

Ground Plane Estimation Based on Dense Stereo Disparity

Nikolay Chumerin, Marc M. Van Hulle

Katholieke Universiteit Leuven, Laboratorium voor Neuro- en Psychofysiologie
 Herestraat 49, B-3000 Leuven, Belgium

E-mail: {nikolay.chumerin, marc.vanhulle}@med.kuleuven.be
<http://simone.neuro.kuleuven.be/>

Abstract – In the paper we present an approach to the ground plane parameters estimation and tracking by the moving observer. The method is based entirely on stereo disparity information and assumes only presence of the textured dominant planar surface in the lower part of the screen, which is not severe restriction in real-world scenarios. The main idea of the proposed approach consists of disparity plane estimation and its conversion into the ground plane. Proposed method does not use a-priori knowledge of the scene appearance (e.g. lane markings, horizon line, presence of the vertical planes) and thus can be applied in the wide range of scenarios.

Keywords – computer vision, structure from stereo, ground plane estimation

I. INTRODUCTION

In many computer vision problems which are related to analysis of the video obtained by moving observer (e.g. obstacle detection, road detection, ego-motion estimation, vehicle/mobile robot navigation, sensors data fusion, etc.) road modeling in 3D is considered as preliminary but very important step. The assumption of the road planarity is commonly accepted in many approaches. Nevertheless some methods impose even more simplifying restrictions on the ground plane: strictly horizontal orientation and even static position with respect to the observer [1]. Vision-based methods rely mostly on the processing of different features attached to the ground plane: texture (lane markings) [2], v -disparity [3], motion (optical flow) [4], presence of vertically oriented planar surfaces (building walls in urban environment) [5], etc. In the paper we describe in details a stereo disparity-based method of the ground plane detection, which is a part of an independently moving objects detection system [6].

The paper is structured as follows: in Section II we describe stereo disparity and its role in 3D structure reconstruction. In the same section we show how to convert disparity plane into ground plane and vice versa. Section III introduces two slightly different methods of the ground plane estimation using estimated disparity plane. In Section IV we propose a simple ground plane tracking procedure. Some experimental results are shown in Section V, where we also show one of the applications of the proposed method. In last Section VI we conclude the

study, present a simple stability comparison of the two approaches presented in Section III and discuss directions of the future work.

II. DISPARITY PLANE ESTIMATION

A. Disparity as a depth cue

In a biological vision system the separation of the eyes causes each eye to see a disparate image of the world. These positional disparities are sufficient to infer depth information from a stereo-pair of the images. Therefore stereo disparity plays role of an important depth cue in the biological visual systems and can easily be adopted by machine vision systems.

We assume a calibrated stereo camera system (see Fig. 1) with baseline L , where the Z axis of the world coordinate system is aligned with camera optical axis, and X and Y axes are aligned with image axes x and y . The focal length f of the cameras and baseline length L are assumed to be known. By p_l and p_r we denote perspective projections of P onto left and right camera image planes.

With this configuration the *disparity* δ of the point P in a 3D space, is defined as the shift between horizontal positions of the points p_l and p_r , which may be expressed as:

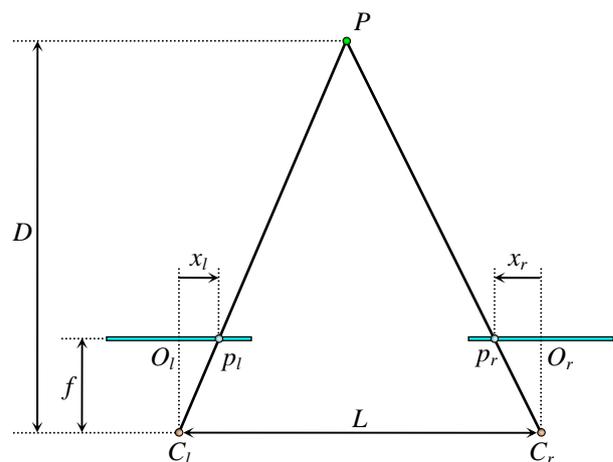


Fig. 1. The geometry of the stereo camera setup.

$$\delta = x_l - x_r. \quad (1)$$

The depth D of the point P can be computed as:

$$D = fL/\delta. \quad (2)$$

We will call *disparity map* the set $\mathcal{D} = \{(x_i, y_i, \delta_i)\}_i$, where δ_i is disparity of the image pixel (x_i, y_i) .

