AVITRACK

Contract n° AST3-CT-2003-502818

Deliverable D6.1 C
Prototype scenes tracking
Evaluation

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

Table of contents

1. INTRODUCTION

This document presents the final performance evaluation results of the AVITRACK tracking prototype. Subject of evaluation is Object Tracking and Object Categorisation.

The framework ViPER has been chosen for the annotation of the AVITRACK apron sequences. The tool ARC Evaluator has been used for the evaluation of the object tracking module.

Object categorisation module has been evaluated by checking frame by frame whether objects present in the scene are properly recognised and classified into the appropriate class.

1.1. VIDEO DATASETS

The prototype scene tracking evaluation assesses the performance of the object tracking and categorisation components on representative test data. The evaluation of the components strongly depends on the choice of the video sequences. The consortium has chosen video datasets containing realistic conditions for an objective evaluation.

2. PERFORMANCE EVALUATION TOOLS

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.

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.

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. However, such generated data does not allow an automatic evaluation process that's why an evaluator tool was developed in order to achieve an automatic evaluation process. The developed ARC Evaluator Tool is completely automatic and no human intervention is required.

2.1. ARC EVALUATOR TOOL

The ARC Evaluator Tool is composed of three modules: One named object-matching module, a second one referred as format conversion and a third module called evaluation.

1. Object-matching module

The object-matching module 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.

The ARC Evaluator tool uses the XML format generated by ViPER.

2. **Format conversion module**

Current module takes as input the internal xml results generated by the object-matching module. These results stored in an xml file contain the information of matched objects ordered by frame.

3. **Evaluation module**

The current module takes the information generated by the format conversion module and produces performance statistics results for each ground truth object (GTO) and tracked object (TO).

The following metrics define by Black are used to characterize the tracking performance:

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

The *TRDR* and the *FAR* metrics characterize the performance of the tracker. The *TDR* metric determines the completeness on individual ground truth objects. The *TF* metric determines the number of object label changes. It is desiderable to have a *TF* value of one.

3. OBJECT TRACKING EVALUATION

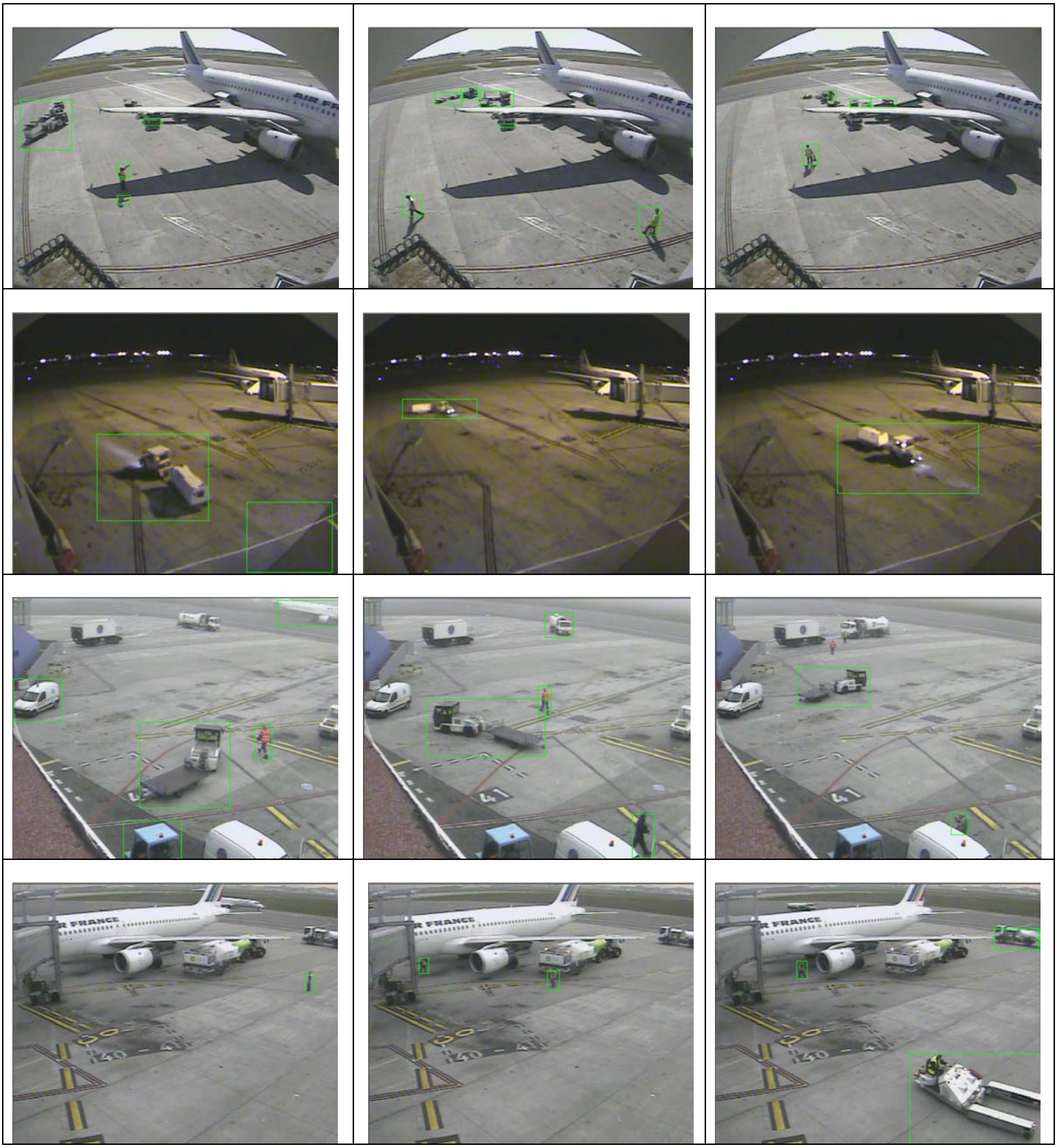
- To evaluate the performance of the local feature tracking method eight apron datasets were chosen. The datasets were taken under a wide range of disturbing conditions such as illumination changes, occlusions, reflections and shadows.

