

eVent
Video

eVi Newsletters
#01 - March 2012

TEC2011-25995 EventVideo (2012-2014)

Strategies for Object Segmentation, Detection and Tracking in Complex
Environments for Event Detection in Video Surveillance and Monitoring

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

Introducing EventVideo

The EventVideo project, officially named “Strategies for Object Segmentation, Detection and Tracking in Complex Environments for Event Detection in Video Surveillance and Monitoring” and with reference TEC2011-25995, is a three years basic research project funded by the Spanish Government under the supervision of the *Ministerio de Economía y Competividad*.

The objective of the EventVideo project is to tackle the problem of event detection in video sequences in complex situations. In this direction, in order to achieve improvements in the detection of state-of-the-art events in the video surveillance and monitoring domain, the workplan proposes research in the design and development of new strategies in the different phases or stages of a video analysis chain.

The main proposed innovation starts from the hypothesis that the most traditional or popular approaches to object segmentation, detection and tracking, while successful for many simple and/or controlled applications, do not achieve satisfactory results in video sequences captured in quite frequent and specially complex situations (e.g., outdoor environments, adverse weather conditions, presence of heavy shadows and reflection, high objects density, low resolution objects, partial occlusions, moving camera, cluttered backgrounds). These situations may require a more thorough consideration of light effects, of robust approaches to global motion estimation, of combined models to tackle object detection and tracking, and of the use of complementary sensors.

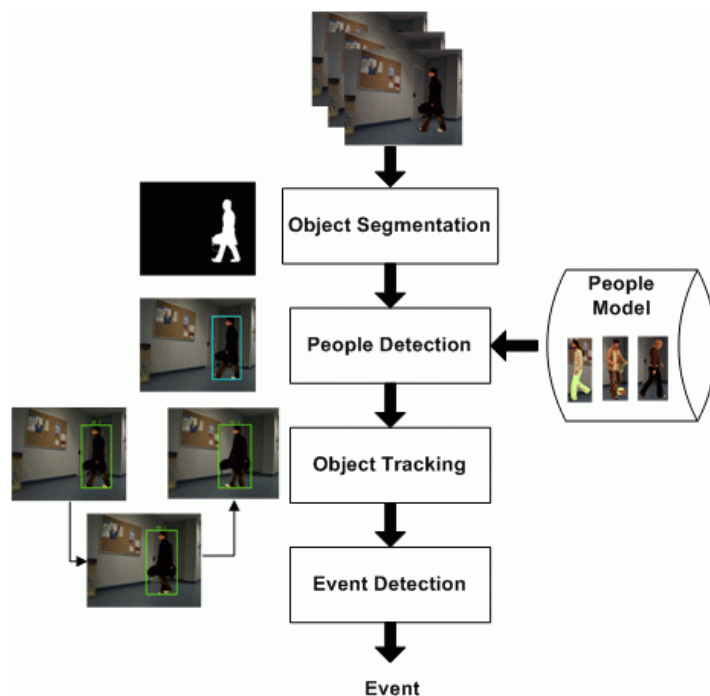
First trimester progress report

During the first semester, the project has been officially launched. Besides the release of the website <http://www-vpu.eps.uam.es/eventvideo/> end of January 2012, the research team has detailed the workplan for the first year of the project.

WP1 “Framework establishment and maintenance” will release during June 2012 the first versions of deliverables D1.1 “System Infrastructure” and D1.2 “DiVA documentation” in order to clearly state the current infrastructure for sequence computing and acquisition and evaluate future actions (e.g., equipment acquisition).

The “pure” algorithmic workpackages, namely WP2 “Video object segmentation”, WP3 “People modeling and detection” and WP4 “Video object tracking”, will work during this year in fine tuning the algorithms designed and developed by the Video Processing and

Understanding Lab (VPULab) and testing them over the video sequences content sets (test sequences and associated ground-truths) defined in WP5, in order to plan future research lines at the end of this year. The main objective during this year will be to validate the performance of the available and under development algorithms in different scenarios and identify future approaches in order to move from the Low Density-Controlled Environment scenario, where algorithms



perform well, to the other ones: Low Density-Uncontrolled Environment, High Density-Controlled Environment and High Density-Uncontrolled Environment.

