AQTR 2004 (THETA14)
2004 IEEE-TTTC – International Conference on Automation, Quality and Testing, Robotics
May 13-15, 2004, Cluj-Napoca, Romania

Dynamic Traffic Description Using Stereovision Equipped Vehicles and Ad-hoc Wireless Networking

Sergiu NEDEVSCHI, Radu DANESCU, Dan FRENTIU, Tiberiu MARITA, Florin ONIGA, Ciprian POCOL

Technical University of Cluj-Napoca, Computer Science Department
E-mail:Sergiu.Nedevschi@cs.utcluj.ro, team@vision.utcluj.ro

Abstract: In this paper an Intelligent Vehicle Interaction System is presented. By connecting neighbor vehicles into an ad-hoc wireless network, each of them can broadcast to the others their GPS position, size and dynamic behavior, allowing them to build a dynamic map of the local traffic. However, some pieces can be missed: not all the vehicles are equipped with such systems, failures can appear, general obstacles can exist on the road. The solution is to use obstacle detection sensors. A stereovision sensor can detect obstacles in front of the vehicle, but it is still not perfect, having limited field of view and limited detection distance. If each stereovision equipped vehicle broadcasts its detection, all the receiver vehicles can improve their perception of the surrounding scene. Each involved vehicle has to assemble its detection results and the results provided by others, as well as the information reported by vehicles about themselves. All the traffic data must be in the same coordinate system, and the GPS coordinates are the most suitable for this. The main aim of the system is to offer assistance to the driver. Since current GPS navigation systems can provide only a static map, such a system can bring dynamic information about the surrounding traffic.

Keywords: stereovision, sensor fusion, symbolic environment description, distributed computation, driver assistance.

1. INTRODUCTION AND PROBLEM DESCRIPTION

Stereovision is becoming more and more popular as a 3D measurement tool, having the advantage of being a passive method and also of providing a rich amount of 3D data. A stereovision sensor can identify objects in front of it, measuring their position and size. The detection can be noisy and can even miss objects, thus a tracking procedure can reduce the noise by averaging measurements and fill in the detection gaps.

Due to the fact that a single sensor covers a limited area, and because of the imperfection of detection, the fusion of multiple sensors becomes necessary.

Note: for the readability purpose of this paper, all the objects detected by the stereovision sensor are subject to be considered and named as “vehicles”.

In figure 1 it is shown a scenario of detected vehicles by the stereovision sensor mounted on vehicle V. Near vehicles (closer than 50m), like vehicle V3, are detected with high accuracy. Further vehicles (from 50m to 100m), like vehicles V4 and V5, are detected with lower and lower accuracy as the distance grows. Vehicles detected further than 100 meters, are considered not to be reliable in terms of their existence, and their size and position accuracy is low. As the boldness of depicted vehicles V3, V4 and V5 shows, the quality of detection is lower for higher distances. Vehicles V1 and V2 are not detected at all, being out of the field of view, which is about 30 degrees. Vehicle V6 is too far. Also, partial or total occlusions can cost the quality of detection.

Having the same scenario, let's consider that in figure 2, vehicles V1 and V4 are equipped with such stereovision sensors.

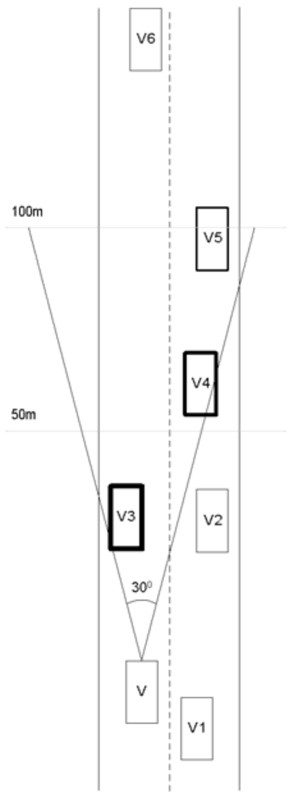


Fig 1. Vehicles V3, V4 and V5 are detected by vehicle V

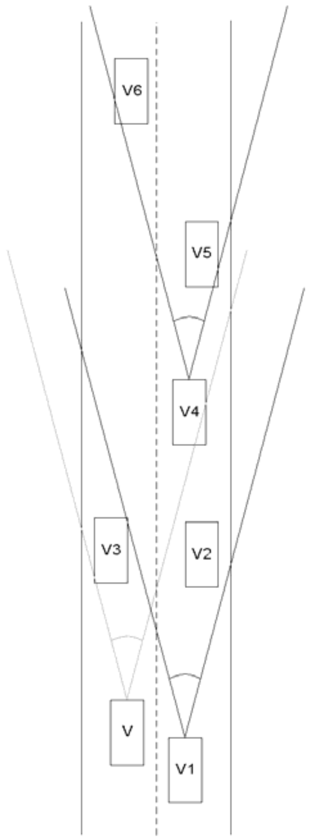


Fig 2. Vehicles V2, V4, V5 and V6 are detected by vehicles V1 and V4

The detection could be as follows:

For vehicle V1:

- V2: very good;
- V4: acceptable;
- V5: totally occluded;
- V and V3: out of field of view;
- V6: too far, not detected.

