HAVi Newsletters

#11 – December 2018

TEC2014-53176-R HAVideo (2015-2017-2018)

High Availability Video Analysis for People Behaviour Understanding

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

Extension period second semester progress report

During this final period several publications have been accepted and published. As planned in the request for the extension period, the main activity has been focused in the development of several applications and demonstrators within WP4:

1. Pedestrian density estimation in multi-camera scenarios
2. Detection of potential distractor in video object tracking
3. People detection evaluation
4. Multi-camera pedestrian detection benchmark
5. Multi-camera video surveillance based on smartphones
6. Live streaming of classes and seminars using automatic video tracking
7. Visualizer and controller for Multi-camera System Simulator (MSS simulator)

These applications have been reported in D4.4.2v2 "Applications" and are briefly described in the corresponding results subsection.

Also, the final version of the report summarizing project results, D4.3.v7 "Results Report", has been published.

The 2018 Dissemination Workshop was held November 2018 at the Escuela Politécnica Superior of the Universidad Autónoma de Madrid.
The Workshop consisted of four parts:

- Project overview and presentation of results
- Presentation of potential extension of results to other areas
- Presentation by companies of required technologies and challenges
- Networking

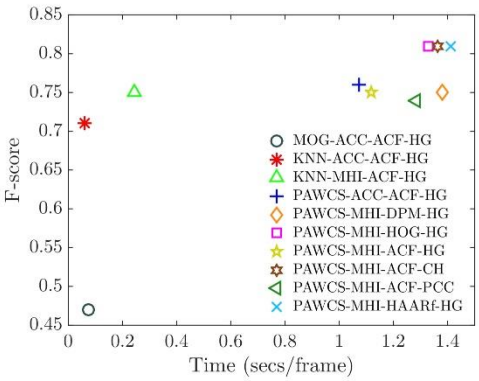
There were 16 participants, including representatives from Treelogic, Vaelsys, and Sigma Technologies

Extension period second semester progress results

Journals

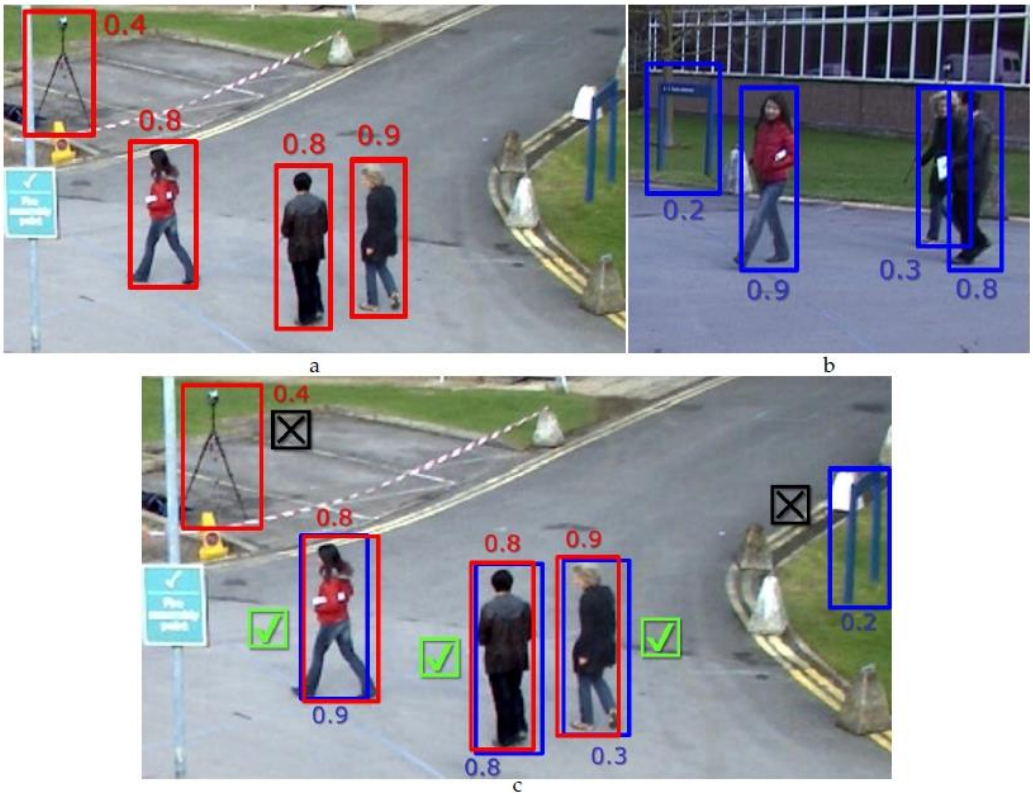
Elena Luna, Juan Carlos San Miguel, Diego Ortego, José M. Martínez, "Abandoned Object Detection in Video-Surveillance: Survey and Comparison", Sensors 18(12):4290, 2018, MDPI, ISSN 1424-8220, (DOI 10.3390/s18124290)