WP5 “Integration, orchestration, testing and evaluation, and dissemination” has already released the EventVideo web site and will release during June 2012 the first version of deliverables D5.3 “EventVideo test sequences, ground-truth and evaluation methodology” and D5.4 “EventVideo results report”. Activities in evaluation will focus in two lines: classical evaluation based on ground-truth and auto-evaluation without the need of ground-truth. Regarding the content sets, we want to add to the classical ones (visible range), one based on laser or infrared patterns in order to evaluate the potential of the algorithms over sequences captured with alternative sensors. Additionally, we want to start thinking, with the help of the EventVideo Observers, about potential applications (e.g., abandoned/stolen object detection, people counting, event detection).

Although in the project proposal we proposed to participate in the TRECVID Event Detection challenge, as this year the challenge has changed in order to deal with Interactive Event Search, we won't participate in 2012. We are looking for other public challenges or competitions in order to evaluate our technology with other SoA techniques.

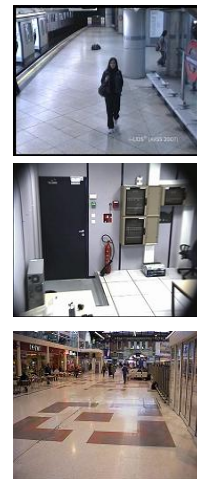
First trimester results

The EventVideo project started officially January 2012, nevertheless the Video Processing and Understanding Lab (VPULab) has been working during part of 2011 in the research lines proposed in the project. Therefore the project has already produced some public results.

Datasets

Abandoned and Stolen Object Discrimination dataset - ASODds

ASODds is a corpus of video sequences that provide a representative test-set for comparing systems devoted to discriminate previously detected stationary regions between abandoned and stolen objects in video surveillance. The dataset is composed of several annotated surveillance sequences of different levels of complexity. Sequences have been extracted from public datasets related with the people detection/object classification task (e.g., PETS, VISOR, WCAM). The test sequences are grouped into three categories (low, medium and high) according to a subjective estimation of the background complexity.



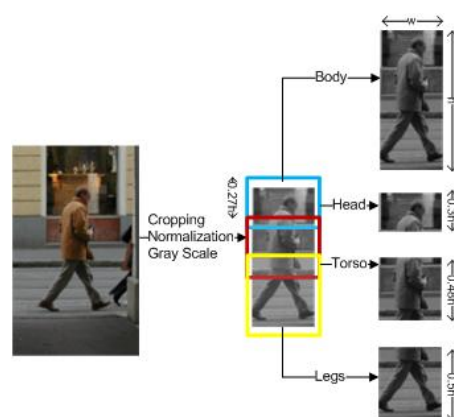
Event Detection dataset - EDds

EDds is a corpus of video sequences created to provide a representative test-set for comparing event recognition algorithms. Moreover, manual annotations of the events of interest are also provided. The dataset is focused on two types of human-related events: interactions and activities. The former refers to actions performed among several humans or environmental objects (e.g., get objects). The latter considers human actions not involving physical interaction (e.g. walking). In particular, two activities (Hand Up and walking) and three human-object interactions (Leave, Get and Use object) have been annotated.



Person Detection dataset – PDds

PDds is a dataset composed of several annotated surveillance sequences of different levels of complexity. Sequences have been extracted from public datasets related with the people detection/object classification task (e.g., PETS, VISOR, TRECVID). All the test sequences are categorized into different complexity categories depending on two aspects: People classification complexity and background complexity.

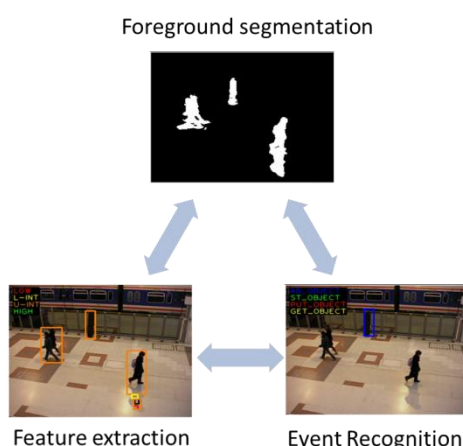


These datasets are freely available for research purposes and can be downloaded following the Content Sets link at the EventVideo web site.

