Road Approximation in Euclidean and v-Disparity Space: A Comparative Study*

Angel D. Sappa[1], Rosa Herrero[1], Fadi Dornaika[2], David Gerónimo[1], and Antonio López[1]

[1] Computer Vision Center, Edifici O Campus UAB
08193 Bellaterra, Barcelona, Spain
{asappa, rherrero, dgeronimo, antonio}@cvc.uab.es
[2] Institut Géographique National
94165 Saint Mandé, France
fadi.dornaika@ign.fr

Abstract. This paper presents a comparative study between two road approximation techniques—planar surfaces—from stereo vision data. The first approach is carried out in the v-disparity space and is based on a voting scheme, the Hough transform. The second one consists in computing the best fitting plane for the whole 3D road data points, directly in the Euclidean space, by using least squares fitting. The comparative study is initially performed over a set of different synthetic surfaces (e.g., plane, quadratic surface, cubic surface) digitized by a virtual stereo head; then real data obtained with a commercial stereo head are used. The comparative study is intended to be used as a criterion for fining the best technique according to the road geometry. Additionally, it highlights common problems driven from a wrong assumption about the scene's prior knowledge.

1 Introduction

Recently, several techniques relaying on stereo vision systems have been proposed in the literature for driver assistance applications. These techniques have to deal with classical 3D processing and modelling problems together with real time constraints. The latter have motivated the use of driver scene's prior knowledge in order to simplify the problem.

A common problem in every on-board vision system (monocular/stereo) is the real time estimation of position and orientation, related to the current 3D road plane parameters—the ego-motion problem (e.g., [1], [2], [3]). Note that since the 3D plane parameters are expressed in the camera coordinate system, the camera position and orientation are equivalent to the plane parameters. Different algorithms have been proposed in the literature for road approximation from

* This work has been partially supported by the Spanish Ministry of Education and Science under project TRA2004-06702/AUT. The first author was supported by The Ramón y Cajal Program. The fourth author was supported by Spanish Ministry of Education and Science grant BES-2005-8864.

R. Moreno-Díaz et al. (Eds.): EUROCAST 2007, LNCS 4739, pp. 1105–1112, 2007.

stereo vision data. These approaches can be broadly classified into two categories depending on the space where the processing is performed: *v-disparity* space and *Euclidean* space.

v-disparity space based approaches, originally proposed by [4], obtain a road approximation by using the scene prior knowledge and avoiding 3D data point computation. Disparity values correspond to the differences between the right and left images, while the corresponding v-disparity representation is computed by accumulating the points with the same disparity value that occur on a given image line. The v-disparity space allows to represent in a compact way the current scene geometry. An appealing feature of v-disparity representations lie in the fact that a plane in the Euclidean space becomes a straight line in the v-disparity space. Therefore, plane fitting becomes segment fitting. Actually, instead of segment fitting the use of Hough transform has been proposed in [4]; since then also adopted by those techniques working in the v-disparity space. The Hough transform is used with a voting scheme to extract the largest sub-set of points that define a straight line; unlike fitting techniques that find the best straight line for the whole set of points (e.g., least squares fitting).

After the original proposal [4], several v-disparity based approaches have been developed for driver assistance: obstacle or pedestrian detection (e.g., [5], [6], [7]), atmospheric visibility measurement system [8], etc. Recently, the v-disparity approach has been extended to a u-v-disparity concept in [9]. In this new proposal, dense disparity maps are used instead of only relying on edge based disparity maps. Working in the disparity space is an interesting idea that is gaining popularity in on-board stereo vision applications, since planes in the original Euclidean space become straight lines in the disparity space. However, it should be noticed that this approach has been proposed under the assumption that the road geometry fit to a plane. If this assumption does not hold, the straight line extracted in the v-disparity space wouldn't correspond to the best fitted plane in the Euclidean space.

