Computer Science Department of The University of Auckland
CITR at Tamaki Campus (http://www.citr.auckland.ac.nz)

CITR-TR-88 April 2001

Dominant Plane Estimation

Kazuhiko Kawamoto [1], Reinhard Klette [2]

Abstract

Video sequences capturing real scenes may be interpreted with respect to a dominant plane which is a planar surface covering more than 50% of a frame, or being that planar surface which is represented in the image with the largest number of pixels.

This note shows a possible way for estimating the surface normal of this plane if just camera rotation is allowed.

[1] Computer Science Division, Department of Information and Image, Chiba University, 1-33 Yayoi-cho, Inage-ku, 263-8522, Chiba, Japan

[2] Centre for Image Technology and Robotics, The University of Auckland, Tamaki Campus, Auckland, New Zealand

You are granted permission for the non-commercial reproduction, distribution, display, and performance of this technical report in any format, BUT this permission is only for a period of 45 (forty-five) days from the most recent time that you verified that this technical report is still available from the CITR Tamaki web site under terms that include this permission. All other rights are reserved by the author(s).

Dominant Plane Estimation

Kazuhiko Kawamoto[1] and Reinhard Klette[2]

[1] Computer Science Division, Department of Information and Image Sciences
Chiba University, 1-33 Yayoi-cho, Inage-ku, 263-8522, Chiba, Japan
kazu@icsd7.tj.chiba-u.ac.jp

[2] CITR, University of Auckland, Tamaki Campus, Building 731
Auckland, New Zealand
r.klette@auckland.ac.nz

Abstract. Video sequences capturing real scenes may be interpreted with respect to a dominant plane which is a planar surface covering more than 50% of a frame, or being that planar surface which is represented in the image with the largest number of pixels. This note shows a possible way for estimating the surface normal of this plane if just camera rotation is allowed.

1 Introduction

Video sequences capturing real scenes may be interpreted with respect to a *dominant plane* which is a planar surface covering more than 50% of a frame, or being that planar surface which is represented in the image with the largest number of pixels. We are interested in analyzing the geometric situation as a case study to support future image analysis work for determining dominant planes in video sequences. In image analysis, we would actually use a method for calculating optical flow to approximate local displacements, and then we may use these optical flow fields to estimate a dominant plane in the image sequence. In our geometric study, see [1] for related models, we restrict ourself on ideal input data. We consider local displacement fields which accurately show the transition in projected surface positions from frame to frame. From all possible camera motions we only study the case of rotation in this note.

2 3D rotation motion

We define an $x_w y_w z_w$ world coordinate system. Let $\boldsymbol{x}_w = (x_w, y_w, z_w)^\top$ be a 3D point defined in the world coordinate system. If the world coordinate system rotates with rotation matrix $\boldsymbol{R}$, a 3D point $\boldsymbol{x}_w$ relatively rotates around the world coordinate system with rotation matrix $\boldsymbol{R}^\top$. In this case of rotation, a 3D point $\boldsymbol{x}_w$ moves to a 3D point $\boldsymbol{x}'_w$ as follows

$$\boldsymbol{x}'_w = \boldsymbol{R}^\top \boldsymbol{x}_w. \tag{1}$$

Assume the world coordinate system rotates around a fixed axis $\boldsymbol{l}$ with a constant angular velocity ω. Such a rotation after Δt seconds is expressed in the matrix form

$$\boldsymbol{R} = \boldsymbol{I} + \boldsymbol{\omega} \times \boldsymbol{I} \Delta t + O(\Delta t^2), \tag{2}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{\omega} = \omega \boldsymbol{l}$, $||\boldsymbol{l}|| = 1$. From eqs. (1) and (2), after Δt seconds, we obtain

$$\boldsymbol{x}'_w = \boldsymbol{R}^\top \boldsymbol{x}_w = (\boldsymbol{I} + \boldsymbol{\omega} \times \boldsymbol{I} \Delta t)^\top \boldsymbol{x}_w + O(\Delta t^2) = \boldsymbol{x}_w - \boldsymbol{\omega} \times \boldsymbol{x}_w \Delta t + O(\Delta t^2). \tag{3}$$

We define an $x_c y_c z_c$ camera coordinate system in such a way that the origin O_c is at the center of the lens and the z_c-axis coincides with the optical axis. Let $\boldsymbol{x}_c = (x_c, y_c, z_c)^\top$ be a

3D point defined in the camera coordinate system. For simplicity, we fix the camera coordinate system with respect to the world coordinate system in such a way the x_c-, y_c- and z_c-axes are in direction of the x_w-, y_w- and z_w-axes, respectively. Therefore, a 3D point $\boldsymbol{x}_w$ is viewed from the camera coordinate system as follows

$$\boldsymbol{x}_c = \boldsymbol{x}_w - \boldsymbol{c}, \tag{4}$$

where $\boldsymbol{c}$ is the position of the origin O_c with respect to the world coordinate system. From eqs. (3) and (4), we obtain

$$\boldsymbol{x}'_c = \boldsymbol{x}_c - \boldsymbol{\omega}\times(\boldsymbol{x}_c + \boldsymbol{c})\Delta t + O(\Delta t^2). \tag{5}$$