Disparity map estimation by itself is a very challenging problem, but in this study we are focused on one of the possible applications of the stereo disparity. Among different disparity algorithms we have chosen a phase-based approach [7] and use its multi-scale extension [8] to compute dense disparity maps (see Fig. 2b).

B. Connection between ground plane and disparity plane

Consider a plane Π , which in 3D world coordinate system (attached to the left camera) can be defined by:

$$\Pi: aX + bY + cZ + d = 0, \quad (3)$$

Without loss of generality we assume that $a^2 + b^2 + c^2 = 1$ (otherwise one can divide coefficients a, b, c by $\sqrt{a^2 + b^2 + c^2}$). In this case vector $\mathbf{n} = (a, b, c)^\top$ represents the normal unity vector of the plane Π and coefficient d represents the distance from the camera nodal point to the plane*. Using simple algebraic manipulations it is easy to show that corresponding disparity model Δ of the plane Π in image coordinate system is linear:

$$\Delta: \delta = \alpha x + \beta y + \gamma, \quad (4)$$

where x, y are pixel coordinates in the frame coordinate system, δ is disparity of the pixel (x, y) and coefficients α, β and γ are defined by:

$$\begin{cases} \alpha = -aL/d, \\ \beta = -bL/d, \\ \gamma = -cfL/d. \end{cases} \quad (5)$$

Inverse mapping from disparity domain to 3D world domain is also possible:

$$\begin{cases} a = -\alpha/\sqrt{\alpha^2 + \beta^2 + (\gamma/f)^2}, \\ b = -\beta/\sqrt{\alpha^2 + \beta^2 + (\gamma/f)^2}, \\ c = -\gamma/(f\sqrt{\alpha^2 + \beta^2 + (\gamma/f)^2}), \\ d = -L/\sqrt{\alpha^2 + \beta^2 + (\gamma/f)^2}. \end{cases} \quad (6)$$

* To avoid ground plane of being fronto-parallel we assume that $b > 0$. In this case vector \mathbf{n} points to upper subspace (with respect to ground plane and taking into account orientation of the coordinate system).

We have investigated two approaches to the disparity based ground plane estimation:

- 1) Reconstruct 3D structure of the scene and then fit the ground plane directly into reconstructed data.
- 2) Fit disparity plane and then use it to reconstruct ground plane;

The first method seems to be more logical, but in practice it renders to be less stable than the second. We explain this by the fact that during 3D reconstruction according to (2), disparity appears in denominator and thus even small disparity noise can cause severe depth fluctuations.

C. Robust plane fitting

At this stage we assume that we have already estimated dense disparity map \mathcal{D} and the current task is to fit linear model (4) into these data. Direct application of the classical linear regression methods (*e.g.* least squares) is useless because their basic assumptions are not met: we do not know anything about underlying noise distribution (which expected to be Gaussian and have zero mean) and there are a lot of *outliers* in input data. One of the possible solutions in this situation involves *robust regression*. Robust regression extends classical regression methods in such a way that they become less sensitive to outliers or other small deviations from the model assumptions. Among popular methods of robust regression (such as *Iteratively Reweighted Least-Squares* (IRLS) [9], *Least Median of Squares Regression* (LMedS) [10], *Random Sample Consensus* (RANSAC) [11]) we have chosen IRLS due to its speed and low computational complexity. In all our simulations for IRLS we have used the weight function proposed by Beaton and Tukey [12] with the tuning parameter $c=4.6851$ and number of iterations fixed to 7.

