DIPLODOC road stereo sequence

Michele Zanin, Stefano Messelodi, Carla Maria Modena
Fondazione Bruno Kessler

FBK Technical Report #ID 164010, April 2013

Abstract

In this note we describe a road stereo sequence acquired, stored and labelled for evaluation purposes in the DIPLODOC project. The sequence, covering different scenarios, and its ground-truth, is made freely available on the web as a common base for the evaluation of road and obstacle detection algorithms.

Keywords: stereo video sequence, ground-truth, road detection, obstacle detection.

1 Introduction

DIPLODOC (DIstributed Processing of LOcal Data for On-line Car services) [1] is a three years project partially funded by the *Provincia Autonoma di Trento* that started on April, 2002. Three research partners are involved: ITC (now called FBK), CRF (*Centro Ricerche FIAT*), and the University of Trento.

The goal of the project is to design and develop a system based on a distributed architecture where intelligent vehicles communicate with a remote traffic control center. Each vehicle integrates different technologies to provide more comfort and driver safety. Speech recognition and synthesis techniques are used to interact with the user. Computer vision and image understanding are applied to the extraction of traffic parameters and to accident avoidance by detection and recognition of obstacles on the road. Wireless telecommunication is used to send and receive traffic data and route planning information to/from the control center.

One of the service, in which TeV, the Computer Vision team at FBK, is strongly involved, is the Front Obstacle Recognition (FOR) service [2, 3]. It aims at warning the driver when pedestrians, vehicles or obstacles are

Figure 1: The dimonstrative DIPLODOC vehicle

in close proximity to the driver's intended path, using information coming from the on-board vision sensors, the vehicle data (actually the speed), and possibly from environmental conditions (ice) or driver activity to modulate the alarm level. This service works fully on-board without exchanges with the control center and one of the main requirements is operating in hard real time.

The vehicle is endowed with two image acquisition devices: a color stereo camera head and a monocular high resolution camera, both mounted in the front of the vehicle and looking ahead. The obstacle detection system receives its input from the stereo device. We employ a couple of IEEE 1394 digital cameras distributed by Videre Design together with the provided software for disparity map computation: Small Vision System (SVS) by SRI International [4, 5]

In the following the real world road stereo sequence stored and labelled for quantitative evaluation porposes is presented.

2 DIPLODOC stereo sequence

In the development of computer vision applications, the availability of common datasets of annotated images and videos (ground truth) plays a fundamental role.

Several stereo sequences were acquired and saved during the development of the DIPLODOC project. They were used mainly for testing and evaluating the algorithms from a qualitative point of view. Portions of a particular sequence, among the available ones, was selected for quantitative measures of the algorithms performances.

The sequence, stored on a frame basis, is the composition of five subsequences, each of them covering different scenarios and environment in traffic and road conditions: from highway-like roads to urban scenarios with crossroads, parking lots and complex environment; from congested traffic to completely free road. See some examples in Figure 2.

Totally, there are 865 image pairs taken from the stereo camera of the moving vehicle (an example is in Figure 3), manually segmented to define the road portion.

Although not very large, the labelled sequences are proposed as ground-truth for performance evaluation of road detection algorithms. The main limitation to collect large scale ground truth is the amount of time and human effort needed to generate high quality ground truth.

The images were captured on July, 16 2004, about 11 a.m., near Trento, Italy, in a sunny day, using the DIPLODOC prototype vehicle. The map in Figure 4 shows a satellite view of the sequence path and the segments where the subsequences have been selected.

The acquisition device is a Videre Design MEGA-D stereo camera pair installed near the rearview mirror. The stereo pair is calibrated (Figure 5) with the SVS software and the camera parameters are included in the downloadable file.

3 Labelling

We have manually annotated the five stereo sequences with the road region using the left image of the stereo couple. The labelling has been performed by means of a Tcl/Tk graphical interface [6], developed for this and similar tasks.