In turn, Euclidean space based approaches are focused on the use of 3D data points directly in the 3D space—dense or sparse representations. For instance, [10] proposes a road approximation technique that works in the Euclidean space, by using a method similar to the Hough transform over a lateral projection of the original 3D points. On the contrary, a least squares fitting approach is used in [11], by previously removing outliers. Hence, a real time estimation of on-board camera extrinsic parameters, related to the plane that minimizes the sum of the squares of the offsets ("the residuals") of the whole set of points, is obtained. In this case, road plane fitting is performed in the 3D Euclidean space by using a compact and structured representation of the raw input data points.

The current paper aims at comparing a v-disparity space based approach [4] with an Euclidean space based approach [11], without going into details on their corresponding implementations; the main objective is to study their validity, while at the same time conclusions are obtained.

The remainder of this paper is organized as follows. Section 2 briefly describes the v-disparity space based road approximation technique [4]. Then, section 3

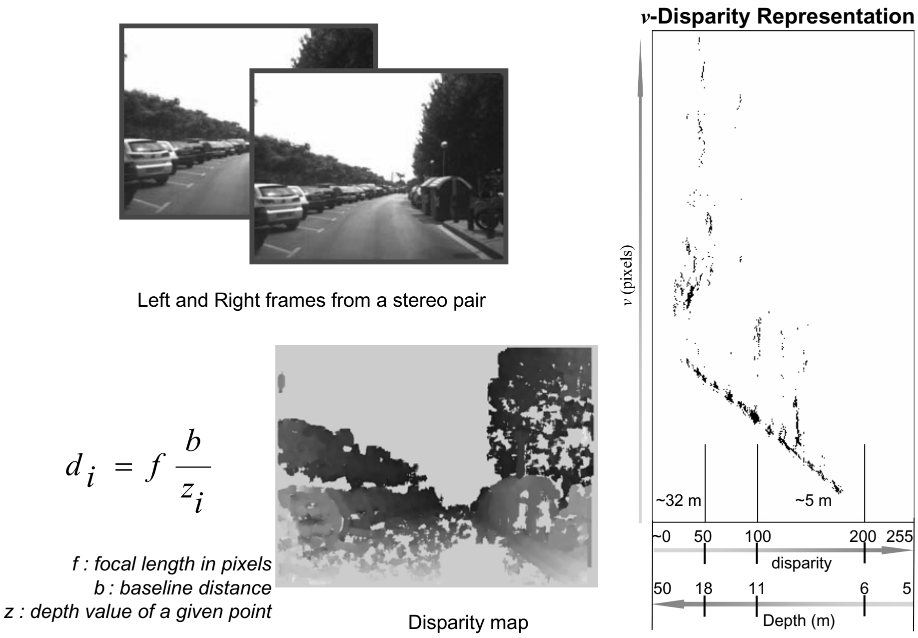


Fig. 1. (*left*) Disparity map from the left and right images of a given stereo head. (*right*) The corresponding v-disparity representation.

introduces the road approximation technique in Euclidean space [11]. Section 4 presents comparisons obtained with synthetic and real scenes. Finally, discussions and conclusions from the comparative study are given in section 5.

2 Road Approximation in v-Disparity Space

As mentioned above several approaches have been proposed based on the fact that a plane in the Euclidean space becomes a straight line in the v-disparity space. Up to our knowledge, all these approaches are based on the use of Hough transform for detecting a straight line, which is assumed to be the road plane projection in the v-disparity representation. Figure 1(*right*) shows the v-disparity representation corresponding to the disparity map presented in Figure 1(*left*).

Hough Transform (HT) is generally used in image processing for finding shapes in images or set of data points. The underlying principle of HT is that there is an infinite number of potential lines that pass through any point, each at a different orientation. Hence if all possible straight lines, for every data point in the image, are considered at once and represented in a (r, θ) space, which correspond to the angle and distance to the origin of a normal to the line in question. The position with most crosses at the (r, θ) space will define the straight line with most points. This voting scheme has been largely used for feature detection as well as for solving regression problems. For instance, HT has been used by [4],

