Generalizing Inverse Perspective

Mohamed Kotb and Steven Beauchemin
Department of Computer Science
The University of Western Ontario
London, Ontario
N6A-5B7

January 31 2005

Abstract

Inverse perspective mapping schemes have been presented in the past as frameworks in which autonomous navigation could be performed on flat surfaces, using optical flow as the main percept. We propose a generalization of these models for uneven terrains. Given an inclinometer indicating the direction of gravity and some simple odometry parameters, the visual and positional information are combined to reconstruct the 3d elevation map of navigational surfaces within the visual field. Camera transformations are applied directly on the image plane. We provide noise and sensitivity analysis using random, zero-mean Gaussian noise to determine the robustness of the proposed model.

Keywords

bird's eye model, autonomous navigation, perspective mapping, rough terrains, optical flow

1 Introduction

Biological vision systems have been the object of growing attention in the past two decades [4, 8, 7]. The robustness of visual systems in primates and man partly stems from their inherent ability to integrate several sensory percepts simultaneously. Researchers are inspired by biological vision systems as they correctly perform considerable amounts of intricate perceptual tasks. In particular, the structure of optical flow can be mathematically complex and relatively difficult to use in reconstructing 3d environmental surfaces. For that reason, it is often useful to impose constraints on the spatiotemporal structure of optical flow to obtain acceptable levels of numerical accuracy. For instance, Mallot's inverse perspective mapping model eliminates the effects of perspective in optical flow, as long as the navigational path on the terrain remains flat [8]. This is accomplished by bringing the line of sight of the visual sensor perpendicular to the navigational surface with a coordinate transform, as depicted in Figure 3. Hence, as the visual sensor moves in a straight line onto a flat surface, the optical flow estimates appear isotropically parallel, with their magnitudes proportional to surface depth.

We propose a bird's eye model that generalizes Mallot's inverse perspective to include uneven navigational surfaces. Our model is supported by additional mathematical transformations which maintain the parallelism of optical flow vectors for uneven terrains, modeled as triangulations of randomly generated surface height points. As we demonstrate, it is possible to maintain correct optical flow structures despite the uncontrolled motions experienced by the visual sensor while navigating on an uneven terrain. We also provide noise analysis to the reconstructed 3d surface heights. Our proposed model is designed to be utilized for outdoor autonomous navigation, without *a-priori* knowledge of the navigational environment. Additionally, the model requires simple odometry capable of recording speed and inclinometers to estimate the direction of the gravitational force.

Our contribution is organized as follows: Section 2 defines the coordinate systems involved; Section 3 presents Mallot's bird's eye model; Section 4 outlines the problems encountered while applying this model on uneven terrains; Section 5 is an extension of Mallot's model that is aimed at maintaining the properties of the optical flow from the bird's eye perspective while navigating on uneven terrains; Section 6 presents a noise sensitivity analysis for the proposed mathematical model; and Section 7 concludes our contribution and outlines future research directions.

Proceedings of the Second Canadian Conference on Computer and Robot Vision (CRV'05)
0-7695-2319-6/05 $ 20.00 **IEEE**

2 Coordinate Systems

The projection of a 3d world point onto the image plane involves three coordinate systems and two transformations. The world coordinate system, W, is described with three primary axes, X, Y, and Z. A point in the 3d world is denoted by P_w and the coordinates of this point are $(P_{w_x}, P_{w_y}, P_{w_z})$.

The point P_w is transformed from the world coordinate system into the visual sensor coordinate system, C, yielding a point $P_c = (P_{c_x}, P_{c_y}, P_{c_z})$, where c_x, c_y, and c_z are the sensor axes defined in the world coordinate system. Given a corresponding point P_w, the transformed point P_c is projected into the image plane coordinate system $I(a, b)$.

A fourth coordinate system, MIP, the *Modified Image Plane*, is used in order to describe the image plane points in the 3d world coordinate system which allows further 3d transformations of image points to obtain the bird's eye projection model. The 3d image coordinate system B is described as $B(M_x, M_y, M_z)$ where M_x, and M_y are descriptions of the values a and b on the image plane, and M_z is the depth of an image point in the world coordinate system.

3 The Bird's Eye Model

Sometimes it is desirable to eliminate perspective effects from imagery as a means to simplify the structure of the resulting optical flow [8]. Figure 1 and 2 show the perspective mapping and the corresponding optical flow for a perfectly flat, square terrain.

