TRACKING CARS IN RANGE IMAGE SEQUENCES

Esther Bettina Meier and Frank Ade

Communication Technology Lab, Image Science

Swiss Federal Institute of Technology (ETH), CH-8092 Zurich, Switzerland

Keywords: autonomous vehicles, optical distance measurement, image region analysis

ABSTRACT

Cars will eventually be equipped with a control system to either warn the driver when he gets too close to another car, or even to keep him automatically at an adequate distance.

In both cases automatic traffic scene analysis will be needed. We want to base this on a new optical range sensor which acquires distance and intensity information. We present a detection and tracking scheme which is fast, simple and operates with a small basic instruction set. Our region growing algorithm ensures robust segmentation and detects objects even in ambiguous data.

Objects are represented by bounding boxes and tracked along the image sequence. The tracking process also provides the velocities for calculating the time to collision with each object.

We show results of some experiments with simulated traffic scenes.

INTRODUCTION

Modern everyday life is characterized by an ever increasing interaction between man and machines, leading to a growing number of potentially harmful situations for human beings. When using public or private transportation systems (cars, trains, buses, etc.), the health of users might be endangered through accidents, malfunctions or human oversight. Traffic surveillance systems and smart cruise control systems are of increasing interest. These problems are being attacked by a number of research groups. Most approaches use black and white or color cameras. Obstacles are detected by their shape (certain edge configurations [1] or symmetries [2]), by stereo [3, 4] or by motion [5, 6]. Others have worked with radar, infrared and range sensors [7].

The MINORA project (a Swiss priority program of OPTIQUE II, funded by the ETH Council) aims to develop a new optical matrix range sensor working in the near infrared which should be fast and *cheap*. It will be mounted on a car and acquires distance and intensity images of the traffic scene in front of it. We introduce a segmentation and tracking scheme that makes use of both sources of information. The developed algorithms have to process the distance and intensity images in milliseconds and must be simple enough to be implemented on dedicated hardware.

Our approach is based on region growing. It is not possible to obtain a successful segmentation on either the distance or the intensity image alone. On the other hand, the simultaneous use of both data sources can result in an oversegmentation of the scene. However, we propose to exploit the respective advantages of the two types of data in a hierarchical scheme.

Each detected object has size, 3D position and velocity vector as attributes. This information is already sufficient to track the objects in consecutive images.

SIMULATING THE SENSOR

Our work is based on novel types of active ranging techniques [8, 9] (coded time-of-flight or AM laser radar) with which we can measure distances of up to $150\ m$ and at an accuracy of about $0.2\ m$. Typically, the range sensor has a low angular resolution and covers an angle of about $\pm 10°$ horizontally and about $\pm 5°$ vertically. The developed segmentation algorithms have to deal with small, spatially undersampled, incomplete and ambiguous data.

As it will need some time until the project partners can provide the first real range data, we have written a simulation package to generate artificial data. This data can be used to develop and test the segmentation and target tracking algorithms. We simulate a virtual world in which a car containing the range sensor is driven through scenery that includes roads, hills, trees, houses, as well as vehicles. The range sensor is virtually placed in the reference car and calculates the distance and intensity images at given time intervals. The distance image is directly calculated from the information in the scene and a noise model can be added to create more realistic data. The intensity image is calculated using a common computer graphics illumination model [10, 11].

SEGMENTATION BY REGION GROWING

Region growing is a simple method for image segmentation [12, 13, 14]. Many algorithms have been proposed based on region growing which use one or several uniformity criteria to compare the characteristics of a pixel and its neighborhood.

In our implementation we use the pixel values (intensity or distance values). $\overline{X}$ is the mean of a region, which is updated whenever a new pixel is added to the region.

$$\overline{X}_{N+1} = \frac{N}{N+1}\overline{X}_N + \frac{1}{N+1} f(x,y) \qquad (1)$$

The value of the current pixel is $f(x,y)$ and N is the number of pixels which already belong to the region. Basically, the growing starts at a seed pixel and new neighboring pixels are included if they fulfill the criterion

$$|\,\overline{X}_N - f(x,y)\,| < \epsilon. \qquad (2)$$

To find all homogeneous regions in an image we start with the pixel at the top left corner. All pixels which belong to the region will be marked. To determine the next region start pixel, we look from left to right, top to bottom for the first non-marked pixel in the image. The segmentation results below are represented by assigning different color levels to different regions.