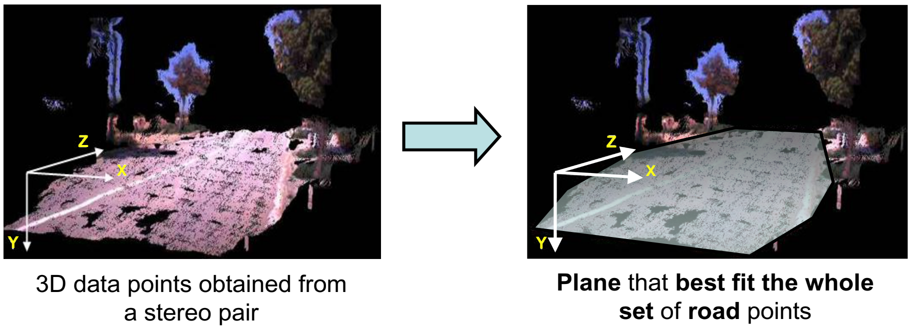


Fig. 2. Road approximation in Euclidean space

[5], [6] and [7] to find the the largest sub-set of points that define a straight line in the v-disparity space. This straight line corresponds to the 3D road plane in the Euclidean space.

3 Road Approximation in Euclidean Space

Approaches proposed to work in Euclidean space essentially consist of two stages. Initially, 3D data points are structured in such a way that a real time processing could be performed. Then, a fitting technique is used for finding the best surface that approximate those data points [12]. The least squares fitting technique is the simplest and most commonly applied procedure for finding the best-fitting surface to a given set of points. It minimizes the sum of the squares of the offsets of the points from the surface. The sum of squares of the offsets is used instead of the offset absolute values because this allows the residuals to be treated as a continuous differentiable quantity.

In [11], the first stage consists in generating a compact 2D representation of the original 3D data points—lateral view of 3D data points. Then a RANSAC based least squares approach is used for fitting a plane to the road candidate points selected in the 2D projection (YZ plane). Figure 2 presents a sketch of road approximation in Euclidean space.

4 Comparisons

The two approaches introduced above have been compared with different road geometries. Initially, synthetic data points were obtained by using a virtual stereo head. Then, data points from real scenes were considered by using a commercial stereo head.

4.1 Synthetic Data

A virtual stereo head has been defined by using two virtual pin-hole cameras. A similar configuration to the one provided by the commercial stereo head used

Table 1. Comparisons: **Synthetic Scenes**

SURFACE	v-DISPARITY SPACE [4] {Mean Sq./Max.} Error [m]	EUCLIDEAN SPACE [11] {Mean Sq./Max.} Error [m]
Plane	0 / 0	0 / 0
(with noise)	0.007 / 0.008	0.007 / 0.008
Quadratic (35 m at 1000m)	2.68 / 34.49	0.44 / 8.47
(with noise)	2.68 / 36.69	0.44 / 8.93
Quadratic (100 m at 1000m)	24.14 /103.48	3.98 / 25.41
(with noise)	24.12 / 110.29	3.98 / 26.99
Cubic (35 m at 1000m)	2.89 / 36.98	0.80 / 11.79
(with noise)	2.89 / 39.44	0.80 / 12.52
Cubic (100 m at 1000m)	20.61 / 98.63	5.69 / 31.45
(with noise)	20.60 / 105.38	5.69 / 33.49

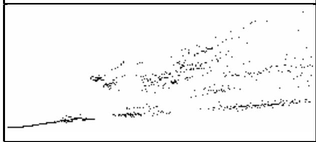
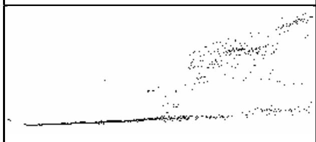
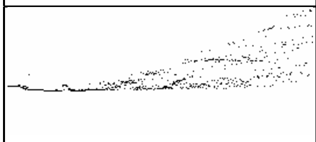
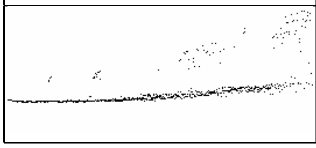
in the real data experiments has been adopted (see 4.2). Different surfaces have been considered and their corresponding images (left/right) have been used for computing disparity values. Disparity values were used for obtaining the corresponding v-disparity representation. For every pixel of the synthetic images 3D information was directly computed by finding the intersection between the ray that passes through that pixel (image plane) and the considered surface. Every surface was used twice; firstly, with the synthetic data computed as indicated above and secondly, by adding noise to the surface. The noise grows according to the depth trying to model accuracy of stereo vision systems—inversely proportional to depth (e.g., [13], [14]).