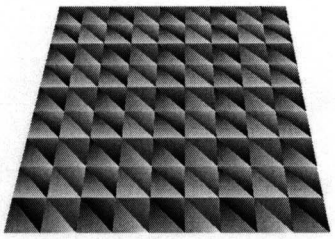

Figure 1: *Perspective projection of a perfectly flat, square navigational terrain.*

The required camera transformation for eliminating perspective effects from the imagery is the bird's eye model. Figure 3 depicts such a transformation, which brings the line of sight of the visual sensor in alignment with the direction of the gravitational force, provided that the flat terrain has no slope. However, the camera coordinate system (expressed in world coordinates) is not accessible in practice and the only available information resides on the image plane and in some intrinsic parameters of the visual sensor, such as its focal length and its height from the ground[1]. In this section, we present a mathematical formulation of the bird's eye model in which we only apply 3d transformations onto the image plane.

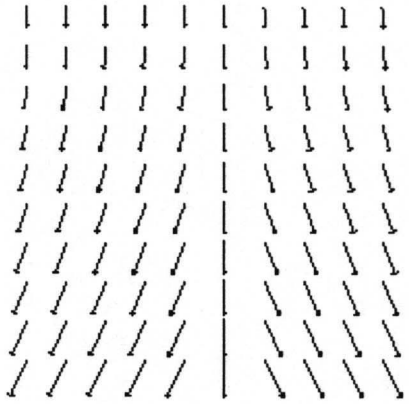

Figure 2: *The corresponding optical flow for Figure 1. The perspective effect appears as a divergent structure of the optical flow field.*

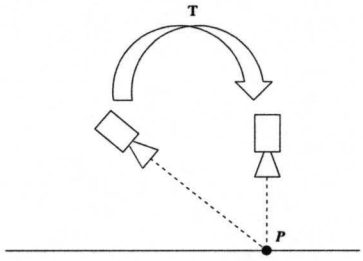

Figure 3: *The required camera transform for the bird's eye model. The transform causes the line of sight of the visual sensor to become perpendicular to the navigational surface.*

Figure 4 illustrates the projection of point P on the flat terrain, onto the image plane. Point L is the surface point intersecting the line of sight of the sensor and point L_p is its projection.

Figure 5 shows the geometric relation between points on the image plane and the nodal point, N, of the visual sensor. Distances L_d and P_d are used to

[1]The height of the visual sensor is described as the distance between the camera nodal point and the base of the mobile platform.

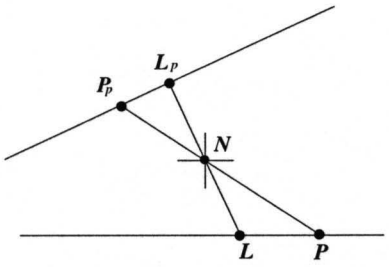

Figure 4: *The projection of point P from a flat terrain onto the image plane.*

recover the relative depths of points P_p and L_p with respect to the camera nodal point.

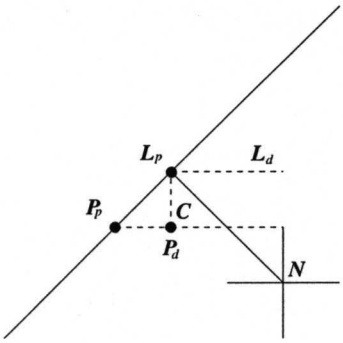

Figure 5: *Geometric relation between the nodal point of the visual sensor and image points.*

Given an angle θ that the sensor makes with the absolute horizon[2], then the distances L_d and P_d are as follows:

$$L_d = f\cos(\theta) \tag{1}$$

$$P_d = L_d + (p_{d_y} - L_{d_y})\cos(\theta) \tag{2}$$

where f is the focal length of the sensor, P_{d_y} and L_{d_y} are the heights of the image points P and L on the image plane. Every point in the image plane is then described in the MIP coordinate system as:

$$M_x = M_1 - \frac{A}{2} + eye_X \tag{3}$$

$$M_y = M_2 - \frac{B}{2} + eye_Y \tag{4}$$

$$M_z = P_d \tag{5}$$

In these equations, M_1 and M_2 are the coordinates of an image point, eye_X and eye_Y are the X, and Y world coordinates of the nodal point, and A and B are the width and height of the image plane.