As preprocessing step before robust fitting we intersect the disparity map (Fig. 2b) with heuristically designed predefined road mask (Fig. 2c). By this step, we filter out the majority of the pixels which belong to sky and objects above the ground plane and are outliers in the disparity plane model (4). Then desired disparity plane parameters α, β , and γ are estimated by IRLS using this reduced disparity map (see Fig. 2d)

III. GROUND PLANE ESTIMATION

D. Direct method

As soon as disparity plane is estimated one can directly estimate parameters of the ground plane using (6). It is a generic approach which does not employ any extra information about the possible ground plane position with respect to the observer. Theoretically (*e.g.* in the case when the ground plane is not a dominant planar structure in the scene), direct method could give completely wrong estimates, but fortunately these situations are rather exotic and the direct method performs relatively well (see Fig. 4).

E. Stabilized method

In order to reduce magnitude of undesired noise in estimates we have developed a stabilized algorithm for the same problem. First we choose a fixed set of nine points (3×3 lattice) in the lower half of the frame (see Fig. 3, red bullets). Disparities for these points are computed in each frame using the estimated disparity plane model (4). Given the disparities and camera calibration data, we project the selected points into a 3D world coordinate system. In addition, we add two so-called *stabilization points* which correspond to the points where the front wheels of the test car are supposed to touch the road surface. For the inverse projection of the stabilization points, we use parameters of the *canonical disparity plane* (it is a disparity plane which corresponds to the horizontal ground plane observed by cameras in a quiescent state, corresponded ground plane we refer as *canonical ground plane*). The parameters of the canonical disparity plane and positions of the stabilization points can be obtained basing on geometry of the camera setup. The full set of 11 points is then used for IRLS fitting of the ground plane (3) in a world coordinate system.

During the disparity plane estimation, we use the estimation from the previous frame for weight initialization in IRLS; for the first frame, for the same purpose, we use the parameters of the canonical disparity plane.

IV. GROUND PLANE TRACKING

We assume that the ground plane is estimated correctly if the following conditions are met:

$$\begin{cases} \|\mathbf{n}_t - \mathbf{n}_0\| < \theta_0, \\ \|\mathbf{n}_t - \mathbf{n}_{t-1}\| < \theta_1, \\ |d_t - d_0|/d_0 < \theta_2, \end{cases} \quad (7)$$

where \mathbf{n}_t and d_t are normal vector and d coefficient of the ground plane for t^{th} frame, \mathbf{n}_0 and d_0 are normal vector of and d coefficient the canonical ground plane.

Thresholds $\theta_0=0.075$, $\theta_1=0.015$ and $\theta_2=0.15$ were chosen empirically. If the estimated ground plane does not satisfy (7), the estimate for the previous frame is used.

V. RESULTS

Some results of the ground plane estimation are shown on Fig. 3. The method has shown not only perfect results on synthetic data, but also robust results in real-world

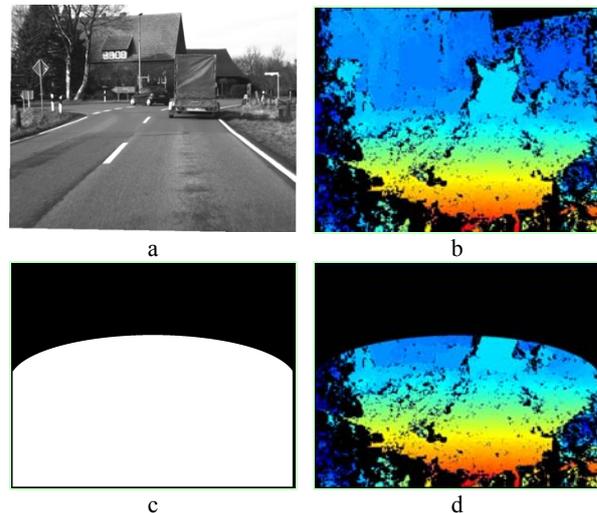


Fig. 2. a) Original (left) image, b) stereo disparity map (color indicates magnitude of the disparity: red – large, blue – small), c) predefined road mask, d) intersection of the road mask with the disparity map.

scenarios. Particularly on Fig. 3a we present results for synthetic sequence (ground truth disparity maps were provided by a ray-tracing engine), on Fig. 3b a typical rural scenario is presented, on Fig. 3c – a typical urban scenario and on Fig. 3d – a cluttered urban scenario with high amount of outliers.

