A Three Resolution Framework for Reliable Road Obstacle Detection using Stereovision

Mathias Perrollaz, Raphaël Labayrade, Romain Gallen, Didier Aubert

A Three Resolution Framework for Reliable Road Obstacle Detection using Stereovision

Mathias Perrollaz, Raphaël Labayrade, Romain Gallen, Didier Aubert

LIVIC (INRETS / LCPC)
Bât. 824, 14 route de la minière
78000 Versailles – France
mathias.perrollaz@inrets.fr

Abstract

Many approaches have been proposed for in-vehicle obstacle detection using stereovision. Unfortunately, computation cost is generally a limiting factor for all these methods, especially for systems using large baselines, as they need to explore a wide range of disparities. Considering this point, we propose a reliable three resolution framework, designed for real time operation, even with high resolution images and a large baseline.

1. Introduction

Stereovision is well suited for road obstacle detection, because it provides a complete three dimensional view of the scene. Many methods have been developed for that purpose and most algorithms use correlation-based stereovision [2]. However, as they systematically present a high computational cost, a compromise is necessary between precision and detection range. More precisely, the resolution of the stereo images and the disparity range are critical parameters. Therefore, many systems use a small baseline to limit the range of disparities [4] or low resolution images [3], leading to reduced detection range and precision. To solve this issue, multi-resolution is an efficient solution. In [1], Kimura *et al* propose to compute a disparity image directly from the lower resolution disparity image, using a pyramidal approach. This method is fast but can propagate errors to the upper resolution.

Once obstacles have been detected, they must be tracked over time to estimate their velocity. Most tracking methods propose to represent the detection results in the Euclidian space, and then to perform the whole tracking task in this coordinate system [8]. However, as reviewed in [9], that means working with an anisotropic and heteroscedastic measurement noise. To avoid dealing with this kind of noise, it is easier to work in the disparity space, in which measurement noise is linear and isotropic.

In this paper, we present a reliable framework for road obstacle detection designed to work in real time with high resolution images, eventually acquired with a large baseline. The presented approach is based on an original three-resolution scheme, in which each resolution has a specific function. The subsequent tracking of the detections is processed in the disparity space.

After a presentation of the geometry of our system (part 2), the detection algorithm will be described in more details (part 3). The tracking step is also presented (part 4). Finally, some experimental results of this algorithm are given (part 5). Computation time is specifically compared to classical approaches.

2. Geometrical Description

The geometrical configuration of our sensor is presented on Figure 1. We assume a perfectly rectified epipolar configuration.

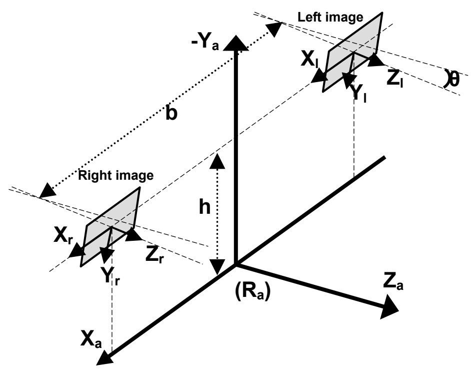

Figure 1. Geometrical configuration of the stereoscopic sensor.

Cameras are described by a pinhole model and characterized by α, their focal length measured in pixels. Given a point P(Xa, Ya, Za) in the absolute coordinate system Ra, its position in disparity space (u_r, Δ , v) can be calculated as:

$$\left\{\begin{array}{lcl} u_r & = & u_0 + \frac{\alpha X_a - \alpha b/2}{(Y_a+h)\sin\theta + Z_a\cos\theta} \\ v & = & \frac{[v_0\sin\theta + \alpha\cos\theta](Y_a+h) + [v_0\cos\theta - \alpha\sin\theta]Z_a}{(Y_a+h)\sin\theta + Z_a\cos\theta} \\ \Delta & = & u_l - u_r = \frac{\alpha b}{(Y_a+h)\sin\theta + Z_a\cos\theta} \end{array}\right. \quad (1)$$

3. The Three Resolution Approach

3.1. Overview

In classical multi-resolution approaches, disparity of a pixel is computed in the lowest resolution images. Then the disparities of the corresponding pixels at the next upper resolution are computed using this result, and are transmitted to the next upper resolution, and so on.