Publications

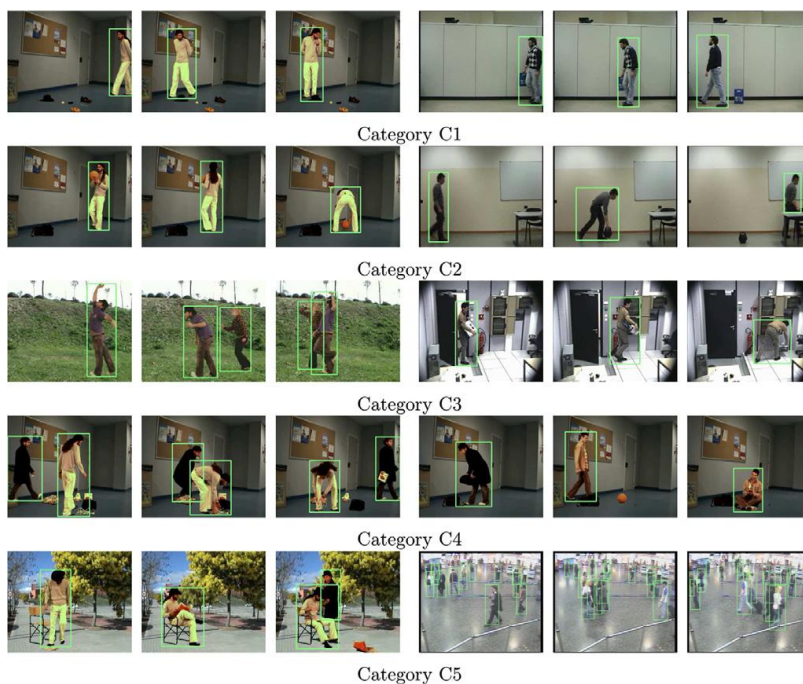
Juan C. SanMiguel, José M. Martínez, “Use of feedback strategies in the detection of events for video surveillance”, IET Computer Vision, 5(5):309–319, Sep. 2011. (DOI [10.1049/iet-cvi.2010.0047](https://doi.org/10.1049/iet-cvi.2010.0047))

Abstract: The authors present a feedback-based approach for event detection in video surveillance that improves the detection accuracy and dynamically adapts the computational effort depending on the complexity of the analysed data. A core feedback structure is proposed based on defining different levels of detail for the analysis performed and estimating the complexity of the data being analysed. Then, three feedback-based analysis strategies are defined (based on this core structure) and introduced in the processing stages of a typical video surveillance system. A rule-based system is designed to manage the interaction between these feedback-strategies. Experimental results show that the proposed approach slightly increases the detection reliability, whereas highly reduces the computational effort as compared to the initially developed surveillance system (without feedback strategies) across a variety of multiple video surveillance scenarios operating at real time.



Álvaro García-Martín, José M. Martínez, Jesús Bescós, "A corpus for benchmarking of people detection algorithms", *Pattern Recognition Letters*, 33(2):152–156, Jan. 2012. (DOI [10.1016/j.patrec.2011.09.038](https://doi.org/10.1016/j.patrec.2011.09.038))

Abstract: This paper describes a corpus, dataset and associated ground-truth, for the evaluation of people detection algorithms in surveillance video scenarios, along with the design procedure followed to generate it. Sequences from scenes with different levels of complexity have been manually annotated. Each person present at a scene has been labeled frame by frame, in order to automatically obtain a people detection ground-truth for each sequence. Sequences have been classified into different complexity categories depending on critical factors that typically affect the behavior of detection algorithms.