After the image points are described in the MIP system, they are rotated around the U axis of the sensor with the same angle the camera makes with the absolute horizon. The perspective mapping here is the pinhole camera projective model, and Mallot's inverse perspective mapping model [8]. Figures 6, 7, and 8 show the bird's eye model perspective mapping, the bird's eye model inverse perspective mapping, and the optical flow of the same scene, respectively. They were captured by the sensor from an arbitrarily chosen location.

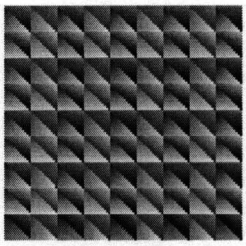

Figure 6: *Perspective mapping of a flat terrain, from a bird's eye perspective.*

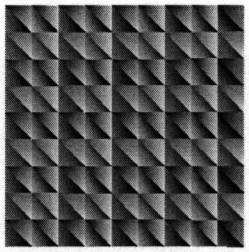

Figure 7: *The bird's eye inverse perspective mapping of the flat terrain from Figure 6.*

4 Rough Terrains

Generally, applying the inverse perspective model with a mobile agent moving on an uneven terrain yields optical flow fields that may not exhibit parallelism among their constituent vectors. This is exemplified by the following case where Figure 9 shows a 3d surface of irregular terrain and Figure 10 displays the optical flow resulting from navigation.

Figure 11 shows the reason behind the incorrect optical flow of Figure 10. The navigational surface

[2]The absolute horizon is defined as the plane perpendicular to the direction of the gravitational force.

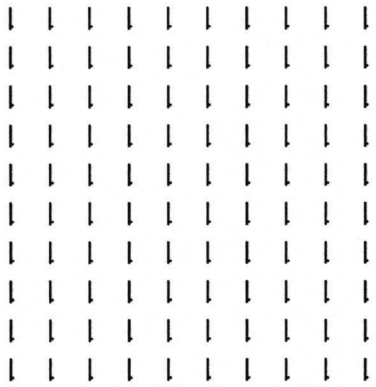

Figure 8: *Optical flow obtained while navigating on the terrain represented in Figure 6.*

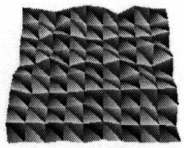

Figure 9: *A perspective view of an uneven terrain.*

point, P, in Figure 11 introduces perspective effects as the mobile agent progresses from one area to the next, due to the fact that the surface is uneven. Consequently, optical flow vectors deviate from the parallelism they should exhibit.

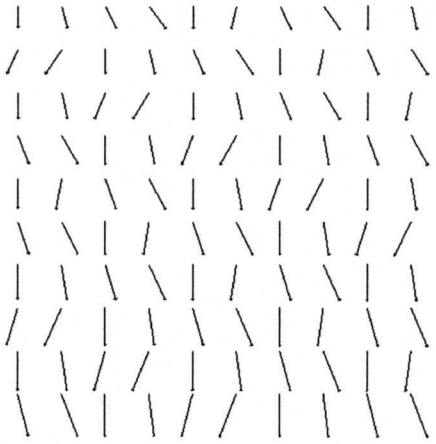

Figure 10: *Optical flow as the inverse perspective model is applied on uneven terrains.*

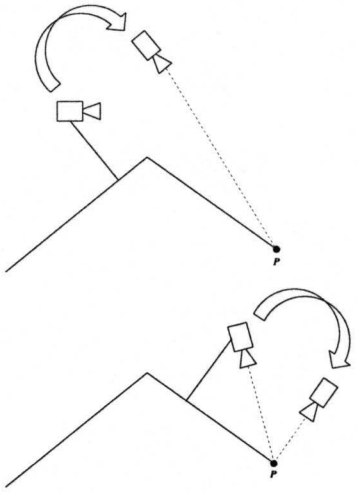

Figure 11: *The visual sensor using the inverse perspective transformation on a rough terrain. Perspective effects are introduced.*

Under such navigational conditions, it is clear the optical flow field cannot be used to construct an inverse perspective mapping that re-samples the bird's eye model. Mallot's model is not applicable on uneven navigational surfaces.