This hierarchical algorithm [7] is very efficient in term of computation time, but an error on one pixel at a low resolution induces errors on several pixels at the direct upper resolution. Therefore, it can lead to highly correlated errors, which are more disturbing than independent errors, especially when seeking for alignments like in the *v-disparity* obstacle detection algorithm [3]. Considering this, we propose a three stages algorithm, in which each step is associated to a given resolution, to benefit from a high reliability with low computation time. This architecture is presented on Figure 2.

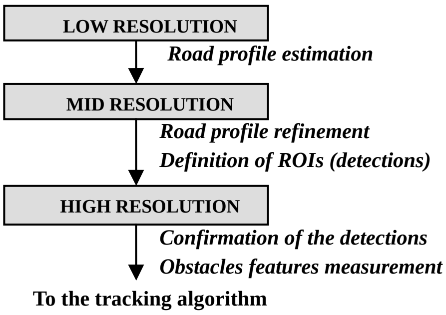

Figure 2. Overview of the detection algorithm.

The longitudinal road profile is estimated from the low resolution step. Then a denser disparity map is computed at middle resolution, to refine the longitudinal profile and define regions of interest (ROI). By finally computing a high resolution stereo matching in these specific ROIs, detections can be confirmed and more precisely located. Let us see in details the description of these three steps.

3.2. Low Resolution

The low resolution step consists in extracting the longitudinal road profile. First, a sparse disparity map is computed by correlation along scanlines, using the ZSSD criteria (Zero-mean Sum of Squared Differences). Then, a *v-disparity* image is built by projecting the disparity map along the lines, with accumulation [3]. In this representation, the road profile appears as straight line, and is estimated using Hough transform.

Thanks to the robustness of the *v-disparity* approach towards errors in the disparity map, this last one can be computed as fast as possible, regardless of its quality. Figure 3 presents the results of the low resolution step.

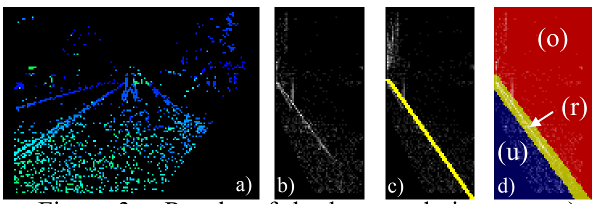

Figure 3. Results of the low resolution step: a) disparity map, b) *v-disparity* projection c) road profile d) the three resulting disparity ranges : "unreachable" (u), "road" (r) and "obstacle" (o).

3.3. Middle Resolution

A disparity map is also computed from the middle resolution images. However, thanks to the result of the low resolution stage, this can be done in a faster and more reliable way than with a classical correlation algorithm. Particularly, it provides a way to positively use the perspective distortion on the road surface.

Dealing with the perspective effect

Correlation based stereovision algorithms are founded on the assumption that all the objects in the observed scene are planar and parallel to the image planes. Road obstacles generally roughly comply with this hypothesis, but never does the road surface. This issue becomes very disturbing when using a large correlation window.

A solution to this problem could be the use of a 1D correlation window [5]. It could also be solved by applying a homographic transformation on one of the stereoscopic images to correct the road perspective as described in [6]. We rather propose to use a parallelogram shaped correlation window (figure 4).