From eq. (5), the velocity of a 3D point $\boldsymbol{x}_c$ is given as follows

$$\dot{\boldsymbol{x}}_c = \lim_{\Delta t \to 0} \frac{\boldsymbol{x}'_c - \boldsymbol{x}_c}{\Delta t} = \lim_{\Delta t \to 0} \frac{-\boldsymbol{\omega}\times(\boldsymbol{x}_c + \boldsymbol{c})\Delta t + O(\Delta t^2)}{\Delta t} = -\boldsymbol{\omega}\times(\boldsymbol{x}_c + \boldsymbol{c}). \tag{6}$$

3 Planar surface local displacement in case of rotational motion

We define an uv image coordinate system in such a way that the origin is on the z_c-axis and the u- and v-axes are parallel to the x_c- and y_c-axes, respectively. For simplicity, we assume the focal length f is $f = 1$ and an image point $\boldsymbol{u}$ is expressed by

$$\boldsymbol{u} = (u, v, 1)^\top. \tag{7}$$

If image point $\boldsymbol{u}$ is a perspective projection of 3D point $\boldsymbol{x}_c$, we have

$$\boldsymbol{u} = \lambda \boldsymbol{x}_c = \frac{\boldsymbol{x}_c}{z_c}, \tag{8}$$

where λ is a nonzero scale factor. Differentiating eq. (8) with respect to time, we obtain the velocity of $\boldsymbol{u}$ as follows

$$\dot{\boldsymbol{u}} = \frac{\dot{\boldsymbol{x}}_c}{z_c} - \frac{\dot{z}_c \boldsymbol{x}_c}{z_c^2} = \frac{\dot{\boldsymbol{x}}_c - \dot{z}_c \boldsymbol{u}}{z_c}. \tag{9}$$

Furthermore, setting $\boldsymbol{k} = (0, 0, 1)^\top$ and substituting eq. (6) into eq. (9), we obtain

$$\dot{\boldsymbol{u}} = \frac{\dot{\boldsymbol{x}}_c - (\boldsymbol{k}^\top \dot{\boldsymbol{x}}_c)\boldsymbol{u}}{z_c} = \frac{(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{k}^\top)\dot{\boldsymbol{x}}_c}{z_c} = -(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{k}^\top)(\boldsymbol{\omega}\times(\boldsymbol{u} + \frac{\boldsymbol{c}}{z_c})). \tag{10}$$

Consider a spatial plane on which 3D point $\boldsymbol{x}_c$ lies. A spatial plane defined in the camera coordinate system is expressed by

$$\boldsymbol{n}^\top \boldsymbol{x}_c = d, \tag{11}$$

where $\boldsymbol{n}$ is the unit surface normal and d is the distance from the origin O_c. Substituting eq. (8) into eq. (11), we obtain

$$z_c = \frac{d}{\boldsymbol{n}^\top \boldsymbol{u}}. \tag{12}$$

Furthermore, substituting eq. (12) into eq. (10), we obtain a local displacement field

$$\dot{\boldsymbol{u}} = -(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{k}^\top)(\boldsymbol{\omega}\times(\boldsymbol{u} + \frac{\boldsymbol{c}\boldsymbol{n}^\top \boldsymbol{u}}{d})) = -(\boldsymbol{I} - \boldsymbol{u}\boldsymbol{k}^\top)(\boldsymbol{\omega}\times(\boldsymbol{I} + \frac{\boldsymbol{c}\boldsymbol{n}^\top}{d})\boldsymbol{u}). \tag{13}$$

4 Classification of local displacements of a planar surface

Consider $\boldsymbol{\omega} = (0, \omega_2, 0)^\top$ and $\boldsymbol{c} = (0, 0, c_3)^\top$, i.e. the camera rotates around the y_w-axis and the center of the lens is on the z_w-axis. In this case, the components of eq. (13) are written as follows

$$\dot{u} = -(\omega_2 u^2 + \omega_2 c_3 \frac{n_1}{d} u + \omega_2 c_3 \frac{n_2}{d} v + \omega_2 c_3 \frac{n_3}{d} + \omega_2), \tag{14}$$

$$\dot{v} = -\omega_2 uv, \tag{15}$$

where $\boldsymbol{n} = (n_1, n_2, n_3)^\top$. Therefore, the spatial plane parameters $\{n_1, n_2, n_3, d\}$ only depend on the velocity $\dot{u}$. Rearranging eq. (14), we obtain

$$\omega_2 c_3 u \frac{n_1}{d} + \omega_2 c_3 v \frac{n_2}{d} + \omega_2 c_3 \frac{n_3}{d} + (\omega_2 u^2 + \omega_2 + \dot{u}) = 0. \tag{16}$$