Intensity Images

Using the region growing algorithm on our intensity images, we obtain a number of homogeneous regions for the background, the bodywork, the license plate and the reflectors of the car. It is desirable to find a homogeneous region for the background which makes object tracking much easier. How many different regions of the car are detected depends on ϵ_I (Fig. 1).

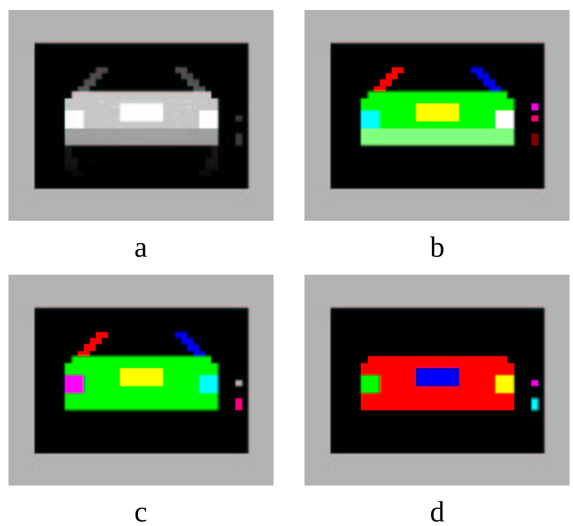

Figure 1: a: Intensity image; b: Segmented image with $\epsilon_I = 0.15$ which is about $8\ \%$ of the maximal received intensity; c: Segmented image with $\epsilon_I = 0.3$; d: Segmented image with $\epsilon_I = 0.45$;

Distance Images

We can apply the region growing algorithm also to the distance image. Here, we detect the car as one object; but we obtain many regions for the road and the ground, which we would prefer to avoid (Fig. 2). We are just interested in detected objects on or beside the road. The advantage of using the distance image is that we obtain a uniform region for the detected car.

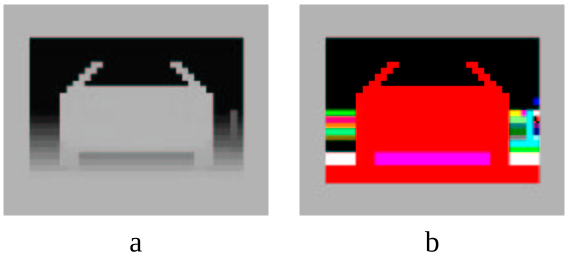

Figure 2: a: Distance image; b: Segmented image with $\epsilon_D = 1.1\,m$;

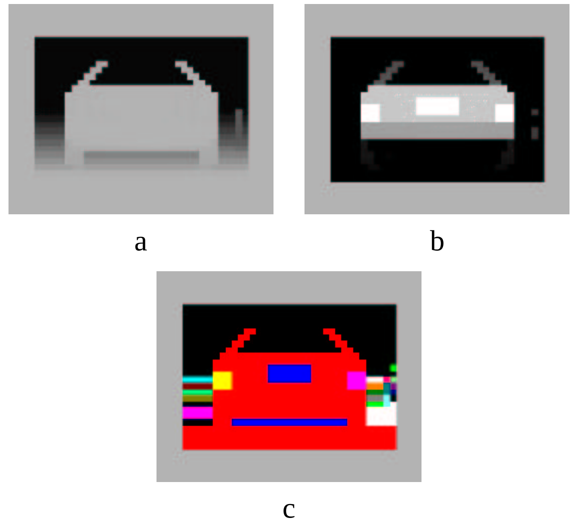

Figure 3: a: Distance image; b: Intensity image; c: Segmentation using both images simultaneously with $\epsilon_D = 0.7\,m$ and $\epsilon_I = 0.4$ which is about $22\,\%$ of the maximal received intensity;

Intensity and Distance Images

Using the intensity and distance information together in the region growing algorithm (two conditions have to be fulfilled), we have to cope with the disadvantages of the segmented distance and intensity images. We get subdivisions of the car, the road and the ground (Fig. 3).

