AVITRACK

Contract n° AST3-CT-2003-502818

Deliverable D6.1 B
Prototype scenes tracking evaluation

Version 1 - approved

Project is co-funded by Industry and the European Commission within FP6 Aeronautics

Table of contents

1. INTRODUCTION

This document presents the intermediate performance evaluation results of the AVITRACK tracking prototype. Subject of evaluation is Object Tracking, 3D Scene Reconstruction and Object Categorisation.

The framework ViPER has been chosen for the annotation and performance evaluation of the object tracking module.

For the 3D scene reconstruction evaluation individual people and a service vehicle were considered. The person class was evaluated by comparing trajectories with well known grid lines of the echo-40 apron. GPS data was generated for the service vehicle driving on the apron in order to evaluate its 3D localisation.

Object categorisation was evaluated by checking frame by frame whether objects present in the scene were properly recognised and classified into the appropriate class.

1.1. VIDEO DATASETS

The prototype scene tracking evaluation assesses the performance of the object tracking, 3D localisation and categorisation components on representative test data. The evaluation of the components strongly depends on the choice of the video sequences.

2. VIDEO PERFORMANCE EVALUATION RESOURCE (VIPER)

ViPER allows the annotation of events through the time and permits to track objects in each frame. The framework contains an evaluation tool ViPER-PE to test tracker's output comparing it to the ground truth done with ViPER-GT.

2.1. GROUND TRUTH TOOL (VIPER-GT)

ViPER-GT is a tool for annotating videos with metadata. Data elements are combined together into objects called descriptors allowing the human operator to define object types. Attributes such as text string (name), bounding-box (position) can be associated to each object.

2.2. PERFORMANCE EVALUATION METHODS (VIPER-PE)

The goal of the ViPER-PE tool is to compare the result data of the tracking system with the ground truth, and to generate data describing the success or failure of the performance analysis. The tool includes three different types of analysis: Object-matching, frame-wise comparison and track comparison.

1. Object-matching analysis

Object analysis attempts to match candidate objects to reference objects counting objects as matches when the 'metric distance', defined by the evaluator, is less than a given threshold.

2. Frame-wise analysis

Frame-wise analysis determines object matches between reference and candidate objects by examining correspondences frame-by-frame. The method calculates the precision and number of matches for each frame.

3. Tracking analysis

Tracking analysis assumes that it exist a unique correspondence between reference and candidate objects.

ViPER-PE generates both a human readable output data and a machine readable output data. The human readable output file is divided into two sections, the first illustrates the parameters used in the evaluation and the second the results.

The object analysis results describe at which evaluation level an object descriptor is false or missed.

The pixel-wise analysis output lists a set of properties for each frame. These are the number of true positive, false negative and false positive pixels in the frame as well as a pixel and object count accuracy, a fragmentation measure, and object precision, average box and localized box precision and recall.

The tracking metric lists results for each reference object. These are temporal precision and recall, positional accuracy, size accuracy and angle difference.

3. OBJECT TRACKING EVALUATION

To evaluate the performance of the local feature tracking method five apron datasets were chosen. The datasets were taken under a wide range of disturbing conditions such as illumination changes, occlusions and shadows.

The ViPER's performance evaluation tool has been used to compare the result data of the local feature tracking method with the ground truth in order to generate data describing the success or failure of the performance analysis. At first, the evaluation tool attempts to match tracked objects (TO) to ground truth objects (GTO) counting objects as matches when the following metric distance is less than a given threshold.

Once the tracked objects and ground truth objects have been matched true positives, false negatives and false positives objects are counted and summed up over the chosen frames. The following metrics defines by Black were used to characterize the tracking performance:

- **Tracker detection rate:** *TRDR*
- **False alarm rate:** *FAR*
- **Track detection rate:** *TDR*
- **Track fragmentation:** *TF*

Next, some representative results of the local feature tracking method are presented.

3.1. EXPERIMENTAL RESULTS

- Weak shadows are detected and tracked as part of the mobile objects
- Strong shadows which are correctly detected and removed from the individuals
- Such objects produce ghosts which remain behind the previous object position.

- Objects in the scene are partially detected due to the achromaticity of the scene. Therefore, fragmentation is presented in objects with the same colour as background.
- The light produced by the siren of the tanker is detected as a mobile object
- Occlusions produce object label changes

3.2. ANALYSIS OF RESULTS

At first, the track detection rate *TDR* and the track fragmentation *TF* were computed separately for each ground truth object.

All ground truth objects of the datasets were matched to tracked objects. The evaluation shows a high track detection rates (close to 95%). Most of the objects present a track detection rate between 83% and 100%. But sometimes several dynamic occlusions causing object label changes was observed.