Following [4], the HT was used for obtaining road plane parameters from the straight line extracted in the v-disparity space. On the other hand the corresponding 3D data points were directly used for computing road plane parameters in the Euclidean space [11]. In both cases, the mean square error and maximum error between the computed planes and the set of 3D data points were computed and used as comparison criteria between the two techniques. Table 1 presents results obtained with different surfaces. As it was expected, both approaches give the same results when planar surfaces are considered (with/without added noise). On the contrary, smaller mean square errors and maximum errors are obtained by using a fitting scheme in the Euclidean space [11] instead of using HT in v-disparity space [4], when non-planar surfaces are considered.

4.2 Real Data

A commercial stereo vision system (`[www.ptgrey.com]` Bumblebee from Point Grey) was used. It consists of two Sony ICX084 Bayer pattern CCDs with 6mm focal length lenses. Bumblebee is a pre-calibrated system that does not require in-field calibration. The baseline of the stereo head is 12 cm and it is connected

Table 2. Comparisons: **Real Scenes**

Road Profile in Euclidean Space	Plane Fitting in v-**DISPARITY SPACE** [4] {Mean Sq./Max.} Error [m]	Plane Fitting in **EUCLIDEAN SPACE** [11] {Mean Sq./Max.} Error [m]
	0.94 / 8.51	0.66 / 5.35
	3.34 / 13.88	1.74 / 8.22
	2.58 / 16.58	0.96 / 11.27
	0.87 / 11.22	0.60 / 9.08

to the computer by a IEEE-1394 connector. Right and left color images were captured at a resolution of 640×480 pixels. Camera control parameters were set to automatic mode to compensate global changes in the light intensity. For every couple of right/left images the corresponding disparity image and 3D data were computed by using the provided 3D reconstruction software. Disparity values were used for computing the v-disparity representation. It is later on used by [4] to obtain the road plane parameters. On the other hand, 3D data points were used for fitting the road plane as indicated in [11]. As in the synthetic case, mean square errors and maximum errors were computed for comparing the obtained results. Table 2 presents some of the results obtained with different road geometries. It can be appreciated that in all the cases working in the Euclidean space, by using a fitting technique, gives better results than those obtained in v-disparity space, by using HT.

5 Discussions and Conclusions

This paper presents a comparative study between two different road approximation techniques generally used on stereo based driver assistance applications.

Although both approaches are valid for road plane extraction, working in the v-Disparity space could drive to wrong results do to the use of Hough transform with non-planar road geometries.

Although voting and fitting schemes would give the same result when a perfectly planar road is considered (ideal case) differences will appear when non planar roads are processed. It is easy to see that the segment that passes through more points will be obtained by using the HT, which does not necessarily correspond to the most representative one for the whole set of points.

The non-linear representation of disparity values (Figure 1($left$)) rises up as an additional drawback when v-disparity space is considered. Figure 1($right$) presents the v-disparity representation of a scene ranging between 5 and 50 meters in depth. Notice that less than a quarter of disparity values (from 0 up to 50, having a total span higher than 200 values) are used for representing more than 70% of depth values (distances from 18 up to 50 meters). This non-linear mapping makes that more attention is paid to nearest points, instead of considering all the points equally (almost half of disparity values, from 100 to 200 are used for representing distances in between 6 and 11 meters, about 11% of depth values). Recently, [15] has also noticed this drawback and proposed a measurement to estimate the quality of the v-disparity image according to road flatness. This image quality value is used for computing the on-board camera pitch orientation.

Although out of the scope of this comparative study, it should be mentioned that the extraction of planar representations with Hough transform in the v-Disparity space is faster than fitting in Euclidean space, since most of the CPU time required by 3D reconstruction algorithms is avoided.