For vehicle V4:

- V5: very good;
- V6: acceptable;
- V, V1, V2 and V3: out of field of view.

If vehicles V, V1 and V4 are interconnected in a wireless network, each of them can get the detection results of the others and improve the perceiving of surrounding traffic. These vehicles are transmitters and receivers in the same time. Each of them, performs a reunion of its own detected vehicles and the vehicles detected by the others, by using a sensor fusion procedure to identify the common vehicles reported by different transmitters. Even more, the description about themselves is communicated to the others, having full reliability (its width and length are known a priori and the GPS information is reliable). If a vehicle doesn't have the stereovision sensor, but it has the GPS sensor and the wireless transceiver, it still can work by sending information about itself and fusing the information from the others.

A GPS sensor can provide information like: latitude, longitude, elevation, velocity, heading and time. The transmission range of wireless networks is 250m. If enough stereovision equipped vehicles run on a 500m portion of road, the central one can have a very good description of the surrounding traffic.

2. DEFINITION OF THE ENVIRONMENT MODEL

The environment is described as a set of objects (vehicles). Each stereovision sensor outputs its own set of objects. The fusion algorithm will join all object sets into a final one.

Each sensor relieves 3D measurements of the perceived objects. The used measurements are: length, width, 2D position on the Earth's surface, speed and direction of movement (fig. 3).

The scene has associated a unique coordinate system. The data from all the stereovision sensors is relative to this unique coordinate system. The GPS coordinate system responds to this request.

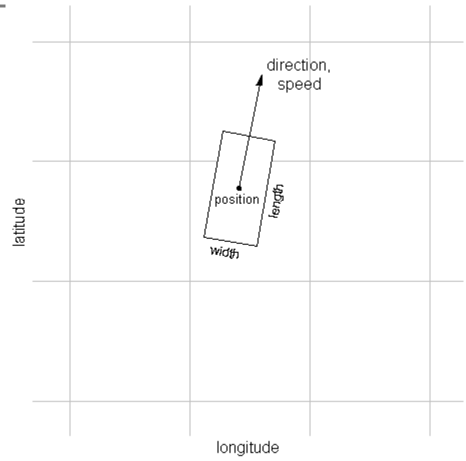


Fig 3. Vehicle's representation

3. THE SENSORIAL SYSTEM ARCHITECTURE AND FUNCTIONS

The stereovision sensor consists of two video cameras mounted on a rigid rig and a computing device (computer) – figure 4. The distance between cameras (base line) is about 30-40 cm. The rig is mounted inside the car, just before windshield. The stereovision detection is done in a coordinate system parallel with the car and with the origin situated where the GPS sensor is placed.

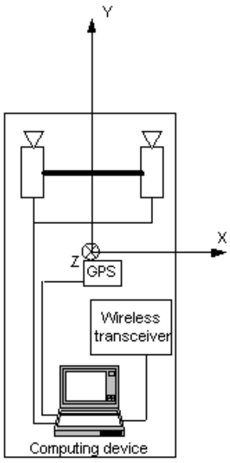


Fig 4. Needed equipment

Using a wireless transceiver, our vehicle sends the set of detected vehicles, and receives the sets from other stereovision equipped vehicles located in nearby.

The image processing computer performs stereo 3D reconstruction cycles on the synchronously acquired image pairs. The reconstructed 3D points are grouped into 3D objects (cuboids) which are tracked. The tracked objects represent the sensor's output and are broadcasted via the wireless transceiver. In the same time, the transceiver gets the sets of detected vehicles from the other stereovision equipped vehicles in nearby.

The synchronization of all the detected vehicles is done based on the GPS time. The GPS time is issued once per second. Aligned to the GPS time, at every 200 milliseconds, each sensor acquires a stereo pair of images, processes it and broadcasts the set of detected vehicles together with the corresponding timestamp. In the period of 200ms from the acquisition moment, all the detection results should be broadcasted by transmitters and received by receivers and fused. The fusion is quite simple, having the same timestamp for all the data.

