computer systems with human capabilities has been a major field of interest and study since the creation of the first computer systems. Due to this, a large number of studies have emerged in last two decades to provide systems with automatic recognition capabilities exclusively on visual information. With the recent advances in depth sensor hardware and imaging, a large number of depth-based image research has appeared, especially with the release of the low-cost sensor Kinect developed by Microsoft. Such sensors provide a new source of information which does not suffer the same problems as RGB data such as lighting changes. The main goal of this project has been to develop a human activity recognition system which allows the integration of different techniques that make use of both RGB and depth data in order to compare them and find out which set of settings provides the best recognition results. The created system has been applied to various state-of-the-art datasets, achieving competitive results with some of the proposed configurations. Moreover, the results allow to discriminate which features are useful for activity recognition.

Integración y evaluación de sistemas de robo–abandono de objetos en video–seguridad (**Integration and evaluation of stolen–abandoned object systems in video–surveillance**), Jorge Gómez Vicente (advisor: Juan C. SanMiguel), Proyecto Fin de Carrera (Master Thesis), Ingeniería de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: This work proposes a configurable abandoned–stolen object detection system in security–video that integrates the most relevant techniques in each one of its stages. A formalization of the problem is presented, followed by a description of the different analysis stages required for the detection. Firstly, this work analyses the state of the art to know the present date problems about the matter. Secondly, the work focuses on the integration of the most recent and relevant algorithms of the literature in every single phase of the system. It also designs the necessary interfaces for its execution in a sequential order. To conclude, the different configurations of the system regarding the detection of static regions as well are evaluated and compared, while it is classified as abandoned–stolen about a compound of heterogeneous videos sequences. Abandoned/stolen object event detection depends on the parameters that modulate the absorption of the blobs from the static foreground by the background model. The results of abandoned/stolen detection stage are affected by the propagation of errors in the earlier stages as it is the last stage of video analysis system.

Graduate thesis

Detección de personas en entornos residenciales y hospitalarios (**People detection in residential and hospital environments**), Jesús Molina Merchán (advisor: Rafael Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2016.

Abstract: There is a large demand in the area of video–surveillance, especially in detecting people, which has caused a large increase in the number of investigations in this field. In the field of detection elderly people, the detector must have into account different positions such as sitting or in a wheelchair. Also is important the cost involved in making these detectors. Therefore, this work has two main objectives. The first has been to develop a model person sitting with the aim of completing a detector on the stage of a nursing home. The second one was based on reducing the amount of resources needed and save the cost of having to record sequences for a detector in this scenario. To achieve this, three 'synthetic images dataset were created in order to perform three different models, evaluating which model is optimal and finally analyzing its feasibility by comparing it with the people detector in wheelchairs.

Detección de objetos abandonados para vídeo-vigilancia a largo plazo (**Abandoned object detection for long-term video-surveillance**), Sergio López Álvarez (advisor: Diego Ortego), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2016.

Abstract: The main objective of this project has been the integration of an abandoned object detection algorithm on a distributed video analysis platform called DiVA. To achieve this, and also improve the processing speed of the system, the base algorithm's operation and all the platforms and tools used in this project have been deeply studied, in order to implement its functionality on C++ later, adapting it so that the system could operate not only using off-line videos, but also in real time. Moreover, this project also proposes a new abandoned objects detection algorithm that improves the capabilities of the base system. These improvements allow the new system to filter false detections using entropy theory and also avoid detections due to static humans. Finally, the performance of the system has been thoroughly evaluated, both according to computational cost and the precision of the detections, proving a noticeable improvement in both aspects.

Reconstrucción de fondo de escena a partir de secuencias de vídeo (**Background reconstruction from video sequences**), Carolina Fernández-Pedraza (advisor: Diego Ortego), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: In this work, several algorithms for background initialization are proposed, all based on temporal and spatial information. First, the state-of-art has been explored. Then, two algorithms have been implemented, one of them uses block-division and the other one pixel-division, where smoothness has been computed to evaluate what spatial region fits best in the new background: the new region of the actual frame or the region that already exists in the background. Finally, these two algorithms and other two that have been already implemented, will be evaluated based on a dataset that is classified in several challenges to overcome for each algorithm.