Sequences were acquired:

- On a sunny day
- By night
- With the presence of fog
- On a cloudy day

The datasets have been manually annotated using ViPER annotation tool. The ARC 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.

Next representative results of the local feature tracking method for the selected datasets are presented.



3.1. EXPERIMENTAL RESULTS

- Weak shadows are detected and tracked as part of the mobile objects,
- Strong shadows are correctly detected and removed from the individuals ,
- The large strong shadow is also correctly removed,
- In some cases, strong shadows are detected as individual mobile objects,
- Reflections are produced from the paints on the ground when objects move close to them.

These reflected paints are detected and tracked as part of mobile objects,

- In the night sequence the lights of both a vehicle are reflected on the ground. These reflections are tracked as mobile objects.

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 apron datasets present high track detection rates (close to 90%). But sometime, dataset contain several dynamic occlusions causing object label changes.

In addition, the tracker detection rate *TRDR* and the false alarm rate *FAR* were calculated for whole frames. The complexity of the scene, which includes a high amount of occlusions, causes a considerable number of false negatives provoking the decrease in *TRDR* (83%). the scene which contain reflections causing the increase in *FAR*. the scene which contain strong shadows are separately tracked from individuals as mobile objects causing the increase in *FAR*.

4. 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, the evaluation of object categorisation was divided into two sub-tasks:

1. Coarse categorisation: This task decides whether the object was correctly classified in the main category or not.

2. Recognition of the object in the category: When the object was correctly classified in its category, the object recognition task evaluates whether the category type of the object was correctly assigned or not.

Table 1 describes the possible categories of the objects in the evaluated scenes and for each category the related subcategories are enumerated. The subcategories are necessary in order to differentiate objects with similar task or purpose (e.g. vehicles).

Category	**Sub categories**
Aircraft	Aircraft
Vehicle	GPU, tanker, transport and dollies, service, loader, conveyor belt, tow tractor, mobile stair
Person	One person, group of people
Equipment	Container, Jet bridge
Other	Other

Table 1: Category and subcategories of the objects.

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

For current evaluation, ten sequences were considered. For each sequence, only one camera was evaluated. The selected sequences, their associated camera, and the objects present in each evaluated sequence are:

The objects considered in the performance evaluation of the categorisation module are:

- Aircraft, Jet Bridge.
- GPU
- Conveyor belt
- Tanker, person
- Tanker, transporter, service, GPU, two people
- Transporter, loader
- Catering vehicle
- Three loaders, transporter, GPU
- Tanker, transporter
- Tow tractor

4.1. EXPERIMENTAL RESULTS

The evaluation procedure was performed 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 by its subcategory was checked. When the classification of the object corresponds with the real meaning of the object, a true positive is counted. When the application assigns an incorrect class to an object, a false positive is counted.

Figure her below, depict categorisation results. Note, that this evaluation method needs implicit knowledge of the human operator.

A common error during the 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 category and subcategory.

The hierarchical model is more consistent keeping the category between frames than the GMM model. However, the hierarchical model is not able in some cases to determine the subcategory of the objects while the GMM model assigns the subcategory to the objects correctly.

The no-consistency between frames might occur due to the following reasons. For example, if the object is partially occluded or the features used during the categorisation change their values from one frame to another one.

It is clear that objects which are not detected and/or tracked properly can not be categorised. These situations are considered either object detection or object tracking errors and not as categorisation errors.

4.2. ANALYSIS OF RESULTS

The errors that occur during the category classification task appear especially in certain sequences. The reason for this is that the bottom-up features used during the categorisation process are not properly detected, and therefore the categorisation algorithm fails. But note the high accuracy obtained on the other evaluated sequences, where a classification rate greater than 85% is achieved, and in many cases this rate is greater than 95%.

Considering classification into subcategories, we note that more errors happened than in the categorisation process. The reasons of these errors are first, the similarity of various types of vehicles; and second an incorrect model fitting by the search algorithm, which finds a local minimum instead of a global one.

Comparing both methods, i.e. Hierarchical and GMM, we note that Hierarchical model produce less false positives than the GMM method in the classification of objects into categories. In the sub-categorisation process is the GMM model, which produces less false positives than the Hierarchical model.

5. SUMMARY

The final 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.

The results are encouraging for the categorisation process. Care must be taken to handle errors propagated during object detection and object tracking modules. In case of sub-categorisation, it is necessary to analyse the detected features and the sub-categories assigned by the fitting model. Further research is 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. Furthermore, the influence of ghosts, reflections and shadows on the tracking procedure might be subject of future research.