Detecting Height from Constrained Motion

Xueyin Lin and Zhigang Zhu

Al Lab, Computer Science Department
Tsinghua University, Beijing 100084, P.R. China

Abstract

In this paper a novel method called Height From Motion (HFM) is developed to estimate the motion and structure under planar motion. By using this method both translational and rotational motion (three degrees of freedom) can be treated in a unified manner. Based on HFM method, correspondence problem becomes easy to deal with, especially under translational motion. Experiments of real scene image sequences and the error analysis (theoretically and experimentally) have shown the efficiency and robustness of the method.

Introduction

A large body of work on the theory of extracting motion and structure of objects has been presented in the field of computer vision. However, as we want to apply them in practice, such as navigation of mobile robot based on motion analysis we would be very disappointed to find out that most of the methods can not be used with confidence. It is obvious that we have to seek alternating method to meet the challenge for practical use.

It is well known that if motion is constrained more or less, such as pure translational motion [1,2,3] or motion with tracking fixation point [4], the estimation of motion and structure parameters can be easier and more robust. However, another kind of constrained motion, planar motion, is more general than either situations. As navigation of mobile robot is concerned, method only applicable to translational motion can not deal with moving objects in the environment and rotational egomotion. And method with fixation point tracking needs sophisticated control mechanism.

We have developed a novel method called Height From Motion (HFM) under planar motion [5]. This method can deal with both egomotion and object motion, and treat translational and rotational motion in a unified manner. In this paper this method is developed further, including feature correspondence and error analysis. Theoretical analysis and experimental results have shown that this method is efficient and robust.

Detecting Height from Planar Motion

Coordinate Systems and Basic Equations

In our motion–vison model there are three coordinate systems: the camera centered coordinate system (CCC), the image coordinate system and the robot centered coordinate system (RCC) (Fig.1). The camera is mounted on the top of a mobile robot with a tilt angle α looking down. The location of origin of CCC with respect to RCC is (O,H,O), and the relationship of (X,Y,Z) in CCC and (x,y,z) in RCC for one space point can be expressed as follows:

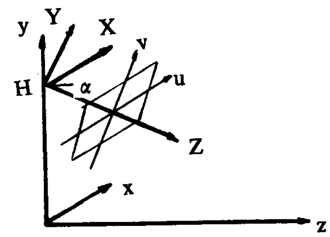

Fig.1. Coordinate Systems of HFM method

$$\begin{cases} X = x \\ Y = z\ \sin\alpha + (y - H)\ \cos\alpha \\ Z = z\ \cos\alpha - (y - H)\ \sin\alpha \end{cases} \tag{1}$$

Based on geometry of pinhole model the image coordinates of a point in space can be expressed as

$$u = f_u X / Z\ , \qquad v = f_v Y / Z \tag{2}$$

where f_u and f_v are used to account for the different scalar factors in u and v respectively.

From equations (1) and (2) the x and z coordinates of a point in RCC can be calculated as function of y as follows:

$$z = G[v](H - y), \qquad x = K[u,v](H - y) \tag{3}$$

where

$$G[v] = (f_v\ \cos\alpha + v\ \sin\alpha) / (f_v\ \sin\alpha - v\ \cos\alpha) \tag{4}$$

$$K[u,v] = (u / f_u)(G[v]\ \cos\alpha + \sin\alpha) = u\ \ K[v] \tag{5}$$

Now that $G[v]$ and $K[v]$ will be used frequently in our algorithm, and they can be calculated in advance, we will used two look–up tables (LUTs) to facilitate on-line calculation.

Unified Height From Motion (HFM) Algorithm

In this subsection the HFM algorithm will be derived under arbitrary (planar) motion. Now that there are three parameters to be determined (one is the rotation angle β about y axis, the other two are the translational components T_x and T_z), therefore more than one point is needed to estimate all of them. Suppose that the coordinates of a surface point P_i in the scene change from $(x_1^{(i)}, y_1^{(i)}, z_1^{(i)})$ to $(x_2^{(i)}, y_2^{(i)}, z_2^{(i)})$ with respect to RCC due to the rotational and translational motion, then we have