Detección de personas utilizando contexto de la escena (**People detection using scene context**), Carlos Chaparro Pozo (advisor: Álvaro García), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: At present, video signal processing is immersed in a constant evolution. The scope of it is wide and includes many fields as, for example, video security. This one is a very relevant and current area. Therefore, research in this area is

vital to improve the existing techniques of analysis. In video security, people detection is one of the most complex tasks. People detection faces many challenges related to pose changes, illumination variations and occlusions. When some of the previous characteristics are present and the scene is complex, the existing algorithms reduce considerably its performance. For it, the aim of this project is the implementation of a framework that improves the performance and the functionality of people detection using context of the scene. At this moment, the models proposed to cope the environments with great variability and complexity need to re-train to offer an optimal performance to a concrete case. In this project, we have implemented a framework in C++ that allows to the users annotate the occlusive zones of the scene, avoiding to re-train the model and improving its functionality and performance. The algorithm implemented is an adaptation of the algorithm known as DTDP (Discriminatively Trained Part Based Models). This allows us to use the context to enhance the people detection related to the occlusive parts of the scene and the relation between the scales of analysis and the size of the detection windows. In order to complete our aim, we realized an evaluation of the implemented framework, comparing it with DTDP detector. In this manner, we can verify that the framework is more efficient, besides the improvement regarding the functionality.

Seguimiento de objetos mediante constricción espacial de puntos TILDE (**Object tracking via spatial constraining of TILDE points**), César Augusto Betancur Cruz (advisor: Marcos Escudero), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jun. 2016.

Abstract: The aim of this work is the design and development of a video object tracker based on point-of-interest (PoI). Specifically, we focus on the potential benefits of including a recently published PoI detector in the core of the tracking process. The design of the algorithm starts by studying existing tracking algorithm emphasizing in methods using PoI in any stage of their tracking algorithm. In order to provide a flexible framework on which to develop potential improvements, the algorithm's design builds on the basic PoI-based tracking scheme. From the state-of-the-art, we propose to use TILDE (Temporally Invariant Learned Detector) as the PoI detection method. TILDE is a train-based PoI detection method which is claimed to provide stable results to change in illumination and appearance. TILDE-driven correspondences are used to spatially constrain the target position between consecutive frames. Final result is refined by means of a classic cross-correlation method. The tracker is experimentally evaluated in a generic evaluation corpus. From this evaluation we aim to discuss on their benefits and drawbacks. Furthermore, we also compare the tracker against state-of-the-art trackers, in order to quantitatively contextualize its operation. Experimental results partially validate the design and development of

the algorithm and suggest that the use of TILDE may help to generate robust constraining schemes for trackers based on cross-correlation.

Diseño y desarrollo de una herramienta para la extracción semi-supervisada de la información de contexto (**Design and development of a tool for semi-automatic extraction of contextual information**), Raúl García Jiménez (advisor: Marcos Escudero), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: A contextual annotation tool has been designed and developed. The tool is specially focused on the annotations of contextual outdoor (e.g. roads, trees or sea) and indoor areas (e.g. furniture or decoration, etc.). Existing tools are designed to ease the annotation of objects of interest in video-analysis applications, e.g.. people or cars, but do not considered the annotation of the contextual objects that surround them. Proposed design builds on an existing annotation tool, which we name base tool. The base tool is modified in order to minimize the number of required user interactions while maintaining the degree of usability of the annotations in similar terms to those obtained via the base tool. To this aim, we take advantage of the intrinsic characteristic of contextual objects: gradual and moderate temporal variation in terms of appearance, shape and position. In general terms, the aim is to propagate an initial user annotation in the first frame along the whole video. User-tool interaction strategies are also included to cope with system failures and new object appearances. These interaction strategies define the designed tool as a semi-supervised annotation tool. The process can be sketched as follows. First, each video frame is segmented in color homogeneous areas via a state of the art region segmentation method. Then, in the first frame and in every frame on which the interaction strategies are activated, the user has to group these regions into regions of interest. These regions on interest are entities of a higher semantic level (close to the object level). This process is devoted to adapt the spatial area to the contours of a particular contextual object. The so-obtained regions of interest are then user-labelled as members of a contextual class. Obtained annotations are used to propagate the information on the next frame by automatically grouping regions without the requirement of user interaction. The so-obtained annotations are subsequently propagated to the next video frame and so on until the end of the video. Through this naive strategy, the tool is potentially able to drastically reduce the number of required user interactions. Furthermore, due to the region-driven scheme, is also able to adapt the annotations to gradual changes of shape, appearance and position of the annotated contextual objects. In order to evaluate the goodness of the designed tool respect to the base tool, two different evaluations have been performed. The first one measures the number of user interactions required in the annotation of a set of sequences when using each of the tools. The second evaluates the quality of the propagated annotations respect to those obtained manually. Results of both experiments

support the tool design and development. Finally, the quality of the user subjective experience is evaluated through a questionnaire. Responses of 10 users suggest that the designed tool ease the annotation of contextual objects in different and varied scenarios.