Abstract: During the last years, abandoned object detection emerged as a hot topic in the video-surveillance community. As a consequence, a myriad of systems have been proposed for automatic monitoring of public and private places, while addressing several challenges affecting detection performance. Due to the complexity of these systems, researchers often address independently their stages such as foreground segmentation, stationary object detection and abandonment validation. Despite the improvement achieved for each stage, the advances are rarely applied to the full pipeline and therefore, the impact on the overall system performance is not studied. In this paper, we formalize the framework employed by systems for abandoned object detection and provide an extensive review of state-of-the-art approaches. We also conduct experimental comparisons of existing approaches over a heterogeneous dataset presenting various challenges such as illumination changes, shadows and high densities of moving objects. For such comparison, we build a multi-configuration system allowing to select a range of alternatives for each stage, which is made available online to the research community. The experimental results identify the most effective configurations and highlight design choices favoring robustness to errors. We conclude the paper by discussing open research challenges arising from the experimental comparison.



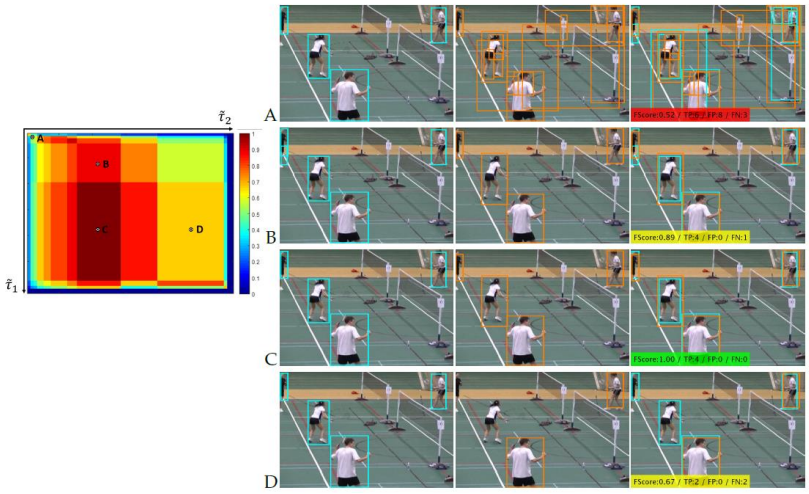
Rafael Martín-Nieto, Álvaro García-Martín, José M. Martínez, Juan C. San Miguel, "Enhancing multi-camera people detection by online automatic parametrization using detection transfer and self-correlation maximization", Sensors, 18(12):4385, 2018, MDPI, ISSN 1424-8220, (DOI 10.3390/s18124385

Abstract: Finding optimal parametrizations for people detectors is a complicated task due to the large number of parameters and the high variability of application scenarios. In this paper, we propose a framework to adapt and improve any detector automatically in multi-camera scenarios where people are observed from various viewpoints. By accurately transferring detector results between camera viewpoints and by self-correlating these transferred results, the best configuration (in this paper, the detection threshold) for each detector-viewpoint pair is identified online without requiring any additional manually-labeled ground truth apart from the offline training of the detection model. Such a configuration consists of establishing the confidence detection threshold present in every people detector, which is a critical parameter affecting detection performance. The experimental results demonstrate that the proposed framework improves the performance of four different state-of-the-art detectors (DPM , ACF, faster R-CNN, and YOLO9000) whose Optimal Fixed Thresholds (OFTs) have been determined and fixed during training time using standard datasets.