$$\begin{cases} y_2^{(i)} = y_1^{(i)} = h^{(i)} \\ x_2^{(i)} + T_x = x_1^{(i)} cos\beta - z_1^{(i)} sin\beta \\ z_2^{(i)} + T_z = x_1^{(i)} sin\beta + z_1^{(i)} cos\beta \end{cases} \quad (6)$$

where (i) denotes the ith point and will be used in the following equations. From equations (3) and (6) we also have

$$\begin{cases} K_2^{(i)} + T_x / (H - h^{(i)}) = K_1^{(i)} cos\beta - G_1^{(i)} sin\beta \\ G_2^{(i)} + T_z / (H - h^{(i)}) = K_1^{(i)} sin\beta + G_1^{(i)} cos\beta \end{cases} \quad (7)$$

where K_j and G_j (j=1,2) are used instead of $K[u_j, v_j]$ and $G[v_j]$ (j=1,2) respectively.

If two point pairs (i=1,2) are given, after some tedious manipulations we get

$$A \; sin\beta + B \; cos\beta = C \quad (8)$$

where

$$\begin{cases} A = K_2^{(1)} K_1^{(2)} - K_1^{(1)} K_2^{(2)} + G_2^{(1)} G_1^{(2)} - G_1^{(1)} G_2^{(2)} \\ B = K_1^{(1)} G_2^{(2)} + K_2^{(1)} G_1^{(2)} - G_2^{(1)} K_1^{(2)} - G_1^{(1)} K_2^{(2)} \\ C = K_2^{(1)} G_2^{(2)} + K_1^{(1)} G_1^{(2)} - G_2^{(1)} K_2^{(2)} - G_1^{(1)} K_1^{(2)} \end{cases} \quad (9)$$

After β is estimated by using (8) and (9), we can also obtained from equation (7) that

$$(H - h^{(1)}) / (H - h^{(2)}) = T_{xc}^{(1)} / T_{xc}^{(2)} = T_{zc}^{(1)} / T_{zc}^{(2)} \quad (10)$$

$$where \quad \begin{cases} T_{xc} = H(K_1 cos\beta - G_1 sin\beta - K_2) \\ T_{zc} = H(G_1 cos\beta + K_1 sin\beta - G_2) \end{cases} \quad (11)$$

Equation (10) indicates that the relative heights of the feature points can be obtained without giving the exact motion parameters.

If the height of one point can be estimated by other method or with a priori knowledge (e.g. some points on the ground can be identified from images), then the translational components can be estimated as

$$T_x = T_{xc}(1 - h/H), \qquad T_z = T_{zc}(1 - h/H) \quad (12)$$

On the other hand, given the translational parameters, we have

$$h = H(1 - T_x / T_{xc}) \quad (13)$$

$$or \qquad h = H(1 - T_z / T_{zc}) \quad (14)$$

After $h^{(i)}$ has been obtained for point P_i, its coordinates can be easily obtained from equations (3).

If there exists only translational motion (ie. $\beta = 0$), it is unnecessary to calculate β using 2 paris of points. In this case equations (10) to (14) also holds true and equation (11) can be reduced as

$$T_{xc} = H(K_1 - K_2), \qquad T_{zc} = H(G_1 - G_2) \quad (15)$$

Feature Correspondence under Constrained Motion

It is well known that feature correspondence is a severe problem to be solved so that the algorithm can be used in practice. In this section we will show that using HFM method the complexity of correspondence problem will be reduced significantly, especially when only translational motion is concerned.

Let $(u_1^{(i)}, v_1^{(i)})$ and $(u_2^{(i)}, v_2^{(i)})$ be the corresponding image points for point P_i in two frames while translational motion (T_x, T_z) is taken place. Without loss of generality, we assume that T_z is not zero. From equations (12) and (15) we can easily obtained that

$$(K_1^{(i)} - K_2^{(i)}) / (G_1^{(i)} - G_2^{(i)}) = T_x / T_z = c \quad (16)$$

where c is a constant dependent on the motion direction and independent of each feature point. It can be seen from equation (16) that if the pictures, captured while the robot moving, are transformed into a KG–space where (K,G) value is used for each feature point, a parallel image flow field will appear instead of a expansion field usually seen (see Fig. 3a and 3b). Obviously the direction of the parallel field can be determined without difficulty and the matching of each pair of feature points in two pictures can be solved by using parallel motion field constraint just as epipolar constraint used in stereo technique.