Diseño de herramientas de apoyo para la detección de logotipos en secuencias de video (**Design of supporting tools for logo detection in video sequences**), Efrén Martín García (advisor: Marcos Escudero), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería Informática, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: This work describes an automatic method for the detection of brand logos in sport sequences. The work starts by studying the solutions existing in the state-of-the-art in the topic. From this study a set of conclusions is derived, and these are used to define the design of the proposed method. The method starts by defining three pre-processing methods which—motivated by design-heuristics—determine the spatial areas on which a logo is prone to be placed. Specifically, the methods use color, structural and saliency based strategies to constrain the areas on which the logo detection process takes place. On the candidate areas—those prone to contain a logo—, a classical point-of-interest matching strategy is used to relate the candidate instances with a preload logo template. From these matches, an affine correction of the template is derived. Logos are detected by measuring the similarity between the transformed template and the candidate areas. Experimental results on a set of candidate sequences partially validate the design and development of the method for soccer sequences. However, results also illustrate the method’s drawbacks and limitations when analyzing sequences of alternative sports. Furthermore, preliminary experiments on the use of the method for the generation of publicity statistics are also included, obtaining promising results. In overall, results suggest that the use of pre-processing techniques may help in the task of automatic logo detection.

Seguimiento automático de objetivos con drones mediante algoritmos de tracking (**Automatic tracking of targets with drones using tracking algorithms**), Roi Rico Díaz (advisor: Fulgencio Navarro), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: The main objective of this Graduate Thesis is to take advantage of the characteristics of using mobile cameras instead of fixed cameras for video-surveillance purposes like automatic tracking. For that purpose, a model of the problems introduced by combining the capture conditions in mobile cameras with the challenges tracking algorithms have to face up has been proposed. Once the problems have been defined, it continues by evaluating them in a result manner on the tracking algorithms from the state of the art through a toy-

example strategy with an own and synthetic dataset. After modeling the effects, two modifications oriented to a reduction of the blurriness in the sequences and a stabilization of the camera by using points of interest have been proposed. These proposals sought to alleviate the effects of the detected problems. For evaluating the desired effects of the improvements the system has been evaluated on the synthetic dataset, and once these improvements have been corroborated, a real dataset with sequences captured by an UAV has been generated in order to evaluate the proposal in a real environment. The obtained results allow to conclude that the objective proposed at the beginning of this thesis has been accomplished with satisfaction. Additionally, it has been considered that this project could be the starting point for vision applications on capture devices aboard UAVs.

Control de cámaras PTZ para la reconstrucción de escena basada en puntos de interés (PTZ cameras control for scene reconstruction based on Points of Interest), Elisa Martín Pérez (advisor: Fulgencio Navarro), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: The aim of this Final Degree Thesis is background initialization through the creation of a panoramic image, under the circumstances of a mobile camera and foreground presence. Since the starting point is a previously implemented system, it is necessary to check its performance in order to choose a strategy to follow. Therefore, a specific dataset has been developed to evaluate said system. Afterwards, several improvements are proposed. Specifically, error presence (or situations in which those could take place) is detected, and some solutions are suggested to achieve a correct background initialization. Thus, the new system contains two differentiated modules: error detection module and error correction module. Two different approaches of this second one are presented. Lastly, the modified system is evaluated and compared with the initial one, to check whether there is an improvement. The two different improvements are compared against each other as well. The conclusion of this project is that the initial system needed to be improved in order to achieve a proper background generation. Moreover, not only the proposed improvements solve the original problem, but could serve as a basis for new improvements with the same approach.

Seguimiento de objetos basado en múltiples algoritmos (Object tracking based on multiple algorithms), Eduardo Moreno de Pablos, (advisor: Juan C. SanMiguel Avedillo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2016.

Abstract: The main objective of this thesis consists on tracking objects in video sequences using a set of search or tracking algorithms. First, an initial study of

existing works is carried out with emphasis on the basic aspects of detection, search, tracking and the combination of various search algorithms. Then, a set of scientific papers that are focused on the same topic have been studied. These papers use different techniques to achieve the desired objective. After this theoretical learning stage, the Matlab tool is used to test a set of selected algorithms and sequences, comparing the effectiveness of the algorithms by the error based on ground-truth also to compare them with a series of measures, or comparing an algorithm himself in different time instants to estimate their reliability. These measures employ the distance between centers and the overlap between areas detected in the corresponding frame. These measures give us an ability to visualize which algorithms are optimal in a possible combination of these, and which are not. Finally, confidence maps are used to estimate results' stability in the possible selection of the optimal algorithms for their combination. These maps are extracted from each algorithm and represent areas of a size similar to the frames which show the probability that the object is located at each point of the map. Experimental results show the strengths and weaknesses of each algorithm on the set of selected sequences using the proposed set of reliability measures. Once the results obtained from the measures described above, we proceed to combine the best algorithms as proposed.

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