Álvaro García-Martín, Juan C. San Miguel, José M. Martínez, "Coarse-to-fine adaptive people detection for video sequences by maximizing mutual information", Sensors, 19(1):4, 2019, MDPI, ISSN 1424-8220, (DOI 10.3390/s19010004)

Abstract: Applying people detectors to unseen data is challenging since patterns distributions, such as viewpoints, motion, poses, backgrounds, occlusions and people sizes, may significantly differ from the ones of the training dataset. In this paper, we propose a coarse-to-fine framework to adapt frame by frame people detectors during runtime classification, without requiring any additional manually labeled ground truth apart from the offline training of the detection model. Such adaptation make use of multiple detectors mutual information, i.e., similarities and dissimilarities of detectors estimated and agreed by pair-wise correlating their outputs. Globally, the proposed adaptation discriminates between relevant instants in a video sequence, i.e. identifies the representative frames for an adaptation of the system. Locally, the proposed adaptation identifies the best configuration (i.e. detection threshold) of each detector under analysis, maximizing the mutual information to obtain the detection threshold of each detector. The proposed coarse-to-fine approach does not require to train the detectors for each new scenario and uses standard people detector outputs, i.e., bounding boxes. The experimental results demonstrate that the proposed approach outperforms state-of-the-art detectors whose optimal threshold configurations are previously determined and fixed from offline training data.



Conferences

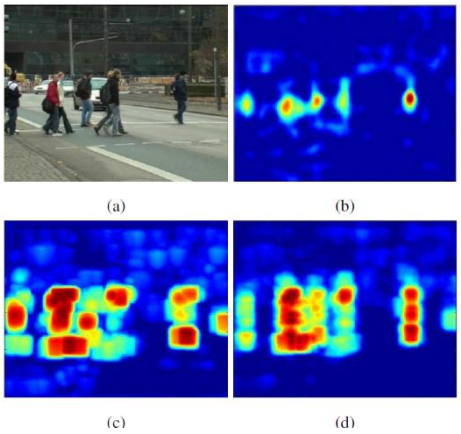
Matej Kristan et al., "**The sixth Visual Object Tracking VOT2018 challenge results**", Proc. of European Conference on Computer Vision Workshops – ECCVW 2018, Munich, Germany, Sept. 2018, (in press).

Abstract: The Visual Object Tracking challenge VOT2018 is the sixth annual tracker benchmarking activity organized by the VOT initiative. Results of over eighty trackers are presented; many are state-of-the-art trackers published at major computer vision conferences or in journals in the recent years. The evaluation included the standard VOT and other popular methodologies for short-term tracking analysis and a \real-time" experiment simulating a situation where a tracker processes images as if provided by a continuously running

sensor. A long-term tracking subchallenge has been introduced to the set of standard VOT sub-challenges. The new subchallenge focuses on long-term tracking properties, namely coping with target disappearance and reappearance. A new dataset has been compiled and a performance evaluation methodology that focuses on long-term tracking capabilities has been adopted. The VOT toolkit has been updated to support both standard short-term and the new long-term tracking subchallenges. Performance of the tested trackers typically by far exceeds standard baselines. The source code for most of the trackers is publicly available from the VOT page. The dataset, the evaluation kit and the results are publicly available at the challenge website.

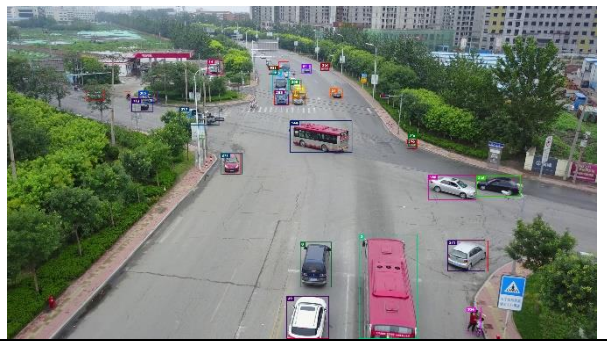
Alejandro Peña, Álvaro García-Martín, Paula Moral, José M. Martínez, Noel O'Connor: "**People-background segmentation based on DeepPyramid Deformable Part Models**", Proc. of 23rd Iberoamerican Congress on Pattern Recognition - CIARP 2018, Madrid, Spain, Nov. 2018.

Abstract: In this work, we propose a deep learning based people-background segmentation approach. A people-background segmentation consists of a traditional two classes segmentation, background and people, but with a totally different purpose, the goal of guaranteeing that no people (or body parts) are miss-classified as background. Our deep learning segmentation approach is based on the use of multiple deformable body parts as Convolutional Neural Networks (CNNs) integrated in a feature pyramid constructed by a second CNN. This scheme, named DeepPyramid Deformable Part Models (DP-DPMs), significantly outperforms DPMs based on traditional Histograms of Oriented Gradients features (HOG).