If there exist both translational and rotational motion, there will be two variable to be solved, one is the rotational angle β, and the other is the direction parameter c.

Several methods, e.g. Hough Transform, can be used to calculate the parameter set (c and β), and other constraints such as edges attached to each feature point can be used to reduce the complexity of matching.

Performance and Error Analysis

In this section we will discuss some issues about the performance of this method. At first the results of a preliminary experiment are described, and then the error analysis is given.

We have been doing experiments with a Hero–2000 mobile robot on which a 16 mm CCD camera is mounted. $512 \times 512 \times 8$ grey level images are processed by Imagebox image system with an IBM PC/AT as host computer. First the optical system is calibrated roughly based on equations (3). The calculated values of the parameters are:

$\alpha = 29.63°$, H=1450.30mm,
fu=950.21 pixels, fv=1266.94 pixels.

After the calibration procedure two LUTs G[v] and K[v] can be obtained with v varies from −255 to +255 (pixels).

Height from Egomotion of the Robot

While the robot moving a distance of about 300 mm, two frames are taken as shown in Fig. 2a and Fig. 2b. Then the edges and corners are extracted from these two frames (Fig. 2c and 2d). All the selected corners for correspondence are connected to some edges or ground marks. Based on parallel motion field constraint, Hough Transform is used to match the corner point. Fig. 3a and Fig. 3b shows the corresponded corner points and optical flow field in (u,v) and (K,G) values respectively. After that the Unified HFM method is used to estimate β using all these pairs, and the computed "translational parameters" (T_{xc} and T_{zc}) are calculated for each point. Then those

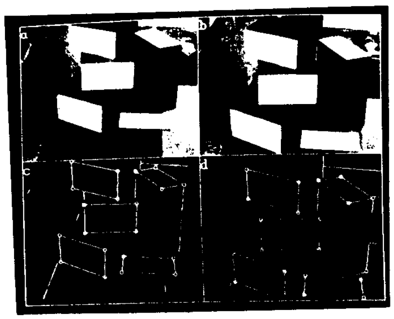

Fig.2. Original image pairs and their features
(a) first frame; (b) second frame; (c),(d) edges and corners for frame l and 2.

points which have minimum T_{zc} and T_{xc} (especially T_{zc} in this experiment, which also approximately equals to 300 mm) are identified as ground points (5 in this experiment), and the real translational parameters (T_x and T_z) are obtained from these points using equation (12). The motion parameters are $\beta = -1.19°$, $T_x = 25.19mm$ and $T_z = 269.72mm$, which coincide with the real motion quite well.

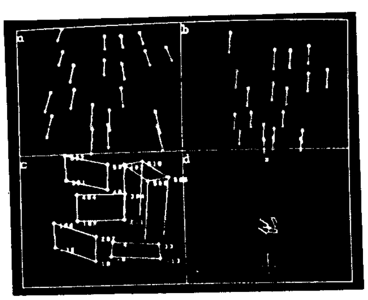

Fig.3. Feature correspondence and HFM calculations
(a) image flow field expressed in (u,v); (b) that in (K,G) ;(c) 3D structures of objects; (d) 2D maps of robot and objects.

Finally the coordinates of each of the points in RCC can be obtained based on equations (14) and (3). Fig. 3c shows the projections of calculated 3D coordinates in RCC of frame 1, with the height of each corner point labeled in the image. Fig. 3d is the 2D map (top view) which indicates the motion of the robot (represented by a circle) and the positions of detected obstacles.

Error analysis

At the mention of error analysis, there are two main factors that affect the performance of this method. One is the error caused by feature point locating error, the other is the calibration error. For simplicity we mainly deal with the case of translational motion.

Point Location Error Analysis Suppose a image point (u,v) is extracted with error (du, dv). By differentiating equation (4) with respect to v we have

$$dG = \frac{f_v}{(f_v \sin\alpha - v\cos\alpha)^2} dv \tag{17}$$

which indicates that the larger the v coordinate is, the larger the absolute errors of G[v] and z are, respective to the same amount of pixel locating error.