Among the direct applications of the disparity plane we want to show one simple disparity-based approach to obstacle detection. Every pixel which belongs to objects above ground plane should be classified as an obstacle pixel. In terms of disparity this condition can be expressed as:

$$\delta(x, y) > \alpha x + \beta y + \gamma + \delta_0, \quad (8)$$

where $\delta(x, y)$ is the disparity value from the disparity map \mathcal{D} of the pixel (x, y) , parameters α, β and γ are estimates of the model (4), and δ_0 is a threshold which adjusts minimal elevation above the disparity plane. Some results of this classification are presented on Fig. 5.

VI. CONCLUSION

The proposed method requires only a disparity map and thus rely only on one assumption about the presence of the textured (not homogeneous) planar structure in the lower part of the scene. This condition is not a severe restriction and usually is met in real-world scenarios.

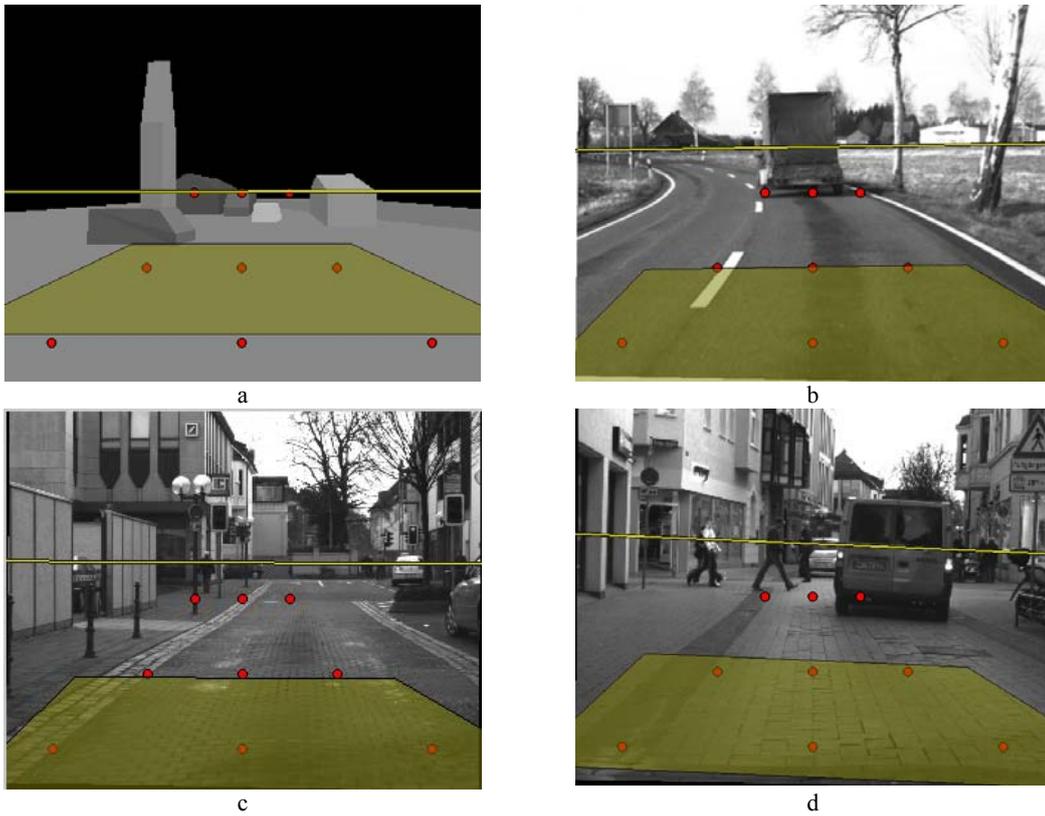


Fig. 3. Results of the ground plane estimation. Red bullets depict points which are used for 3D reconstruction and fitting the ground plane. Yellow patch in the lower part of the screen shows ground plane position for the current frame. Yellow horizontal line shows position of the horizon line (see text).