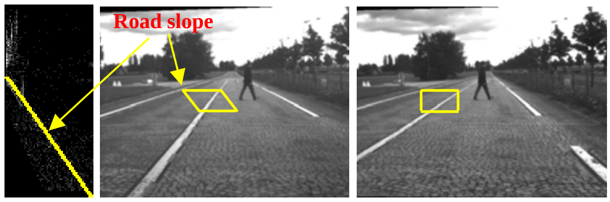

Figure 4. Principle of a sheared correlation window adapted to the road surface.

Such a method needs the estimation of the slope of the sheared window, which is related to the vehicle pitch and height. In our case, it is directly given by the slope of the plane road profile estimated from the low resolution *v-disparity* image.

Disparity map computation

Thanks to the results of the low resolution step, three ranges of disparities can now be defined for each image line, as presented on Figure 3-d. The first one (r) represents disparities which are close to the road surface. Pixels having this disparity can either belong to the road or to obstacles. The lower disparities (u) are simply impossible to reach as they are situated under the road surface. At last, the higher disparities (o) can only represent depth corresponding to obstacles. Considering this knowledge, we propose an efficient stereo matching algorithm:

For each pixel:

- the best correlation score (minimal for the ZSSD criteria) is computed for "road disparities" using the sheared window,
- the best correlation score is computed for "obstacle disparities" and "road disparities" using the rectangular window,
- the pixel is classified as "road" or "obstacle", considering that rectangular or sheared window providing the best correlation score,
- if it is an "obstacle pixels", it is reported on a middle resolution disparity map,
- else if it is a "road pixels", it is directly accumulated on a middle resolution *v-disparity* image.

At the end of this process, we obtain a disparity map containing only "obstacle pixels", which will be used for obstacles detection, and a *v-disparity* image used for road profile refining.

Obstacles detection

The resulting disparity image is used to find regions of interest , *i.e.* regions where there might be an obstacle. For this purpose, the Euclidian space is divided into voxels, whose size corresponds to the smallest detectable object. Each voxel is then projected into the disparity space, and the number of "obstacle pixels" inside is computed. Voxels containing a sufficient number of "obstacles pixels" are kept as small volumes of interest. Since there might be many volumes like this, the neighbor volumes are merged to build the final ROIs. Figure 6 shows the results of this middle resolution step of the algorithm.

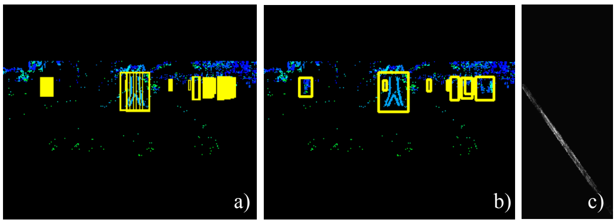

Figure 6. Results of the middle resolution step: a) "Obstacle pixels" disparity map with represented small ROIs, b) same image with merged ROIs, c) "road pixels" *v-disparity* image.

In this stage, disparity map computation and obstacle detection are parameterized to obtain an overabundance of detections, even if false detections appear. Using this technique, the detection rate is maximized. False detections will be removed in the high resolution stage.

3.4. High Resolution

Disparity map computation

The high resolution images are used to compute a local disparity map in each of the previously defined ROIs. Therefore, it is possible to benefit from high precision even with reasonable computation time. Indeed, ROIs are small against the images size, and very few disparity values are explored in each of them.

The disparity is computed by using the same algorithm as in middle resolution images, but with very strong requirements on the quality of matching (no ambiguities and no bad matching costs are accepted), so that only very reliable "obstacle pixels" appear on the image.

A bounding box is finally fitted round this "obstacle pixels".

Confirmation of the detection

To ensure to our system a maximum robustness against false positives, the local disparity map is also used to perform an "*a posteriori*" confirmation of the detections. This action is realized by using two of the confirmation algorithms presented in [10]:

- "number of obstacle pixels": ensures that the number of strong "obstacle pixels" inside the bounding box is high enough,
- "Prevailing alignment": checks that the projection of the "obstacle pixels" of a detected object in the *v-disparity* space forms an alignment which is roughly vertical. This method is designed to remove false positives on the road surface.