Elena Luna, Juan Carlos San Miguel, Diego Ortego, José M. Martínez: "**Multi-target Tracking from Unmanned Aerial Vehicles**", Proc. of 23rd Iberoamerican Congress on Pattern Recognition - CIARP 2018, Madrid, Spain, Nov. 2018.

Abstract: Unmanned Aerial Vehicles have experimented a recent commercial advent that has lead to a new and growing research area in computer vision scope, where the goal is to empower drones to automatically detect and track targets. Existing Multi-Target Tracking (MTT) algorithms are not usually optimal for dealing with the challenges of images captured by UAVs, such as the viewpoint, tiny



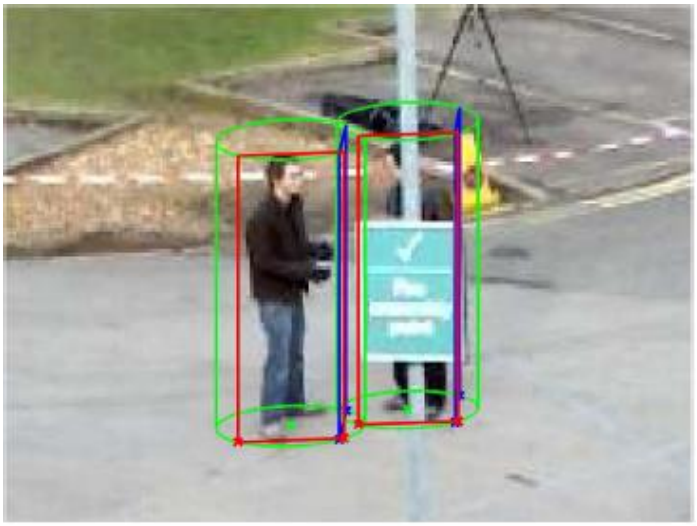
scales, camera movement, etc. We present ongoing research on MTT in images captured from UAVs and provide preliminary results for the Visdrone 2018 benchmark to drive future work. In the proposed framework, where targets are modeled by their visual appearance (via deep features) and their spatial location (via bounding boxes), feature extraction, data association, track management, model update, and spatial prediction are carried out.

PhD thesis

Object detection for video-monitoring using fixed multi-camera systems, Rafael Martín Nieto (advisors: Álvaro García-Martín, José M. Martínez), Tesis Doctoral (PhD Thesis), Univ. Autónoma de Madrid, Sep. 2018.

Abstract: Object detection is one of the most important tasks in computer vision. This is a very complex task due to the difficulty of modelling objects, which contains a high degree of variability, and its performance is also very dependent on the data used for training. There are multiple detection algorithms in the state of the art, but all them present problems with one or multiple factors like: occlusions, illumination changes, perspective changes, etc. This thesis addresses tasks related to object detection: training and evaluation framework, detection approaches and applications, and detection improvements in multi-camera scenarios. In the first part of this thesis, we focus on the training and evaluation framework. We analyze existing datasets in the state of the art that meet the requirements we need to evaluate the different developed systems. These datasets must be multi-camera datasets, in which the cameras have an orientation that generate overlap between the points of view. To complete these existing datasets, two new datasets have been designed, recorded and published: one containing wheelchair users, and another one which contains vehicles in a parking lot. Continuing with the evaluation framework, we present the metrics commonly used for the evaluation of object detectors. First, the "classical" evaluation metrics are formulated, precision and recall, and their combinations. For the evaluation of some of the different developed applications, we adapt these metrics for, on the one hand, considering a third dimension (depth) in the scenarios, and, on the other hand, evaluating the capacity to detect occupied or empty parking spots. To finish this part, we present a technique for the generation of synthetic datasets to be able to train a detection model without having enough training data. We train a wheelchair user model considering synthetic datasets from empty wheelchairs images and standing people images. Three synthetic image datasets have been created in order to train three different models, evaluating which model is optimal, and, finally, analyzing its feasibility by comparing it with a people detector model for wheelchair users trained with real images. In the second part, this thesis presents two different detection approaches with a final application. With the idea of providing an existing object detector model the capacity to detect variants of the desired object, which have

not been considered in their initial design, we present a wheelchair users model and we include it in a generic people detector, providing a more general solution to detect people in environments such as houses adapted for independent and assisted living, hospitals, healthcare centers and senior residences. As an application of the presented work, an example of a room in a nursing home is shown in which the detections are mapped on the ground plane in order to monitor people. To conclude this part, we present an automatic multi-camera system for vehicle detection and their corresponding mapping into the parking spots of a parking lot. The results clearly show that the proposed system works correctly in challenging scenarios including almost total occlusions, illumination changes and different weather conditions. Finally, the third part of the thesis takes as starting point the output of the detection algorithms executed on the images and sequences, adding performance improvements and autoparameterization of the algorithms. We combine information obtained from different cameras in order to enhance object detection algorithms performance. Using multiple cameras and information from the recorded scenario, called contextual information (distances between detected objects and cameras, position of the cameras, etc.), the detection performance is improved taking advantage of the results of the other cameras, transferring information from one camera to another, and then combining it. This technique also allows, using an additional correlation framework, to automatically adapt (determining an optimal threshold for each camera) and improve any detector in multi-camera scenarios, during runtime detection.