5 An Extension to the Model

In order to compensate for the errors of the inverse perspective model on elevated terrains, the points in the MIP coordinate system are rotated with the negative values of the encountered angles around the camera axes while navigating. These pitch and roll angles are obtained through an inclinometer describing the orientation of the mobile agent with respect to the absolute horizon. The inverse perspective model can then be generalized in the following way:

$$P = (Q + Tr_h)\mathrm{P}(-\alpha)\mathrm{R}(-\phi) \qquad (6)$$

where Q is an image point described in the MIP coordinate system, α and ϕ are the respective pitch and roll angles, and Tr_h is a translation:

$$P(-\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(-\alpha) & \sin(-\alpha) \\ 0 & -\sin(-\alpha) & \cos(-\alpha) \end{pmatrix}$$

$$R(-\phi) = \begin{pmatrix} \cos(-\phi) & \sin(-\phi) & 0 \\ -\sin(-\phi) & \cos(-\phi) & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

As it navigates on an uneven terrain, the mobile agent experiences height variations with respect to any arbitrarily determined reference point on the terrain. This, of course, introduces unwanted perspective effects, even while tilt and yaw are being corrected in the imagery acquired by the sensor. Therefore, a third transformation, this time requiring both the inclinometer and the speed odometry of the agent as inputs, needs to be formulated.

Figure 12 shows the agent moving on such a rough terrain. As the camera moves further down, its height with respect to a terrain point P decreases, thus creating a perspective effect. The following equation shows the transformation, Tr_h, which compensates for this effect:

$$Tr_h = \begin{pmatrix} 0 \\ -h' \\ 0 \end{pmatrix} \tag{7}$$

The quantity h' is obtained by assuming that the agent is moving with velocity V. The distance per time unit interval covered by the robot can then be computed as

$$\delta S = Vt \tag{8}$$

Given that the angle of the terrain surface is known to be θ by way of the inclinometer, then the change in camera height h' with respect to the virtual plane is calculated as follows:

$$h' = \delta S \sin\theta \tag{9}$$

The height compensation vector Tr_h keeps the terrain in the field of view with respect to a virtual plane passing through a reference point. The normal to this plane is parallel to the direction of the gravitational force.

Figures 13, 14, and 15 show the perspective, optical flow, and inverse perspective of the rough terrain displayed in Figure 9. The differences in magnitude of the optical flow vectors in Figure 14 contain the required information to determine the 3d heights of the corresponding 3d points from the navigational surface.

6 Noise and Sensitivity Analysis

The orientation and magnitude of ground-truth optical flow fields are corrupted by two independent, zero-mean Gaussian distributions. Consider ϵ_{angle}, a randomly generated number from a zero-mean Gaussian distribution with standard deviation σ_{angle}. We formed the disturbance angle θ_d as

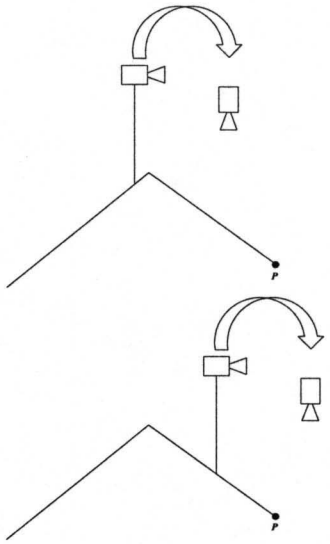

Figure 12: *The extension for the proposed camera on a rough terrain. The visual sensor remains perpendicular to the absolute horizon.*

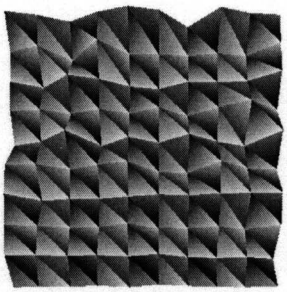

Figure 13: *A bird's eye perspective mapping of a rough terrain after applying the proposed extension.*

$$\theta_d = \epsilon_{angle} 2\pi \tag{10}$$

Consider ϵ_{mag}, a randomly generated number from zero-mean Gaussian distribution with standard deviation σ_{mag}. We formed a disturbance value to be added to the magnitudes of optical flow vectors as

$$\ell_{noisy} = \epsilon_{mag} \times \ell_{orig} \tag{11}$$

The output noise in the terrain reconstruction process is represented by the SSE (*Sum of Squared Errors*) between a noise-free inverse perspective mapping and the one reconstructed with the corrupted optical flow field. The following equation rep-

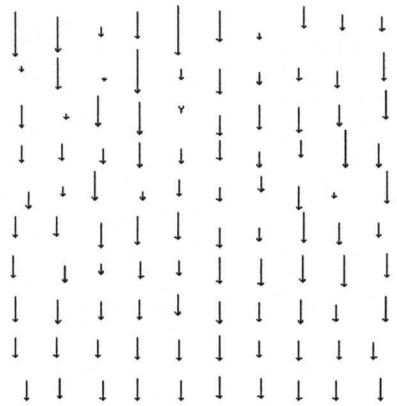

Figure 14: *Optical flow obtained while navigating on the terrain represented in Figure 13.*