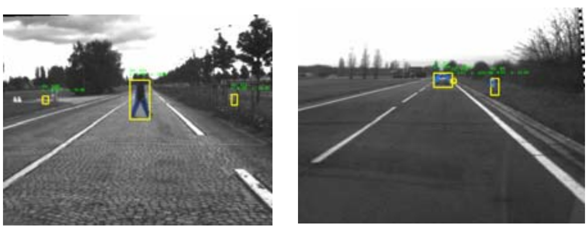

Figure 7. Results after the high resolution step.

4. Tracking the Detections

Once the objects of the scene have been detected using the three resolution algorithm, they are tracked to estimate their evolution over time.

4.1. Tracking algorithm

The tracking algorithm is founded on a very classical approach, using Kalman filtering. For each previously detected object (track):

- its state (position and speed) is predicted for the new frame,
- if possible, this prediction is associated with one of the detected object for this frame,
- the current state of the track is observed,
- its state is corrected thanks to the filter.

4.2. Design of the Filter

Each detected object is tracked by its own Extended Kalman Filter. We decided to represent its state in the vehicle Euclidian coordinate system (R_a), as:

$$X = [x, z, V_x, V_z]^t$$

Then the evolution is estimated through a linear model:

$$\begin{cases} x_{k+1} & = & x_k + V_{xk}.dt \\ z_{k+1} & = & z_k + V_{zk}.dt \\ V_{xk+1} & = & V_{xk} \\ V_{zk+1} & = & V_{zk} \end{cases} \quad (2)$$

As the measurement error in the images is directly related to the sampling process, it induces an anisotropic and heteroscedastic noise in R_a [9]. To solve this issue, the observations are given to the Kalman filter directly from the disparity space. Moreover, to ensure a maximum reliability to the estimation step, we chose to use both position and speed measurement. Finally, the observed variables are: $Z = [u_r, \Delta, v, V_{ur}, V_\Delta, V_v]^t$

Using equation (1), a non linear equation system can be defined to perform observation. This system is locally linearized by computation of its Jacobian matrix.

After the prediction step, the object state is given in R_a. Its predicted position in the image is computed using equations (1), with θ and h newly estimated from the *v-disparity* image.

4.3. Observation of the system

Measuring the position

The position of a detected object is simply measured by taking the image coordinates of the center of the lowest and nearest segment of its bounding box.

Measuring the speed

Measuring the relative speed of a track is achieved through a template matching strategy. This is performed by matching quickly and precisely the *u-disparity* projection of its "obstacle pixels" from successive frames.

As 2D matching techniques would be too expensive in term of computation time, we chose instead to perform two 1D correlations: we first determine lateral displacement with correlating the vertical projection histograms of the successive local *u-disparity* images. Vertical translation between the frames is found by correlation of the horizontal projection histograms of the successive local *u-disparity* images.

5. Experimental Evaluation

The performances of our algorithm have been evaluated on our experimental vehicle. The stereoscopic sensor is composed of two VGA video cameras. The baseline is 1.03 *m*. 255 disparity values are explored, so that a perception range from 3.5 *m* to about 100 *m* can be covered. 7x7 correlation windows are used.

5.1. Precision

Precision and detection range are directly related to our sensor features. By comparing measurements with lidar data, the experimental values appear coherent with the attended values:

- obstacles are detected up to 95 meters.
- precision of detection is about 5 *cm* at 6 *m* and 2.7 *m* at 50 *m*.

5.2. Reliability

Our three resolution detection strategy shows good results in terms of reliability. The overabundance of detection in middle resolution permits to obtain a correct detection rate. Most obstacles (vehicles, pedestrians and boxes) have been correctly detected and tracked.

Thanks to the confirmation stage used during the high resolution step, the false detection rate remains low: 3 false detections for 4763 frames processed.

Some weakness remains on the stability of the bounding boxes.