Graduate thesis

Detección de ritmo cardíaco mediante análisis de secuencias de vídeo en color **(Color video sequences analysis for heart rate detection)**, Ana Martín Doncel (advisor: José M. Martínez), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: The aim of this Bachelor Thesis is to develop a system which estimates heart rate from a random person analysing variations un skin colour by using non-invasive techniques. This study has been motivated by the importance of

our heart's health since it is in charge of pumping blood to every single part of the body. Image processing, in particular information extraction, is the order of the day in medical applications, due to technological advances nowadays. For this reason, video sequence analysis seems to be a good initiative to detect heart rhythm in a noninvasive way. Blood circulation causes volumetric changes in blood vessels. This makes the light path vary in a periodic way throughout the cardiac cycle. Furthermore, hemoglobin has a fixed and visible spectral range so it would be a good idea to work with colorful images to extract heart rate. This thesis has been divided into two parts. In the first one we have designed algorithms which are based on the method that has been described. They have been compared to an already existing algorithm which has been tested so we can observe their reliability, reaching a percentage rate from 0.339 y 3.543. Once we have completed this experiments and results have been obtained, we proceed to the second part. This part consists in an evaluation of different parameters through a sweep of the signal. In this study we have achieved good results. recent years, video-surveillance systems, and more specifically camera positioning systems, are now being deployed widely thanks to their ability. In this context, optimal camera configuration will reduce the total number of cameras used achieving the same or greater level of utility, as well as better results and lower cost associated with future modifications. The objective of this project is the development of a Matlab toolbox which allows the calculation of the camera's position according to the constraints associated to the problem's formulation: continuous or discrete domain. The toolbox makes use of an interface that allows the user to select the floor map that is to be analyzed, the coverage's restrictions and the type of optimization to be used. Additionally, a graphical interface functionality has been included which allows the user to insert objects on the floor map by directly clicking on the interface, as well as fixing their height relative to the ceiling's. The final result will allow the user to visually examine the location of the cameras and the field of view.

Detección de ritmo cardíaco mediante análisis de secuencias de vídeo en modalidad sin color **(Non-Color modality video sequences analysis for heart rate detection)**, Claudia Fernández Refoyo (advisor: José M. Martínez), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: The objective of this Final Degree Thesis is to improve an algorithm that calculate the heart rate. This heart rate is detected by using depth images based on movement changes. The algorithm has been adapted to be applied on depth images. It has been decided to apply this algorithm on depth images because they can't be influence by the lighting conditions and they could ensure the individual privacy. It has been implemented in the first instance an algorithm that could detect the heart rate manually. Once this algorithm has been valued on depth images by obtaining satisfactory results, it has been decided to add

new implementations. These implementations have been incorporated on the first stage of the algorithm: the detection of the area of interest. This region of interest is detected automatically and considering the distance between the subject and the camera. Lastly, it has been realized a comparative among the two methods. It is concluded that even though the results obtained manually are more precise, it is better to use the algorithm automatically since it doesn't require the interaction of the person to identify the area of interest.

Detección de vehículos en entornos multi-cámara utilizando información contextual **(Vehicles detection in multicamera setups using contextual information)**, Juan Enrique de Santiago Rojo (advisor: Rafael Martín-Nieto), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: Due to the high demand present nowadays in the video-surveillance area, the number of researches related to this field has been increased. In particular, with the purpose of making parking access easier, a multi-camera system is proposed for the vehicle detection and its corresponding mapping into the parking spots of a parking lot. Therefore, we will obtain the knowledge of which parking spots are free or occupied. Thanks to this system, the use of the common deployed sensors can be substituted by computer vision, which also reduces system costs. In order to make more efficient systems, many algorithms have been developed for object detection. In this work, two object detection algorithms have been used, Deformable Parts Model (DPM) and a Faster Regions with Convolutional Neural Network (Faster RCNN). Both tested in previous works. The main goal in this work is to improve a vehicle detector system by fusing the information captured from the environment, from a set of cameras. The detections performed in the parking are improved, so some detections which were not possible to detect from one of the cameras now obtained thanks to another camera´s information. This helps to get detections from those objects that are far away or occluded as different points of view are taken into consideration.