HIERARCHICAL APPROACH

For the image segmentation we would like to combine the advantages of the two data modalities.

Our approach is to segment first the intensity image and then refine the result using the distance image. Two types of errors occur:

1. A single object is segmented into different regions because its parts have different reflectance properties.
2. Several objects are recognized as one region since they have similar intensity values and are close to or occlude each other.

The first problem is solved by calculating a mean distance for each region and then combining close regions with a similar mean distance (see Fig. 4).

For the second problem the regions obtained from the intensity image are segmented again, this time using the distance image (see Fig. 5). Regions belonging to the background or road are identified as such and therefore not subject to a second segmentation step. On the one hand we can extract them with a threshold or we are able to localize the ground, as we know the sensor geometry with respect to the road.

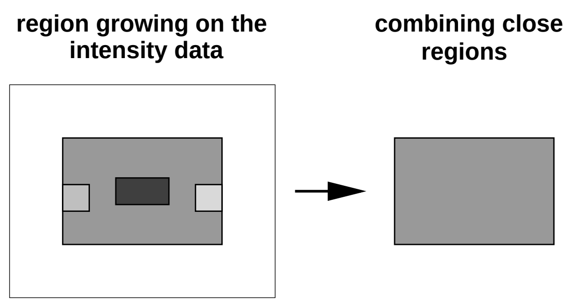

Figure 4: Combining neighboring regions with approximately the same mean distances.

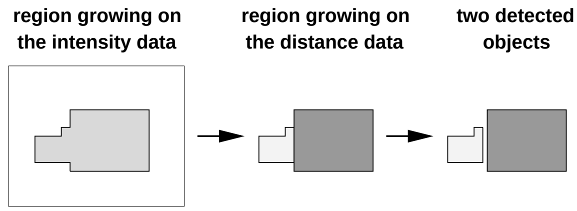

Figure 5: From one homogeneous region – which we get from the region growing algorithm on the intensity image – we get two detected regions on the corresponding distance image.

TRACKING

Two factors make tracking difficult in our case: spatial undersampling and mutual occlusion of vehicles.

1. Undersampling has the effect that we can not always detect small objects (e.g. a road post) in consecutive frames. When the object reappears later it will be regarded as a new object.
2. The positions and sizes of occluded objects change considerably and the tracking process becomes more difficult.

We use the bounding box of a region for object representation thus simplifying object management. Additionally we calculate the 3D center of gravity and the velocity vector $\vec{v}$. Two objects in successive frames are matched if their 2D bounding box corners and the 3D centers are not further apart than a specified distance δ_{2D}, respectively δ_{3D} . We compare the expected 3D center of gravity $\vec{r}_t$ and the actual center $\vec{s}_t$ at time t.

$$\vec{v} = \frac{\vec{s}_{t-1} - \vec{s}_{t-2}}{\Delta t} \tag{3}$$

$$\vec{r}_t = \vec{s}_{t-1} + \vec{v}\,\Delta t \tag{4}$$

$$\|\vec{r}_t - \vec{s}_t\| < \delta_{3D} \tag{5}$$

For a new object where we can not yet calculate the velocity vector, we look at the displacement between the 3D centers of gravity $\vec{s}_{t-1}$ and $\vec{s}_t$, and compare this to a threshold θ.

$$\|\vec{s}_t - \vec{s}_{t-1}\| < \theta \tag{6}$$

The high frame rate and the comparatively small relative velocities limit the distance a vehicle can move between frames. The tracking in a 3D object coordinate system increases the stability of the method.

The tracking process provides the velocities of the objects. Together with the distance this information can be used to calculate the time to collision which triggers a warning signal for the driver or reduces the speed of the car.

RESULTS ON THE SIMULATED RANGE IMAGES

We have tested our system on several simulated image sequences and present here the results of one 250-frame sequence.