If we know the parameters $\{\omega_2, c_3, u, v, \dot{u}\}$, eq. (16) is a linear equation in three variables n_1/d, n_2/d and n_3/d. Therefore, with more than three image points and their velocities, we can compute the three variables n_1/d, n_2/d and n_3/d. Normalizing the vector $(n_1/d, n_2/d, n_3/d)^\top$ into a unit vector, we can determine the unit surface normal

$$\boldsymbol{n} = \frac{1}{\sqrt{(n_1/d)^2 + (n_2/d)^2 + (n_3/d)^2}} \begin{pmatrix} n_1/d \\ n_2/d \\ n_3/d \end{pmatrix}. \tag{17}$$

Furthermore, noting the following identity

$$\left(\frac{n_1}{d}\right)^2 + \left(\frac{n_2}{d}\right)^2 + \left(\frac{n_3}{d}\right)^2 = \frac{n_1^2 + n_2^2 + n_3^2}{d^2} = \frac{1}{d^2}, \tag{18}$$

we can determine the distance

$$d = \frac{1}{\sqrt{(n_1/d)^2 + (n_2/d)^2 + (n_3/d)^2}}. \tag{19}$$

Consider a scene with multiple spatial planes. In this case, a classification of image points is required, because image points corresponding to one plane are outliers for another plane. In eq. (14), we observe two properties as follows:

(a) For fixed u in eq. (14), a pair of the v-coordinates and the velocities of the u direction, $(v, \dot{u})$, lies on the following straight line

$$\dot{u} = av + b, \tag{20}$$

where

$$a = -\omega_2 c_3 \frac{n_2}{d}, \quad b = -(\omega_2 u^2 + \omega_2 c_3 \frac{n_1}{d} u + \omega_2 c_3 \frac{n_3}{d} + \omega_2). \tag{21}$$

Furthermore, from eq. (21), the different three variables n_1/d, n_2/d, and n_3/d give the different parameters a and b for fixed u.

(b) For fixed v in eq. (14), a pair of the u-coordinates and the velocities of the u direction, $(u, \dot{u})$, lies on the following quadratic curve

$$\dot{u} = cu^2 + du + e, \tag{22}$$

where

$$c = -\omega_2, \quad d = -\omega_2 c_3 \frac{n_1}{d}, \quad e = -(\omega_2 c_3 \frac{n_2}{d} v + \omega_2 c_3 \frac{n_3}{d} + \omega_2). \tag{23}$$

Furthermore, from eq. (23), the different three variables n_1/d, n_2/d, and n_3/d give the different parameters c, d and e for fixed v.

From (a), the classification problem for each u can be reduced to straight-line detection in $(v, \dot{u})$-space. Also, from (b), the classification problem for each v can be reduced to quadratic-curve detection in $(u, \dot{u})$-space.

One of the methods to detect straight lines and quadratic curves is the Hough transform [2]. By using the Hough transform, straight-line or quadratic-curve detection is transformed into peak detection in the parameter space. For straight line detection, each point $(v, \dot{u})$ maps the straight line $b = -va + \dot{u}$ in (a, b)-space. Also, for quadratic curve detection, each point $(u, \dot{u})$ maps the spatial plane $\dot{u} = u^2c + ud + e$ in (c, d, e)-space. Furthermore, by transforming peaks in the parameter space into the original space, each point in the original space can be labeled with the corresponding peak. If this labeling procedure is carried out for both vertical and horizontal lines of the image, each point (u, v) in the image is labeled with the corresponding straight line and quadratic curve. Then each point (u, v) in the image is labeled as follows

$$(a, b) \rightarrow (u, v), \quad (c, d, e) \rightarrow (u, v). \tag{24}$$

According to the label (a, b), each point is classified vertical line by line. Also, according to the label (c, d, e), each point is classified horizontal line by line. By combining the results of these two classifications, all points in the image are classified according to the corresponding spatial planes.

If the classification of all image points is achieved, the dominant plane in scene is determined by detecting the spatial plane supported by the largest number of image points.

5 Conclusions

This note shows a possible way for a new task formulated for video sequences: assume that the captured scene has a dominant plane such as the surface of a lake, a plane meadow, or a wall of a building. If just camera rotation is allowed then a Hough transform approach is suggested for solving the problem of calculating the surface normal of a dominant plane in the scene. Experiments will be necessary to illustrate this for real scenes where optical flow will be used to approximate local displacement.

References

1. K. Kanatani. *Geometric Computation for Machine Vision.* Oxford University Press, Oxford, 1993.
2. P. V. C. Hough. *A Method and Means for Recognizing Complex Patterns.* U.S. Patent. 3,069,654, 1962.