Detección jerárquica de vehículos **(Vehicles hierarchical detection)**, Raúl Palacios Ramos (advisor: Álvaro García-Martín), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: The establishment of visual capture system in both public and private spaces, makes analysis and control at the processing level very complicated. Owing to the great importance of these aspects, it would be necessary to study and expand systems that update what has been settled down so far. The current vehicle detection algorithms, in general, earn good results in controlled spaces. But this property is undeniable that it is not usually fulfilled. Due to that, a system

that allows the location of vehicles in variability of environments has been proposed. Specially, we are going to focus on the case of occlusions between models. The main objective will be to develop a detector that enables locating vehicles, by obtaining information from different mechanical configurations and increasing the search area around them. The vehicle could be defined at a general level or using some of its parts. By last the proposed algorithm on video sequences will be reviewed, showing there is an increment in the detection of vehicles thanks to the improvements made.

Localización automática de cámaras en sistemas de análisis de vídeo con múltiples vistas **(Automatic location of cameras in multiview video analysis systems)**, Enrique Sepúlveda Jorcano (advisor: Juan Carlos SanMiguel), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: The objective of this Final Degree Thesis is to implement an algorithm to automatically calibrate a camera and estimate the parameters that define that camera without human interaction needed. Until now, it was necessary to manually introduce a previously known pattern in the scene. We will use prior knowledge of human relative heights distribution in pedestrians from the scene to efficient estimate those parameters, and make the camera calibration automatic. This algorithm will be implemented in C++ and we will use OpenCV library for image processing. We will put to test videos from public datasets and will analyze the results in each step of the process.

Generación automática de conjuntos de evaluación de camuflage **(Automatic generation of camouflague evaluation datasets)**, Sergio Álvarez Balanya (advisor: Marocs Escudero Viñolo), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

Abstract: Background subtraction has become a key step in several computer vision algorithms. There are plenty of studies proposing different and varied approaches. However, the problem of background subtraction is not yet fully addressed. One reason might be the fact that each method has been developed for different tasks, e.g. video surveillance or optical motion capture. The recent appearance of comprehensive datasets provides a common framework for evaluating background subtraction algorithms. These datasets present a balanced repertoire of sequences in which common challenges are present. This leads to extensive overall scores in which robustness against different challenges is considered, but not particularized to these challenges. A particularly barely studied challenge, and the focus of our work, is camouflage: the resemblance between background and foreground samples. The research community agrees that there isn't yet a commonly accepted approach to handle camouflage. In this

work, we propose a novel solution for modeling camouflage based on the Jung's theorem. Based on this solution, we generate camouflage likelihoods for every foreground pixel in a sequence using available ground-truth information to discriminate the background from the foreground. The evaluation of the proposed solution is performed in discrepancy terms by thresholding the camouflage likelihoods to obtain a binary mask on which we apply classical classification metrics. Thereby, we are able to further analyze the effect of the features selected by different background subtraction algorithms in handling camouflage. Furthermore, the proposed solution also permits the ranking of a set of sequences in terms of camouflage. The experiments carried out on the popular CDNET2014 dataset suggest that the use of certain alternative features to color—e,g, motion—is beneficial to robustly handle camouflage.

Segmentación de objetos en vídeos no controlados **(Object segmentation in uncontrolled videos)**, Ariana Vicario Arroyo (advisor: Diego Ortego Hernández), Trabajo Fin de Grado (Graduate Thesis), Grado en Ingeniería de Tecnologías y Servicios de Telecomunicación, Univ. Autónoma de Madrid, Jul. 2018.