We can conclude that [4] can be used in highways environments where road vertical curvature could be neglected. However, if the on-board stereo system is intended to be used on urban driver assistance applications [4] could drive to wrong results, particularly those urban scenarios in non-flat regions. In this case, a technique such as the one presented in [11] is better suited.

References

1. Bertozzi, M., Broggi, A.: GOLD: A parallel real-time stereo vision system for generic obstacle and lane detection. IEEE Trans. on Image Processing 7(1), 62–81 (1998)
2. Milella, A., Siegwart, R.: Stereo-based ego-motion estimation using pixel tracking and iterative closest point. In: Proc. IEEE Int. Conf. on Computer Vision Systems, IEEE Computer Society Press, New York, USA (2006)
3. van der Mark, W., Fontijne, D., Dorst, L., Groen, F.: Vehicle ego-motion estimation with geometric algebra. In: Proc. IEEE Intelligent Vehicles Symposium, Versailles, France, pp. 58–63. IEEE Computer Society Press, Los Alamitos (2002)
4. Labayrade, R., Aubert, D., Tarel, J.: Real time obstacle detection in stereovision on non flat road geometry through 'V-disparity' representation. In: Proc. IEEE Intelligent Vehicles Symposium, Versailles, France, pp. 646–651. IEEE Computer Society Press, Los Alamitos (2002)

5. Bertozzi, M., Binelli, E., Broggi, A., Del Rose, M.: Stereo vision-based approaches for pedestrian detection. In: Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition, IEEE Computer Society Press, San Diego, USA (2005)
6. Broggi, A., Fascioli, A., Fedriga, I., Tibaldi, A., Del Rose, M.: Stereo-based preprocessing for human shape localization in unstructured environments. In: Proc. IEEE Intelligent Vehicles Symposium, pp. 410–415. IEEE Computer Society Press, Columbus, OH, USA (2003)
7. Labayrade, R., Aubert, D.: A single framework for vehicle roll, pitch, yaw estimation and obstacles detection by stereovision. In: Proc. IEEE Intelligent Vehicles Symposium, pp. 31–36. IEEE Computer Society Press, Columbus, OH, USA (2003)
8. Hautière, N., Labayrade, R., Aubert, D.: Real-time disparity contrast combination for onboard estimation of the visibility distance. IEEE Trans. on Intelligent Transportation Systems 7(2), 201–212 (2006)
9. Hu, Z., Uchimura, K.: U-V-Disparity: An efficient algorithm for stereovision based scene analysis. In: Proc. IEEE Intelligent Vehicles Symposium, pp. 48–54. IEEE Computer Society Press, Las Vegas, USA (2005)
10. Nedevschi, S., Danescu, R., Frentiu, D., Graf, T., Schmidt, R.: High accuracy stereovision approach for obstacle detection on non-planar roads. In: Proc. IEEE Inteligent Engineering Systems, Cluj Napoca, Romania, pp. 211–216. IEEE Computer Society Press, Los Alamitos (2004)
11. Sappa, A., Gerónimo, D., Dornaika, F., López, A.: On-board camera extrinsic parameter estimation. Electronics Letters 42(13), 745–747 (2006)
12. Rousseeuw, P., Leroy, A.: Robust Regression and Outlier Detection. John Wiley & Sons, New York (1987)
13. Toulminet, G., Bertozzi, M., Mousset, S., Bensrhair, A., Broggi, A.: Vehicle detection by means of stereo vision-based obstacles features extraction and monocular pattern analysis. IEEE Trans. on Image Processing 15(8), 2364–2375 (2005)
14. Scherba, D., Bajcsy, P.: Depth map calibration by stereo and wireless sensor network fusion. In: Proc. IEEE Int. Conf. on Information Fusion, Philadelphia, USA, pp. 1540–1547. IEEE Computer Society Press, Los Alamitos (2005)
15. Broggi, A., Caraffi, C., Porta, P., Zani, P.: The single frame stereo vision system for reliable obstacle detection used during the 2005 DARPA Grand Challenge on TerraMax. In: Proc. IEEE Int. Conf. on Intelligent Transportation Systems, Toronto, Canada, pp. 745–752. IEEE Computer Society Press, Los Alamitos (2006)