ETISEO Performance Evaluation for Video Surveillance Systems

F. Bremond and A.T Nghiem

ETISEO [1] was a two year project (ended in December 2006) on performance evaluation for video surveillance systems. ETISEO aims at helping algorithm developers to identify algorithm weaknesses and to underline the dependencies between algorithms and their conditions of use.

1. ETISEO project

More precisely, ETISEO aims at evaluating video processing algorithms given a video processing task (i.e. object detection, classification, tracking and event recognition), a type of scene (e.g. road) and a global difficulty level (e.g. contrasted shadows). The ultimate goal is to study the dependencies between a video processing task and video characteristics (e.g. shadows), which are called in the following, video processing problems. The methodology of ETISEO is as follows:

Firstly, ETISEO addresses separately each video processing problem that have been accurately defined and classified. For instance, handling shadows can be studied within at least three different problems: (1) shadows at different intensity levels (i.e. weakly or strongly contrasted shadows) with uniform non color background, (2) shadows at the same intensity level with different types of background images in terms of color and texture and (3) shadows with different illumination sources in terms of source position and wavelengths.

Secondly, ETISEO collects video sequences illustrating only a given problem. The video sequences were intended to illustrate the video processing problem at different difficulty levels. For instance, for the problem of shadows and intensity levels, we have selected video sequences containing shadows at different intensity levels (more or less contrasted). On these selected sequences, the appropriate part of the ground truth is filtered and extracted to isolate video processing problems. For instance, for the object detection task, we can evaluate the algorithm performance relatively to the problem of handling occluded objects by considering only the ground truth related to the occluded objects.

Thirdly, ETISEO provides three types of associated data for each video sequence. The first one is the ground truth (e.g. object bounding box, object class, event etc.) given by human operators using VIPER tool [2] at each level of the four video processing tasks. The second one is the general annotation on the video sequences concerning video processing difficulties (e.g. weak shadows) or concerning recording conditions (e.g. weather conditions such as sunny day). The third information is the camera calibration and contextual information about the empty scene describing the topology of the scene (e.g. zone of interest).

Fourthly, ETISEO has defined various metrics to evaluate the performance of a video surveillance system for every video processing task (object detection, tracking, object classification and event recognition). The ETISEO metrics are described in a report document [1]. An analysis of the evaluation metric advantages and drawbacks can be found in [4].

Finally, ETISEO provides a flexible and automatic evaluation tool to accurately analyse and visualise how a given algorithm addresses a given problem.

2. ETISEO limitations and future work

ETISEO had to face several shortcomings. In this section we review these shortcomings and propose few research directions to go beyond.

Firstly, there were still inconsistencies among participants, particularly in defining the objects and events of interest. For instance, several participants processed the stationary objects differently.

Some participants considered the objects not moving for a certain period of time as part of the background and eliminated them from the algorithm results while others detected these objects up to the end as it was requested. Therefore it was difficult to compare the algorithm results of these participants. The solution was to create a filter that removes these objects from both the ground truth and the algorithm results. After applying the filter, the algorithms were ranked differently. This filter has enable us to distinguish two different problems: handling stationary objects which were previously mobile and mobile object detection. However more needs to be done (more filters) to get really convenient and easy to use evaluation platform.

Secondly ETISEO did not set up a limit on the processing time to satisfy the real-time requirement. Hence some participants have applied sophisticated algorithms with a learning stage and have obtained good evaluation results. Moreover, ETISEO did not require the participants to keep the same algorithm parameters for all the video sequences or at least for each type of scene. Consequently, they have tuned their algorithms to achieve better results on each video sequence. Then, the evaluation results do not reflect the algorithm performance in real conditions which change arbitrarily but rather the partner involvement in ETISEO. To mitigate the performance results, the participants were asked to fill up questionnaires indicating the algorithm requirements (e.g. how many parameters have been tuned). ETISEO has set up a workshop to demonstrate the real time capability and the dynamic configuration of the systems. However too few partners agreed for real time demonstration. A web-based evaluation platform could be useful to let developers evaluating their algorithm online whenever they can with a seamless integration protocol.

Thirdly, the evaluation results communicated through numbers and curves mostly help to compare the algorithm performance between themselves. In the user point of view, it is difficult to answer the question of how significant are these values. For instance, using metric ``number of objects'', is the F-Score value equal to 0.8 good enough? is the difference of 0.1 between two algorithms significant? There is no absolute answer to these questions because the answer depends on the specific application. We should perform also an end user evaluation on a selection of applications to establish the significance of the evaluation results (i.e. numbers).

Finally, although ETISEO has tried to estimate the difficulty levels of the video processing problems in each sequence, this estimation is still very rough. For instance, ETISEO uses the terms ``normal'' or ``dark'' to describe the intensity levels of video sequences. Therefore, the selection of video sequences in ETISEO according to their difficulty levels is not sufficient because the comparison among video sequences is subjective and imprecise. Moreover, the prediction of algorithm performance on new scenes based on these evaluation results is difficult because we have to compare these new scenes with the ETISEO video sequences. To solve this problem, we are currently working on defining objective and quantitative metrics to measure automatically the difficulty levels of video processing problems [3].

Once these issues are addressed, we will be able to predict the performance of a given algorithm on new videos. To predict algorithm performance is a crucial point for video surveillance system certification which has not been addressed up to now in the scientific community.

Reference

[1] ETISEO: Video understanding Evaluation, http://www-sop.inria.fr/orion/ETISEO/
[2] ViPER-GT, the ground truth authoring tool, http://vipertoolkit.sourceforge.net/docs/gt/
[3] A.T. Nghiem, F. Bremond, M. Thonnat, R. Ma, “A New Evaluation Approach for Video Processing Algorithms”, In Proceedings of IEEE International Workshop, on Motion and Video Computing, Austin, Texas 23-24 February
[4] A.T. Nghiem, F. Bremond, M. Thonnat, V.Valentin, ETISEO, performance evaluation for video surveillance systems, AVSS 2007, IEEE International Conference on Advanced Video and Signal based Surveillance, September 5-7, 2007, London - Great Britain