4. STEREO 3D RECONSTRUCTION

The calibration process estimates the camera's intrinsic parameters (which are related to its internal optical and geometrical characteristics) and extrinsic ones (which are related to the 3D position and orientation of the camera relative to a global world coordinate system) [1,2,3].

Using the stereo system geometry determined by calibration, the stereo reconstruction algorithm finds pairs of left-right correspondent points and map them into the 3D world [4].

Constraints, concerning real-time response of the system and high confidence of the reconstructed points, must be used. In order to reduce the search space, only edge points of the left image are correlated to the right image points. For robust detection of the image edges, a Canny-based [5] edge detector was implemented. By focusing to the image edges, not only the response time is improved, but also the correlation task is easier, since these points are placed in non-uniform image areas. The sum of absolute differences (SAD) function [6] is used as a measure of similarity, applied on a local neighborhood. For a given left image point the search is performed along the epipolar line computed from the stereo geometry.

After this step of finding correspondences, each left-right pair of points is mapped into a unique 3D point [4]. Using the camera geometry, two 3D projection rays are traced, one for each point of the pair. By computing the intersection of the two projection rays, the coordinates of the 3D point are determined.

The result of reconstruction is a set of 3D points that must be clustered into objects. The grouping is performed mainly based on the local density of the points and the vicinity criteria: a local group of points must be dense enough to be considered as candidate and two points are considered to be in the same group if they are close to each other. Both these criteria are adapted to the fact that the density of reconstructed points per object decreases with the distance (due to the perspective projection) and their positioning error increases with the same distance. For each cluster of points, the circumscribing box is built.

5. OBJECT TRACKING

Tracking is employed in order to estimate the dynamic parameters of an object. The dynamic parameters are the positions and the speeds on both X and Y axes. The geometric sizes can also be established filtering them against the detection noise. The geometric sizes are: width, length and height.

The position and speed of the object are tracked through a linear Kalman filter. The speed vector is established as the resultant of the speeds on individual axes. The modulus of the speed vector gives the speed value itself and the vector's orientation shows the direction of movement.

The tracker will output the set of objects in the environment model format, being ready to be broadcasted.

6. DEALING WITH DIFFERENT COORDINATE SYSTEMS

The detection results of the stereovision sensor are relative to our vehicle. Some transformations have to be done, to move them into the global coordinate system.

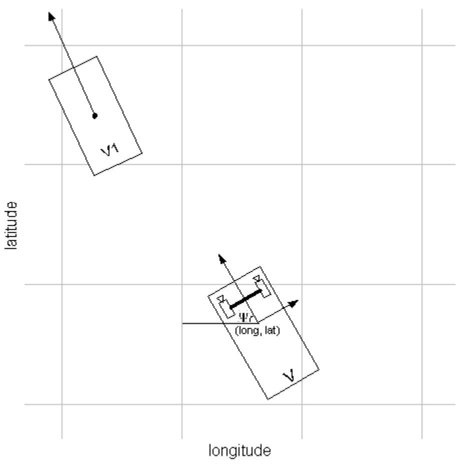


Fig 5. Our vehicle (V) and the detected vehicle (V1), in the global coordinate system.

The detected vehicles are in a coordinate system parallel with the car and with the origin situated where the GPS sensor is placed. By knowing the heading (orientation angle - ψ) and the global position (longitude and latitude) of vehicle V (more precisely of the GPS sensor), the detected vehicles are easily converted in the global coordinate system.

Our vehicle is added to the set of detected vehicles. Thus, its center, which differs by the GPS sensor's place, has to be correctly expressed by knowing the GPS sensor placement in the vehicle's geometry.

7. SENSOR FUSION ALGORITHM

Not all of the received detected vehicles are usable from temporal and spatial reasons. Some messages can arrive to the receiver too late, being outdated, and will be discarded. Other vehicles can be out of our space of interest. On highways, the broadcasted messages from the other driving direction are not important. Another similar situation appears at the intersection of highways, where one passes over other. Vehicles which are too far behind, don't represent a danger and can be disregarded. The space of interest can be considered 100m behind and 200m in front of us, even if messages can be transmitted over longer distances.

When transmitting the set of detected vehicles, data about the ego vehicle is transmitted first, in the same format as the detected vehicles. In this way, it is known that the first vehicle, in a set, represents a fully reliable one.

Each detected vehicle has associated a confidence degree, depending on its distance, number of 3D points inside it and the number of frames it was observed by tracking, all of these being weighted with their importance. Our vehicle has a 100% confidence degree.

When attempting to fuse the results of the sensors into a global result, we must make the difference between the case when an object is detected by only one sensor, and the case when it is detected by two or more sensors. In the first case, the act of fusion is simply to add this object to the global result set. In the second case, the result must be a combination of the sensor readings, taking into consideration the confidence degrees of each sensor.