As in many segmentation tasks, problems arise in conceptually defining the target, in our case "the road". People can give many different definitions and variations of what the road region is:

- anywhere a car can drive (asphalt or non-asphalt)
- everywhere a car could drive without going up a step
- zone whatever is asphalt

Figure 2: Some representative images of the considered stereo sequence: Highway-like road, urban scenario, parked cars, a complex environment scene, congested traffic, completely free road

Figure 3: Example of a couple of stereo images (frame number 202)

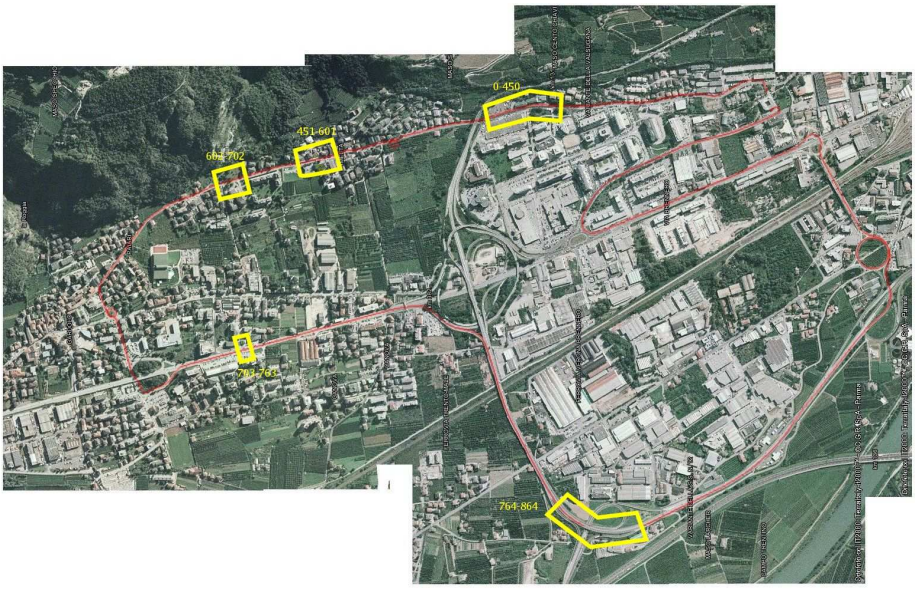

Figure 4: Trip map and sequence segments

Figure 5: Extrinsic camera calibration with a checkerboard rigidly attached to the vehicle. a) lateral view (b) internal view of the three cameras looking at the calibration device.

- asphalt zone delimited by line marking
- any asphalt zone where a car can drive on (parking lots, bus stop, entrace to private garages included)

These examples show how challenging is the ground-truth definition for evaluating a road segmentation algorithm. A consequence is that the comparison of road detection algorithms is also difficult, even on the same sequence, if a common ground-thruth is not provided.

In our case, the definition of road given to the operator was: "*everywhere a car could drive without going up a step*". The ground truth for an image is saved as a set of polygons. Some of them represent ideal road regions, some of them represent objects, like vehicles, that occlude portions of the road. From these polygons it is possible to compute the road region actually visible in the considered image. Figure 6 presents an example (frame no. 202), with one road region and two occluding regions. Figure 7 depicts the derived visible road region, computed as the difference from ideal road and occluding objects.

4 Download

The stereo sequence is available for download at the address: `http://tev.fbk.eu/DATABASES/road.html`

The compressed archive file has a dimension of 224 330 461 bytes. In the uncompressed folder we find a text file with the description of the package content, a `.ini` text file with the camera calibration data, obtained by the