In Fig. 6, from left to right, we see the simulated traffic scene, the calculated intensity and distance images, and the detected objects represented by their bounding boxes. Tracked objects can be recognized by having the same color value of the bounding box in consecutive frames. Our reference car has a speed of 40 km/h, the red car drives at 45 km/h and the velocity of the white car is 50 km/h. This sequence has a frame rate of 25 images per second and an image size of 24×35 pixels. The range sensor has a relatively narrow field of view ($12° \times 17.5°$) whereas the driver has a wider field of view.

Exactly as the real intensity the simulated intensity depends on the distance of the viewed objects. Therefore, with increasing distance the car gets darker in the intensity column. The intensity of the reflected light limits the seeing distance of the range sensor (not considering resolution effects).

At frame 27 we get approximately the same intensity for the dark and bright car. Our system solves this ambiguity by using the distance information and continues tracking without any difficulties.

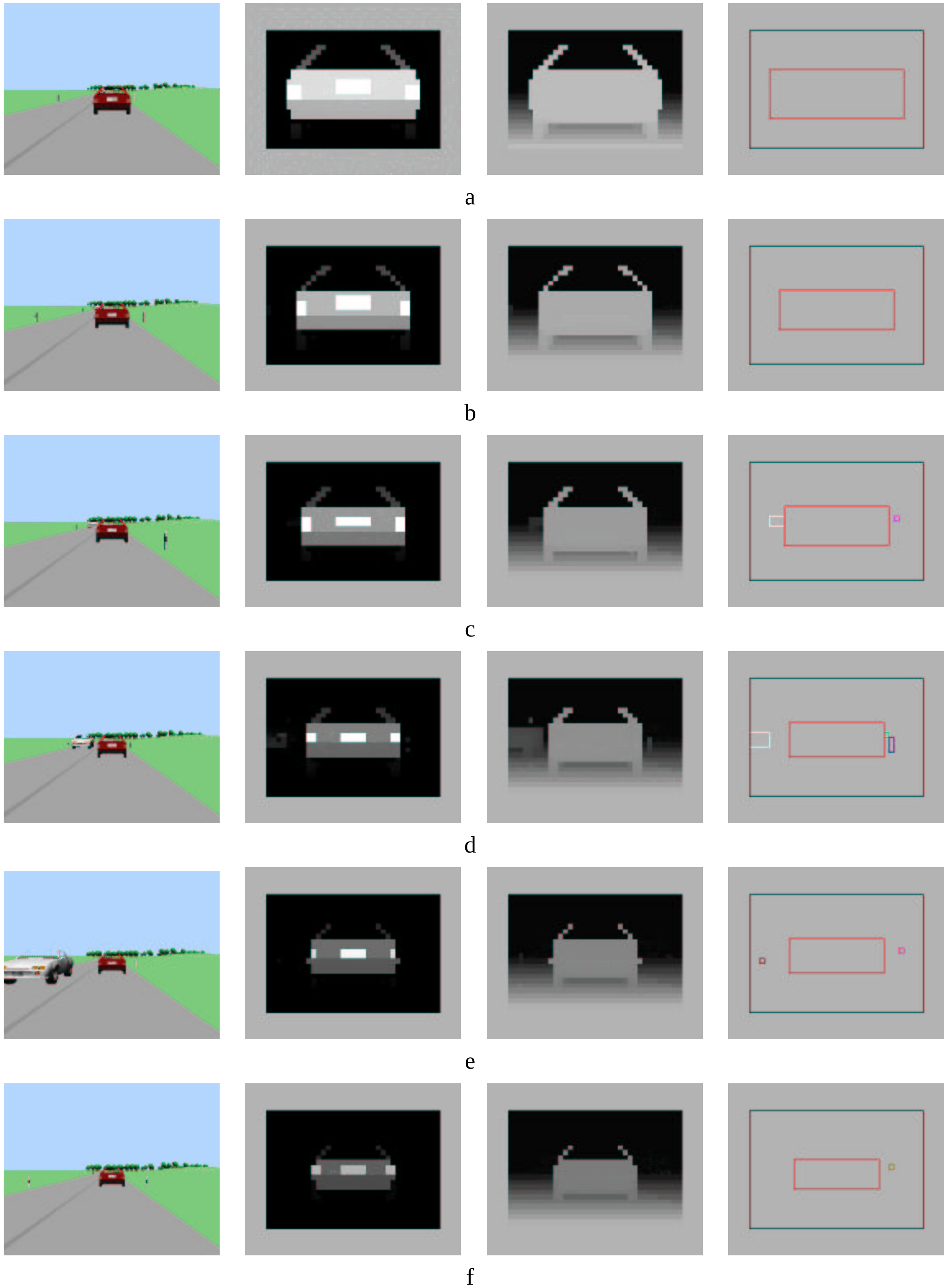

Figure 6: a: Frame 3; b: Frame 15; c: Frame 27; d: Frame 39; e: Frame 51; f: Frame 63;

CONCLUSION

We designed a robust system for detecting and tracking vehicles in front of a reference car. The problems of undersampled, incomplete and ambiguous data is solved by a hierarchical approach with a simple region growing algorithm, which is fast and allows to keep the instruction set of the processor simple. The objects are represented by their bounding boxes. The 3D center of gravity, the velocity vector and the 4 corners of the box help to track the objects.

References

[1] Dieter Koller, Joseph Weber, and Jitendra Malik. Towards realtime visual based tracking in cluttered traffic scenes. *Proc. 2nd IEEE Symposium on Intelligent Vehicles*, pages 201–206, 1994.

[2] Thomas Zielke, Michael Brauckmann, and Werner von Seelen. Intensity and edge-based symmetry detection with an application to car-following. *CVGIP: Image Understanding*, 58(2):177–190, September 1993.