In the next future, the effectiveness of the method will be validated on a large set of images.

5.3. Computation time

The computation time of the various steps of the algorithm has been measured on a 2.4GHz Pentium 4 computer. These values are the average times measured among a set of 700 images, including various number and size of obstacles. As a reference, times are compared to the computation time for a complete VGA disparity map using two classical methods:

- Low resolution step: 14.3 *ms*
- Middle resolution step: 94.6 *ms*
- High resolution step: 77.9 *ms*
- Tracking step: 6.7 *ms*
- **Total computation time: 193.5** *ms*
- Complete VGA disparity map: **492.4** *ms*
- Complete VGA disparity map, using hierarchical approach : **176.0** *ms.*

As we can see from these results, computation time of our complete detection algorithm is very low compared to the time needed for a full resolution disparity map. Moreover, it is quite the same than the computation time of a disparity map using multiresolution hierarchical approach. Even so, it includes the whole detection stage, providing reliable results.

6. Conclusion and Outlook

We presented in this paper a complete algorithm for road obstacle detection and tracking.

Thanks to the overdetection / confirmation strategy and to the sheared window, the three-stage detection algorithm provides flexibility and reliability. The implementation of this technique would be difficult and uncertain with a single high resolution disparity map, because it would need compromises between density and correctness of this map.

By combining this algorithm with a three resolution approach, the system is designed for real time operation, even with high resolution images and a large disparity range.

In parallel, the tracking algorithm solves some issues related to the non linearity of the image projection transform.

Now, the complete algorithm will be more intensively evaluated on large data sets. Then, it will be implemented on a specific hardware to run at high framerate on VGA images.

Acknowledgments The presented work has been realized as a part of D.O.30, a French research project dealing with obstacle detection using stereovision.

References:

[1] Y. Kimura, T. Kato, M. Ohta, Y. Ninomiya, Y. Takagi, M. Usami and S. Tokoro. “Stereovision for Obstacle Detection”. In *Proc. of the 13th ITS World Congress*, London, UK, 2006.

[2] T. Kanade, H. Kano, S. Kimura, A. Yoshida and K. Oda. “Development of a Video Rate Stereo Machine”. In *Proc. of IEEE IROS Conference*, Pittsburg, USA, 1995.

[3] R. Labayrade, D. Aubert, and J.P. Tarel. “Real time obstacle detection on non flat road geometry through ‘v-disparity’ representation”. In *Proc. of IEEE IV Symposium*, Versailles, France, 2002.

[4] K. Konolige. “Small vision system: Hardware and implementation”, In *Proc. of 8th ISRR Symposium*, 1997.

[5] S. Lefebvre, S. Ambellouis and F. Cabestaing. “Obstacles Detection on a Road by Dense Stereovision with 1D Correlation Windows and Fuzzy Filtering”. In *Proc. of the IEEE ITS Conference*, Toronto, Canada, 2006.

[6] T. Williamson. “A High-Performance Stereo Vision System for Obstacle Detection”. *PhD thesis, Carnegie Mellon University*, USA, 1998.

[7] C. Sun. “A Fast Stereo Matching Method”. In *Digital Image Computing: Techniques and Applications*, Massey University, Auckland, New Zealand, December 1997.

[8] S. Nedeveschi and R. Schmidt. “High Accuracy Stereo Vision System for Far Distance Obstacle Detection”. In *Proc. of IEEE IV Symposium*, Parma, Italy, 2005.

[9] G. Deparnis and P. Chang. “Closed-form Linear Solution to Motion Estimation in Disparity Space”. In *Proc. of IEEE IV Symposium*, Tokyo, Japan, 2006.

[10] M. Perrollaz, R. Labayrade, C. Royère, N. Hautière and D. Aubert. “Long Range Obstacle Detection Using Laser Scanner and Stereovision”. In *Proc. of IEEE IV Symposium*, Tokyo, Japan, 2006