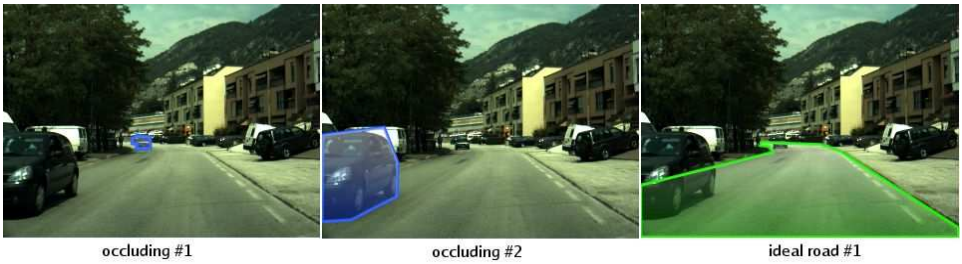

Figure 6: Two road occluding objects, the ideal road region

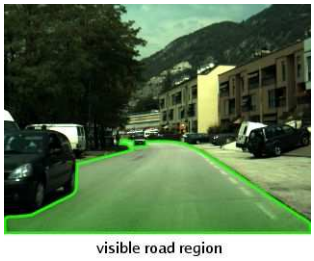

Figure 7: The real visible road

SVS calibration module, the 865 couple of images and their ground-truth, and finally a text file with the timestamps of the images, acquired at 15 fps. The color images are saved in a lossless format `png` without any pre-processing. Their dimension is 320×240. Every image pair is represented by 3 files:

```
diplo%06d-L.png
diplo%06d-R.png
diplo%06d-L.txt
```

where: `%06d` ranges from `000000` to `000864`, `diplo%06d-L.png` represents the left color image, `diplo%06d-R.png` the right one, `diplo%06d-L.txt` is a text file encoding the segmentation ground truth, manually performed on the left image.

Every line of the `diplo%06d-L.txt` file represents a polygon and has the following structure:

```
type N x1 y1 x2 y2 ... xN yN
```

where:

type can be 'road' (an ideal road region) or 'occl' (a region that overlaps the road occluding a part of it);

N is the number of vertices of the polygon;

xi yi are are the coordinates of the i-th vertex (for $i = 1, \ldots N$) normalized into the $[0, 1]$ interval.

Given an image, if R_i, with $i = 1, \ldots N_r$, are the regions labelled as 'road' and C_j, with $j = 1, \ldots N_c$ are the regions labelled as occluding the road ('occl'), the visible road region of the image is obtained as:

$$V = \bigcup_i R_i \setminus \bigcup_j C_j \tag{1}$$

The five subsequences are represented in the following frame intervals:

1 0 - 450 (450 frames)

2 : 451 - 601 (100 frames)

3 : 602 - 702 (100 frames)

4 : 703 - 763 (60 frames)

5 : 764 - 864 (100 frames)

5 Conclusions

The five subsequences, together with their ground truth, are available for download in a single package at [7].

Please, cite the present technical report when referring to the DIPLODOC Road Stereo Sequence in your publications.

References

[1] `http://www.dit.unitn.it/~diplodoc` The DIPLODOC project web site, 2001.

[2] C. Corridori, D. Giordani, P. Lombardi, S. Messelodi, C.M. Modena, and M. Zanin. An in-vehicle vision system for dangerous situation detection. In A. Milani, editor, *Conferenza Italiana sui Sistemi Intelligenti - CISI*, page 10 p., Perugia, Italy, September 2004. Morlacchi Editore.

[3] M. Zanin. *Vision-based Road Recognition and Obstacle Detection for Intelligent Vehicles*. PhD thesis, International Doctorate School in ICT - University of Trento, March 2005-2006.

[4] `http://www.ai.sri.com/~konolige/svs` SRI Stereo Engine web site , 1997.

[5] K. Konolige. Small vision systems: Hardware and implementation. In *Eighth International Symposium on Robotics Research*, pages 209–214, Hayama, Japan, October 1997.

[6] N. Peroni. Localizzazione della strada in immagini digitali riprese da bordo veicolo: algoritmi di segmentazione e sviluppo di un ambiente di valutazione. Master's thesis, Ingegneria delle Telecomunicazioni, Università degli Studi di Trento, 2002-2003.

[7] `http://tev.fbk.eu/DATABASES/road.html` DIPLODOC road stereo sequence , 2005.