[3] Q.-T. Luong, J. Weber, D. Koller, and J. Malik. An integrated stereo-based approach to automatic vehicle guidance. *5th International Conference on Computer Vision*, pages 52–57, 1995.

[4] J. Weber, D. Koller, Q.-T. Luong, and J. Malik. New results in stereo-based automatic vehicle guidance. *Proc. 3th IEEE Symposium on Intelligent Vehicles*, pages 530–535, 1995.

[5] Andrea Giachetti, Marco Cappello, and Vincent Torre. Dynamic segmentation of traffic scenes. *Proc. 3th IEEE Symposium on Intelligent Vehicles*, pages 258–263, 1995.

[6] B. Heisele and W. Ritter. Obstacle detection based on color blob flow. *Proc. 3th IEEE Symposium on Intelligent Vehicles*, pages 282–286, 1995.

[7] K. Sobottka, X.-Y. Jiang, and H. Bunke. Spatiotemporal segmentation of range image sequences into planar surfaces for collision avoidance. *30th International Symposium on Automotive Technology and Automation (ISATA), Special Session on Machine Vision and Intelligent Vehicles and Autonomous Robots*, pages 69–76, 1997.

[8] T. Spirig, M. Marley, and P. Seitz. The multitap lock-in CCD with offset subtraction. *Submitted for Publication in IEEE Transactions on Electron Devices*, 1997.

[9] T. Spirig, P. Seitz, O. Vietze, and F. Heitger. The lock-in CCD — two-dimensional synchronous detection of light. *IEEE Journal of Quantum Electronics*, 31(9):1705–1708, Sept. 1995.

[10] James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics, Principles and Practice*. Addison-Wesley Publishing Company, Inc., 1993.

[11] Nadia Magnenat-Thalmann and Daniel Thalmann. *Image Synthesis, Theory and Practice*. Springer-Verlag, 1987.

[12] Robert J. Schalkoff. *Digital Image Processing and Computer Vision*. John Wiley and Sons, Inc, 1989.

[13] Yu Xiaohan, Juha Ylä-Jääski, Olli Huttunen, Tuomo Vehkomäki, Outi Sipilä, and Toivo Katila. Image segmentation combining region growing and edge detection. *International Conference on Pattern Recognition*, C:481–484, 1992.

[14] R. M. Haralick and L. G. Shapiro. Image segmentation techniques. *Computer Vision, Graphics, and Image Processing*, 29:100–132, 1985.