The proposed method does not depend on *a-priori* knowledge of the scene appearance (*e.g.* lane markings, horizon line, presence of the vertical planes) and thus can be applied in the wide range of scenarios.

Even though there is an iterative fitting procedure during estimation, the method remains relatively fast and can be easily optimized for real-time purposes.

As a drawback we should mention that the proposed

approach rely heavily on the density and the accuracy of the disparity map estimation. Nevertheless, multiscale phase-based disparity algorithm [8], in most cases produces reliable enough data for the robust ground plane estimation.

On Fig. 4 we present two plots on which basis one can judge about stability of the two methods of estimation ground plane from disparity plane proposed in Sec-

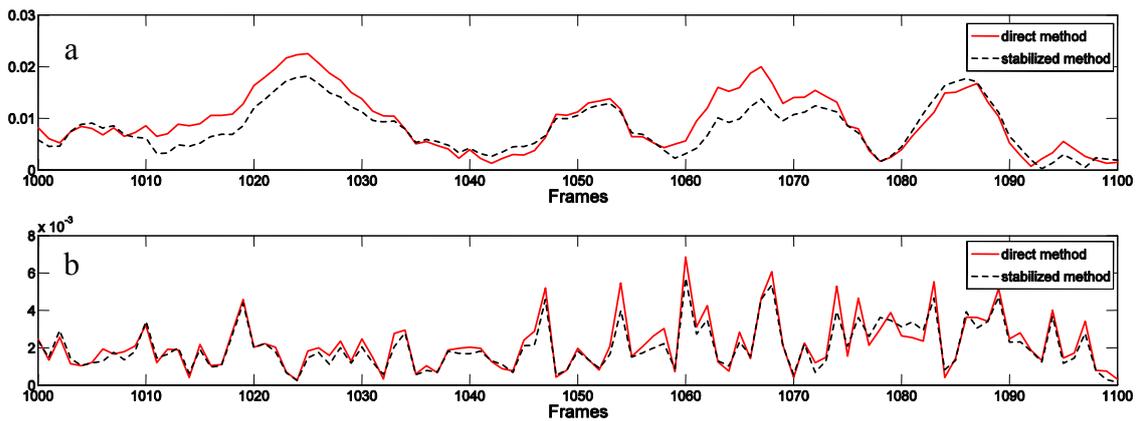


Fig. 4. Comparison of the direct and stabilized methods by means of: a) deviation of the estimated ground plane normal vector from canonical ground plane normal vector; b) deviation of the estimated ground plane normal vector in current frame from the same vector estimated in the previous frame (see text).

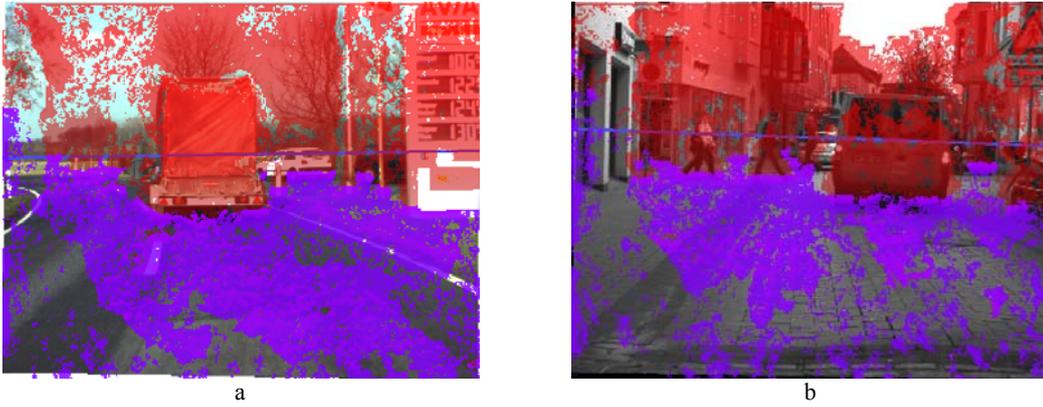


Fig. 5. Obstacle detection based on estimated disparity plane. By red color we depict obstacle pixels and using magenta we show pixels near disparity plane.