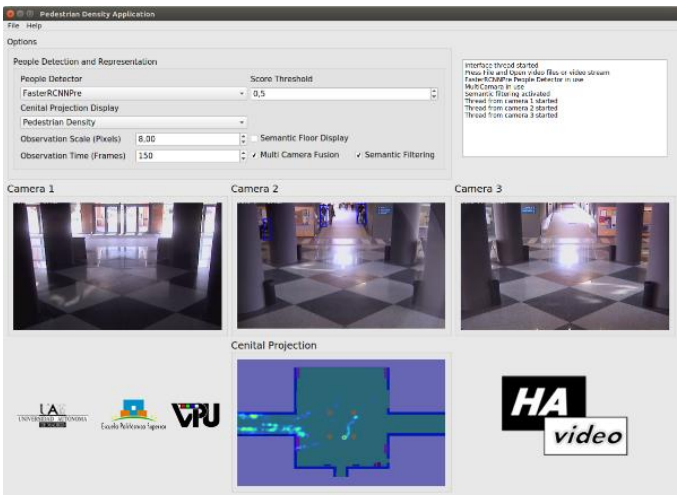
Abstract: In this work, we propose an algorithm which is able to automatically segment objects from sequences recorded by camera motion. In order to develop it, techniques from the state of art have been used to identify relevant space-temporal patterns and to classify them in two categories: foreground and background. The development and implementation has been divided into three steps: Firstly, an initial segmentation in which optical ow and saliency techniques have been applied to extract motion information and to estimate a localization model. Secondly, a medium segmentation has been done. With a localization model, an appearance model is estimated by using mixture of gaussian models and a CRF. Finally, a final segmentation, whose quality of the actual model is estimated and it is compared with the best model obtained in previously loops, has been done. Then, the best model is updated when it has a higher quality. Otherwise, if the quality is worst, the old model is used to obtain another segmentation. In order to validate the algorithm, a set of experiments have been performed to measure the contour accuracy, the region similarity and the temporal stability of the segmented object in each video. As a conclusion, the results of the experiments have been compared with other segmentation algorithms from the research community.

Applications

Pedestrian Density estimation in multi-camera scenarios

This application provides new functionalities to estimate and demonstrate pedestrian density estimation in multi-camera scenarios. Unique scene-wise pedestrian detections are generated projecting detections onto a common ground plane by means of homography transformations.

Pedestrian density estimation is driven by a novel density estimation method which operates on the fused detections on an image representation of the ground plane.

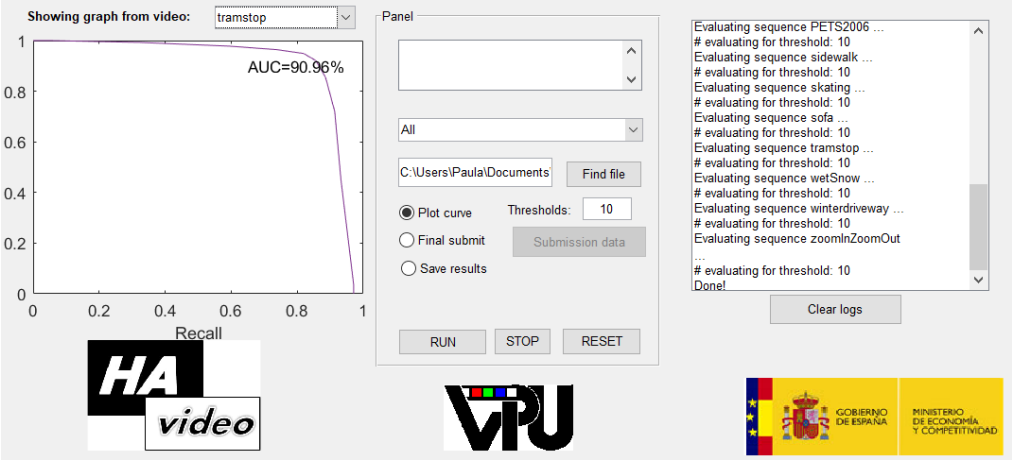


Application for potential distractors detection in video object tracking

This application aims to demonstrate the capabilities of the methods created along the HA-Video project to detect, given a target object, a scored-set of potential distractors in the same video sequence. To this aim, some of the methods have been adapted to this scenario, and a new application with associated interface is developed.