4. 3D SCENE RECONSTRUCTION EVALUATION

For the evaluation of the 3D localisation module individual people and a service vehicle have been considered.

3D localisation output data has been generated for each of the test cameras installed at the airport's apron. The co-ordinate Z is equal to 0 because the objects are constrained to lie on the known ground plane.

For the evaluation of the vehicle trajectory we consider a single trajectory estimate generated by the data fusion module.

The results demonstrate that the estimated vehicle location is reasonably accurate close to the camera sensors. In the far field the estimate diverges from the measured GPS signal due to the perspective effect and the uniform quantisation of the sensor pixels.

Also a study of 2D graphs illustrating the people trajectories along the apron's grid generated by the data fusion module was realised.

4.1. EXPERIMENTAL RESULTS

- Occlusion leads to loss of 3D data information causing errors on 3D trajectory reconstruction.
- The accuracy of the localisation module depends on the distance between the camera and the object due to the perspective effect and the uniform quantisation of sensor pixels.

4.2. ANALYSIS OF RESULTS

With the statistic results for six and five cameras respectively we can demonstrate that the accuracy of the person localisation is approximately 35 centimetre average over all cameras. Due to the general inaccuracy in the far-field of all cameras these results show that the use of multiple overlapping cameras is justified for this surveillance system to ensure the objects are accurately located on the airport apron.

5. OBJECT CATEGORISATION EVALUATION

Evaluation task of object categorisation consists of recognising objects properly classifying them into predefined classes. Aims of this classification are: First, to provide a coarse classification, and second to differentiate objects for each categorised class. Due that, object categorisation task was divided into two sub-tasks:

- Categorisation itself: The categorisation itself task decides whether the object was correctly classified in the category or not.
- Recognition of the object in the category: When the object was correctly classified in its category, the object recognition task evaluate whether the category type of the object was correctly assigned or not.

For recognition task, only the vehicle category was evaluated. Although person category might be subdivided into one person or a group of people, such division was not considered in current evaluation.

5.1. EXPERIMENTAL RESULTS

The evaluation procedure was done as follows:

For each sequence, the evaluation was done frame by frame checking whether objects present in the scene were properly classified into the appropriate category or not. At the same time, the recognition of the object into the subcategory was checked. When the classification of the object corresponds with the real meaning of the object, a true positive was counted. In case that the application assigns wrongly a class to some object, a false positive was counted. Note, that this evaluation method needs implicit knowledge of the human evaluator.

The achieved categorisation is done although the complexity of the scene.

Object categorisation:

- A common error during categorisation process was the no-consistency between consecutive frames. While the hierarchical model keeps the consistency between the frames, the GMM model suddenly changes the subcategory.
- Some times category and subcategory were not assigned correctly.

Categorisation of one object might change from one frame to the next ones. This situation is considered as a special kind of error of categorisation process. Such errors might occur due to different reasons. For example, if the object is partially occluded or features used during the categorisation change their values from one frame to another one.

Obviously objects which are not detected and/or tracked properly can not been categorised because there is nothing to categorise. This type of situations is considered as an object detection of object tracking error and not a categorisation error.

It is shown in which category objects were wrongly classified. Such information can be useful to improve the object detection and object recognition.

5.2. ANALYSIS OF RESULTS

We can observe that some classification errors during classification into the categories happen. This occurs especially in two sequences. We think that the reason of this is that the features used during the categorisation process were not properly detected, therefore the categorisation fails.

Considering classification into subcategories, we note that more errors happened than in the categorisation process. Again, the reason of that might be that the correct features to perform the classification were not properly detected.

Comparing both methods, i.e. Hierarchical and GMM, we note that GMM produce less false positives than the Hierarchical method, while both true positive rates have the same order.

6. SUMMARY

The intermediate prototype scenes tracking evaluation demonstrates that the object tracking module detects a high proportion of the objects in the scene and these objects are tracked over extended time periods.

The prototype has illustrated the capability to process tracking on all cameras field of view. To demonstrate it, qualitative tests have been performed on several video sequences considering the complexity and the number of vehicles and individuals in the scene.

Under severe partial occlusions we have found that the tracks become fragmented and lose the track ID.

Track localisation has been shown to be accurate for vehicles and people, although naturally the accuracy reduces further from the camera sensor.

Considering object categorisation, it is necessary to improve the categorisation method in order to avoid misclassification into the subcategories. A possible method to apply is to consider information of previous frames when an object has been previously detected and it was categorised. Such information might be possible to carry on from one frame to another like a history of the object's category. Further research is also necessary to keep consistency between frames.

Future work on the object tracker is to improve the prediction of the bounding boxes when object are undergoing occlusion and to retain the object ID's during this period. We also plan to work on the reduction of the influencing of ghosts, reflections and shadows on the tracking procedure.