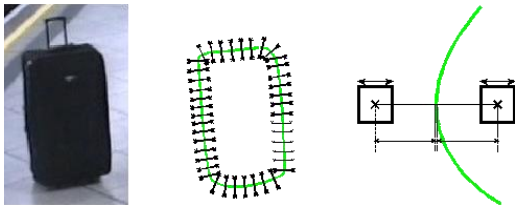


The resulting corpus, which exceeds other public pedestrian datasets in the amount of video sequences and its complexity variability, is freely available for benchmarking and research purposes under a license agreement.

Juan Carlos San Miguel, Luis Caro, José M. Martínez, "Pixel-based colour contrast for abandoned and stolen object discrimination in video surveillance", *Electronic Letters*, 48(2):86–87, Jan. 2012. (DOI [10.1049/el.2011.3160](https://doi.org/10.1049/el.2011.3160))

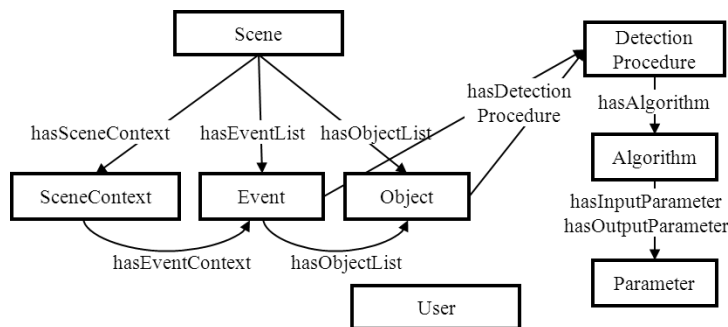
Abstract: A novel approach is proposed for discriminating between abandoned or stolen previously detected stationary foreground regions in video surveillance. It is based on measuring the colour contrast of the contour of the stationary object under analysis at pixel level. Two contrasts are computed by analysing such a contour in the current and background frames. Then, both are combined for performing the discrimination.

The experimental results over a heterogeneous dataset containing real scenarios demonstrate that this approach outperforms the related literature and greatly reduces the computational cost of the discrimination task, allowing real-time operation.



Juan Carlos San Miguel, José M. Martínez, “A semantic-guided and self-configurable framework for video analysis”, *Machine Vision and Applications* (Online, Dec. 2011). (DOI [10.1007/s00138-011-0397-x](https://doi.org/10.1007/s00138-011-0397-x))

Abstract: This paper presents a distributed and scalable framework for video analysis that automatically estimates the optimal workflow required for the analysis of different application domains. It integrates several technologies related with data acquisition, visual analysis tools, communication protocols, and data storage. Moreover, hierarchical semantic representations are included in the framework to describe the application domain, the analysis capabilities, and the user preferences.