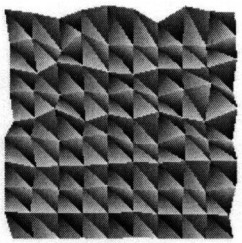

Figure 15: *Inverse perspective mapping of the terrain in Figure 13.*

resents our noise metric:

$$\mathrm{SSE} = \sum_{i=1}^{n} \sqrt{(x_i - \bar{x}_i)^2 + (y_i - \bar{y}_i)^2} \qquad (12)$$

where x_i and y_i are the reconstructed coordinates of a point P_i in the image, and $\bar{x}_i$ and $\bar{y}_i$ are the corresponding noisy ones.

Figure 16 shows the relation between the two standard deviations, σ_{angle} and σ_{mag}, within the range 0.0001 and 0.05 with step 0.01 and the SSE metric. As the Figure demonstrates, the error increases non-linearly with respect to the standard deviation corrupting the magnitude of the optical flow vectors. However, the output error behaviour for the input optical flow directional error appears to be linear.

It is apparent from this analysis that linear input noise generates non-linear output noise in the terrain reconstruction process. We believe this effect to be mainly due from expected sources, including the non-linear behaviour of perspective projection equations, and the relationship between optical flow from a bird's eye perspective with respect to the depth of environmental surfaces.

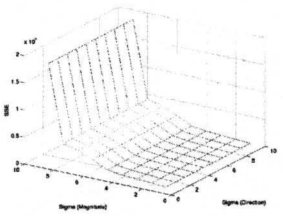

Figure 16: *SSE Vs standard deviation representing the noise in magnitudes and directions, every unit in the graph represents 0.01 of standard deviation. The experiment takes a standard deviation range from 0.0001 to 0.05.*

7 Conclusion

We propose a mathematical model for optical flow-based autonomous navigation on uneven terrains. We provide a mathematical formulation of the bird's eye model on flat terrains requiring image transformations only. We extend the model to uneven navigational surfaces and show that simple odometry and an inclinometer are required for our generalization. In addition, our model can be extended further to compensate for acceleration and deceleration, as long as this information is made available to the vision system through odometry.

The extension relies on the knowledge of the direction of the gravitational field and the speed of the mobile agent. We believe that an adequate level of robustness for navigational systems may only be achieved if visual processes are coupled with other sensory systems such as a sense of balance, which provides humans and mammals in general with inertial and gravitational information.

We are currently working towards generalizing our approach to stereo vision systems in order to fuse multiple channels of visual information. We hope this addition will enhance the robustness of the approach.

References

[1] J.L. Barron, D.J. Fleet, and S.S. Beauchemin. Performance of optical flow techniques. *IJCV*, 12(1):43–77, 1994.

[2] P.H. Batavia, S.A. Roth, and S. Singh. Autonomous coverage operations in semi-structured outdoor environments. In *IEEE International Conference on Intelligent Robots and Systems*, October 2002.

[3] S. Baten, M. Lutzeler, E.D. Dickmanns, R. Mandelbaum, and P.J. Burt. Techniques for autonomous, off-road navigation. *IEEE Intelligent Systems*, 13(6), 1998.

[4] S.S. Beauchemin, M.T. Kotb, and H.O. Hamshari. Navigation and gravitation. In K. Daniilidis, R. Klette, and A. Leonardis, editors, *to appear in Lecture Notes in Computer Science, Springer-Verlag, Berlin*, Dagstuhl Seminars, June 2004.

[5] J.S. Chahl and M.V. Srinivasan. A complete panoramic vision system, incorporating imaging, ranging, and 3-dimensional navigation. In *IEEE Workshop on Omnidirectional Vision*, June 2000.

[6] Y. Kim and H. Kim. Dense 3d map building for autonomous mobile robots. In *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, 2003.

[7] J. Lobo and J. Dias. Vision and inertial sensor cooperation using gravity as a vertical reference. *IEEE PAMI*, pages 1579–1608, 2003.

[8] H.A. Mallot, H.H. Bülthoff, J.J. Little, and S. Bohrer. Inverse perspective mapping simplifies optical flow computation and obstacle detection. *Biological Cybernetics*, 44(3):177–185, 1991.