The main simplification of the problem comes from the fact that the cuboids are defined in the same coordinate system.

The criterion that two objects occupy the same space (and therefore they could be joined) is literally interpreted: the intersection area of their rectangles is computed (figure 6). If this area, represents at least 50% of the smallest rectangle, the two objects can be joined. Because of imperfection of detection, tracking and GPS position, it is possible to have two very close objects obeying our criterion, apparently the same one, but if their direction of movement differs too much they will not be joined.

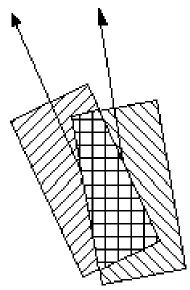


Fig 6. Intersection area of two vehicles from different result sets

If they are declared as joinable, the resulted one is computed from the two as a weighted sum of their parameters, using the confidence degree as the weight.

8. RESULTS

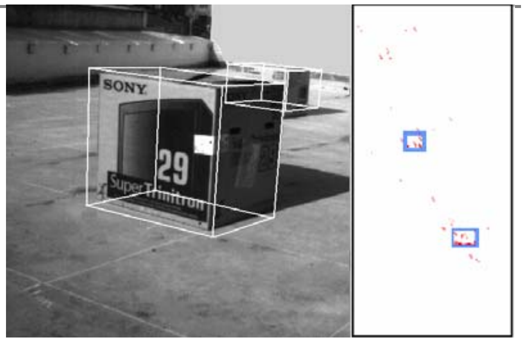
For testing of the algorithm we have used two stereovision setups. The two setups were calibrated using the method described in the calibration section, having the coordinate system on the ground, just under the stereo rig and oriented as the cameras. The perspective views of the scene for each stereovision sensor are presented in the left side of (fig. 7.a and 7.b). The reconstruction results for each stereovision sensor is presented as a bird-eye view of the scene in the right part of the same images, and as white cuboids projected on the original perspective image.

The two sets of results were gathered and, by knowing the relative position of the sensors, the fusion could be done. The position of the first sensor was considered as being the origin of GPS coordinate system.

The fusion results are displayed in fig. 7.c as a bird-eye view. The final result fulfils the aim of the algorithm: combining together the scene description of different sensors and refining the measurements of each sensor against each other, in the case where the same object is viewed by more than one sensor and in the case where an object is not viewed by one sensor but it is viewed by the other one.

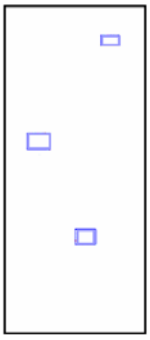
9. CONCLUSIONS

A method for extracting the 3D scene description from multiple stereovision sensors has been presented. The stereovision sensors are able to perform real-time image pair processing and extract 3D points of the environment. These points are then grouped into cuboids which are tracked in successive frames. The scene description of each sensor is sent to the other sensors through a wireless ad-hoc network. Fusion of scene descriptions is performed in order to obtain a more complete description of the scene. The fused description has the advantage of increasing global field of view by uniting the fields of view of each sensor, and the advantage of refining the description of individual objects, if they are viewed by more than one sensor.



a – Results of the first stereo sensor

b – Results of the second stereo sensor



c – Results of the sensor fusion

Fig. 7. Stereo reconstruction and object fusion results

REFERENCES

[1] Jean-Yves Bouguet (2003), “Camera Calibration Toolbox for Matlab”, MRL – Intel Corp.,http://www.vision.caltech.edu/bouguetj/calib_doc/

[2] S. Nedevschi, T. Marita, M. Vaida, R. Danescu, D. Frentiu, F. Oniga, C. Pocol, D. Moga (2002), “Camera Calibration Method for Stereo Measurements”, *Journal of Control Engineering and Applied Informatics (CEAI)*, Vol.4, No. 2, pp.21-28, Bucuresti, Romania.

[3] S. Nedevschi, T. Marita, R. Danescu, F. Oniga, D. Frentiu, C. Pocol (2003), "Camera Calibration Error Analysis in Stereo Measurements", *microCAD International Scientific Conference, March, Miskolc, Hungary*, pp. 51-56.

[4] Trucco E., Verri A (1998), “Introductory techniques for 3D Computer Vision”, New Jersey: Prentice Hall.

[5] J. Ramesh, R. Kasturi, B. G. Schunk (1996), “Machine Vision”, New York: McGraw-Hill Inc.

[6] Todd A. Williamson (September 1998), “A High-Performance Stereo Vision System for Obstacle Detection”, Ph.D. Thesis. Carnegie Mellon Technical Report