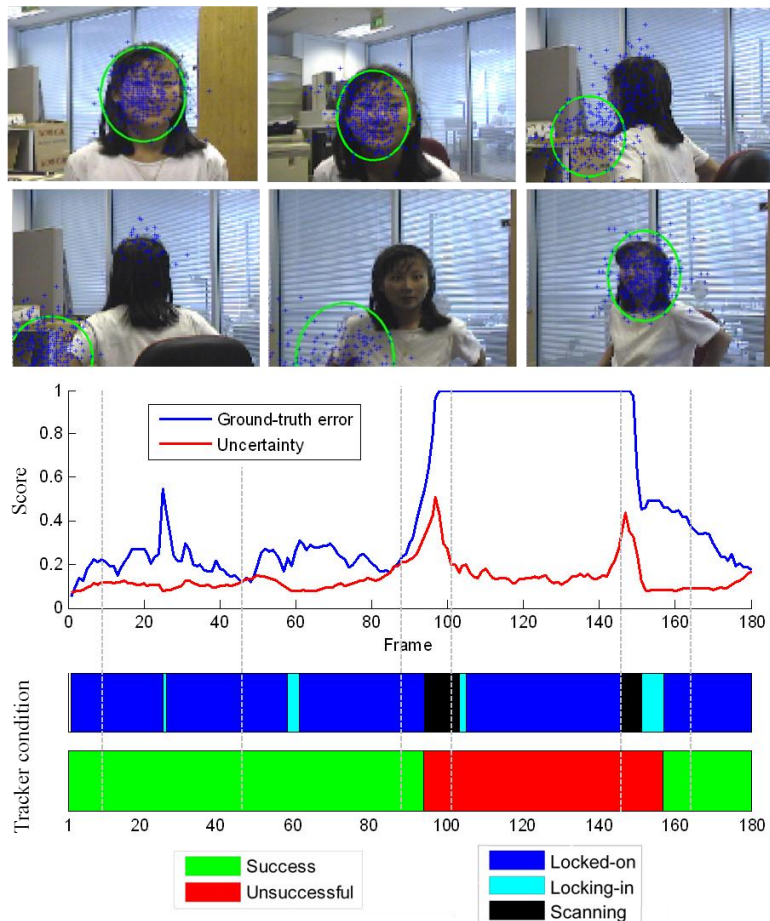


The automatic determination of the analysis workflow is performed by selecting the most appropriate tools for each domain among the available ones in the framework by means of exploiting the relations between the semantic descriptions. The experimental results in the video surveillance domain demonstrate that the proposed approach successfully composes optimal workflows for video analysis applications.

Juan Carlos San Miguel, Andrea Cavallaro, José M. Martínez, “Adaptive on-line performance evaluation of video trackers”, *IEEE Transactions on Image Processing* (Online, Jan. 2012). (DOI [10.1109/TIP.2011.2182520](https://doi.org/10.1109/TIP.2011.2182520))

Abstract: We propose an adaptive framework to estimate the quality of video tracking algorithms without ground-truth data. The framework is divided into two main stages, namely the estimation of the tracker condition to identify temporal segments during which a target is lost and the measurement of the quality of the estimated track when the tracker is successful. A key novelty of

the proposed framework is the capability of evaluating video trackers with multiple failures and recoveries over long sequences. Successful tracking is identified by analyzing the uncertainty of the tracker, whereas track recovery from errors is determined based on the time-reversibility constraint. The proposed approach is demonstrated on a particle filter tracker over a heterogeneous dataset.



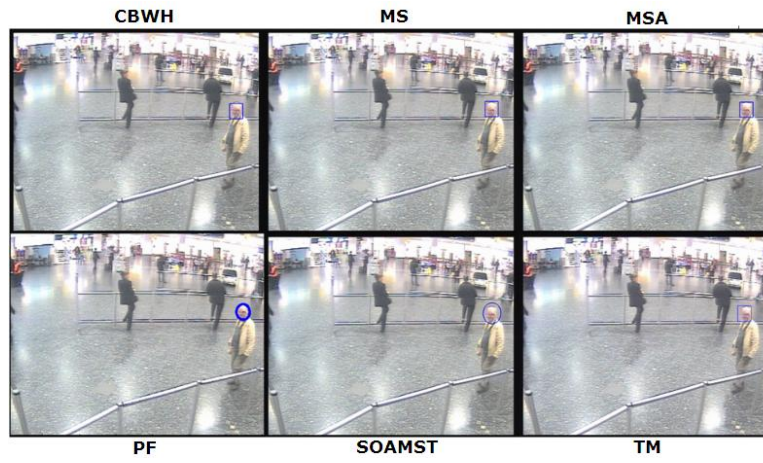
Experimental results show the effectiveness and robustness of the proposed framework that improves state-of-art approaches in the presence of tracking challenges such as occlusions, illumination changes and clutter, and on sequences containing multiple tracking errors and recoveries.

Master thesis

Evaluación comparativa de algoritmos de seguimiento de objetos (tracking), Mónica Cruz Lozano (tutor: Juan C. San Miguel, ponente: José M. Martínez), Proyecto Fin de Carrera, Univ. Autónoma de Madrid, Escuela Politécnica Superior, Feb. 2012.

Abstract: The main objective of this master thesis is the evaluation of video tracking algorithms. A protocol is designed for such evaluation task including

the following aspects: accuracy, robustness to initialization errors, computational cost and optimum parameterization. In order to support the evaluation, a dataset is defined considering the most common problems in video tracking. Finally, the proposed protocol is tested on representative approaches of the state-of-the-art.



The EventVideo (TEC2011-25995) project is supported by