tion III. Unfortunately, there is no ground truth for the real-world sequences, so we are forced to use indirect stability measures: the norm of the deviation of the estimated ground plane normal vector \mathbf{n} , from canonic ground plane normal vector \mathbf{n}_0 (9), and the norm of the deviation of the two consecutive estimations of the ground plane normal vectors \mathbf{n}_t and \mathbf{n}_{t-1} (10).

$$S_1(t) = \|\mathbf{n}_t - \mathbf{n}_0\|. \quad (9)$$

$$S_2(t) = \|\mathbf{n}_t - \mathbf{n}_{t-1}\|. \quad (10)$$

On Fig. 4 we show plots of S_1 and S_2 for 100 frames of the real-world sequence (tour003#003). Clearly, the stabilized method is more accurate than direct one. Surprisingly, this accuracy level difference is not dramatic.

As future steps we consider to involve Kalman filtering for the ground plane tracking and adapt the proposed method to piecewise linear surface reconstruction.

ACKNOWLEDGMENT

N.C. is supported by the European Commission (NEST-2003-012963). M.M.V.H. is supported by research grants received from the Excellence Financing program of the K.U.Leuven (EF 2005), the Belgian Fund for Scientific Research – Flanders (G.0248.03 and G.0234.04), the Interuniversity Attraction Poles Programme – Belgian Science Policy (IUAP P5/04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), and the European Commission (NEST-2003-012963, STREP-2002-016276, IST-2004-027017, and IST-2007-217077).

REFERENCES

- [1] Z. Zhang, R. Weiss, A. Hanson “Qualitative obstacle detection” Computer Vision and Pattern Recognition, 1994. Proceedings CVPR’94., 1994 IEEE Computer Society Conference on, 1994, pp. 554–559.
- [2] M. Bertozzi, A. Broggi “GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection” Image Processing, IEEE Transactions on, **7**, 1998, pp. 62–81.
- [3] R. Labayrade, D. Aubert and J. Tarel “Real Time Obstacle Detection on Non Flat Road Geometry through ‘V-Disparity’ Representation” Proceedings of IEEE Intelligent Vehicle Symposium, 2002.
- [4] M. Betke, E. Haritaoglu and L. Davis “Real-time multiple vehicle detection and tracking from a moving vehicle” Machine Vision and Applications, Springer, 2000, **12**, pp. 69–83.
- [5] N. Cornelis, K. Cornelis, and L. Van Gool “Fast compact city modeling for navigation pre-visualization”. In CVPR 2006, volume **2**, 2006, pp. 1339–1344.
- [6] N. Chumerin, M.M. Van Hulle “Cue and Sensor Fusion for Independent Moving Objects Detection and Description in Driving Scenes.” In: Signal Processing Techniques for Knowledge Extraction and Information Fusion, D.P. Mandic, M. Golz, A. Kuh, D. Obradovic, and T. Tanaka (Eds.), Springer, Boston, USA, 2007, pp. 161–180.
- [7] F. Solari, S.P. Sabatini, and G.M. Bisio, “Fast technique for phase based disparity estimation with no explicit calculation of phase,” Elect. Lett., vol. 37, no. 23, pp. 1382–1383.
- [8] K. Pauwels, “Computational Modeling of Visual Attention: Neuronal Response Modulation in the Thalamocortical Complex and Saliency-based Detection of Independent Motion”, PhD thesis, Katholieke Universiteit Leuven, 2008.
- [9] P.W. Holland, R.E. Welsch, “Robust Regression using Iteratively Reweighted Least Squares”, Commun. Statist. Theor. Meth., **A6**, 1977, pp. 813–827.
- [10] P.J. Rousseeuw, “Least Median of Squares Regression,” J. Am. Statistical Assoc., vol. 79, 1984, pp. 871–880.
- [11] M. Fischler, R. Bolles “Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography”, CACM, **24**(6), June 1981, pp. 381–395.
- [12] A. Beaton, J. Tukey “The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data,” Technometrics **16**(2), 1974, pp. 147–185.