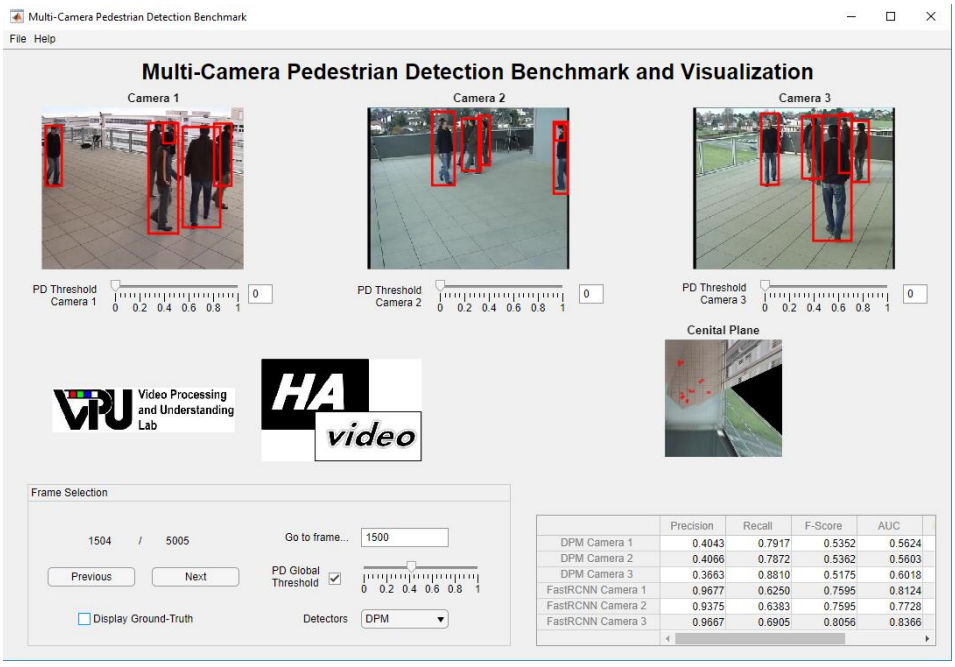
Application for people detection evaluation

This new approach for evaluation of people detection provides a standalone application which allows users to evaluate their algorithms from their own computer, and compare the output with the results of the detectors published in the state of the art.



Multi-Camera Pedestrian Detection Benchmark Application

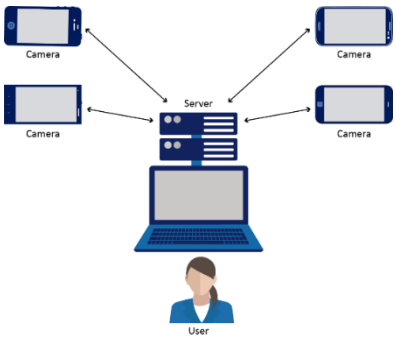
The application allows the users to quantitatively evaluate both state-of-the-art and new pedestrian detectors, providing per-frame and synchronized performance indicators for all the evaluated algorithms. Evaluation routines are included within the



application. Specifically, the following multi-camera pedestrian detection performance indicators have been included: Precision, Recall, F-Score, Area Under the Curve, N-MODA, N-MODP.

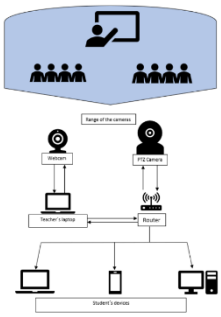
Multi-camera video surveillance system based on smartphones

The proposed system uses Android smartphones as cameras, which are controlled remotely by the user with a client application installed in a computer. In addition, a server is implemented, in the same computer, in order to control and manage every communication between users and cameras.



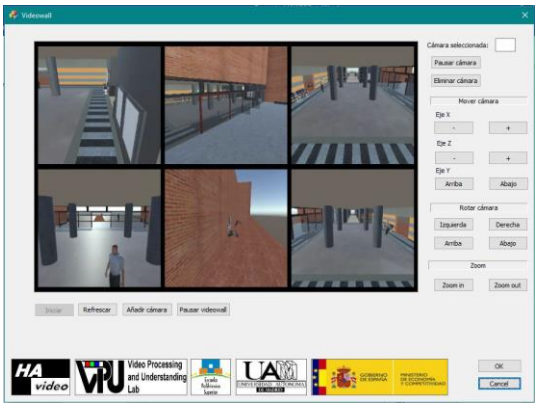
Live streaming of classes and seminars using automatic video tracking

The main motivation of this work is to automate the camera tracking of a lecturer in a classroom to generate a portable system that could stream classes in real-time using on-demand video streaming and automatic tracking. It makes uses of the Long-term Tracking with Target Re-identification applications developed earlier in the project.



Visualizer and controller for Multi-camera System Simulator (MSS simulator)

The main motivation of this work is to automate the camera tracking of a lecturer in a classroom to generate a portable system that could stream classes in real-time using on-demand video streaming and automatic tracking.



The portable system is composed of a cameras system, a router and a laptop. It uses two cameras; a PTZ IP camera and a webcam, both with a broad range of vision, which will be connected to the router and the laptop, respectively.

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