On the relation of precision of height estimation with respect to point locating error, we can find from equation (14) that

$$\frac{h}{H} = 1 - \frac{T_z}{H(G[v_1] - G[v_2])} \tag{18}$$

and

$$dh / H \doteq H(dG[v_2] - dG[v_1]) / T_{zc} \tag{19}$$

which indicates that errors of G values of each feature point affect the precision of height estimate.

From equation (17) and (19) we can see that high precision is expected near the observer (the robot) and in order to gain high performance the moving distance of the robot should be as large as possible. For the parameter set we use, the errors of z for the nearest and farthest visible ground points are 3.0 mm and 10.5mm, with respect to 1 pixel locating error in v.

For the above experiment, random noise of 0~ 1 pixels is added to each image point in frame 1 and frame 2. We use (du, dv) (pixels) as the pixel locating errors, $d\beta$(degrees) the error of angle β, $(dT_x, dT_z)(mm)$the error of translational vector (T_x, T_z), (dx,dy,dz)(mm) as the error of 3D coordinate vector (x,y,z) in RCC and D−size (mm) representing the error of distance (size) between two adjacent points of the objects. The average errors, errors of a nearby point P_{near}(u=75.27, v=−2.87, x=187.99, y=189.62, z=1733.13) and a far point P_{far} (u=−130.67, v=218.40, x=−388.88, y=510.45, z=2132.32) are given in Table 1.

Table 1 . Experimental results for random pixel error (|n| means n is a absolute value)

Cases	du dv	$d\beta$	dT_x	dT_z	dx	dy	dz	D-size
Average error	0.52 0.50	-0.24	7.16	1.80	|2.24|	|11.58|	|23.52|	|6.96|
Point P_{near}	-0.37 0.87	/	/	/	-1.32	3.49	-2.62	-2.67
Point P_{far}	0.67 -0.40	/	/	/	-1.81	-10.19	25.73	7.58

Table 2 . Experimental results for system error

Cases	$d\alpha$	$d\beta$	dT_x	dT_z	dx	dy	dz	D-size
Average error	1.00	-0.52	12.62	-11.18	|4.16|	-25.81	-50.21	|9.50|
Point P_{near}	1.00	/	/	/	-1.54	-28.35	-39.06	-0.06
Point P_{far}	1.00	/	/	/	1.64	-37.92	-37.29	21.14

System Error Analysis The camera parameters would be inaccurate due to the calibrating errors. If the tilt angle α has a small error $-d\alpha$(i.e. the actual angle $\alpha^* = \alpha + d\alpha$), and the optical system exists a small pan angle $d\beta$ and a small swing angle $d\gamma$, then the relation between CCC and RCC becomes

$$\begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & d\gamma & -d\beta & 0 \\ -d\gamma & 1 & d\alpha & 0 \\ d\beta & -d\alpha & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & cos\alpha & sin\alpha & 0 \\ 0 & -sin\alpha & cos\alpha & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y-H \\ z \\ 1 \end{bmatrix} \quad (20)$$

If $d\beta$ and $d\alpha$ can be negligible, then we have from (20) that

$$\begin{cases} X = x \\ Y = (y-H)(cos\alpha - sin\alpha\ d\alpha) + z(sin\alpha + cos\alpha\ d\alpha) \\ Z = -(y-H)(cos\alpha\ d\alpha + sin\alpha) - z(sin\alpha\ d\alpha - cos\alpha) \end{cases} \quad (21)$$

Hence from equation (2) and (21) we have

$$Z = G'[v](H-h)$$

where

$$G'[v] = \frac{fv\ cos\alpha + v\ sin\alpha - fv\ sin\alpha\ d\alpha + v\ cos\alpha\ d\alpha}{fv\ sin\alpha - v\ cos\alpha + fv\ cos\alpha\ d\alpha + v\ sin\alpha\ d\alpha} \quad (22)$$

Combining equation (22) with equation (4) we have

$$\frac{dG}{G[v]} = \frac{G[v] - G'[v]}{G[v]} = \frac{(G[v] + 1/G[v])d\alpha}{1 + G[v]d\alpha} \quad (23)$$

From equation (23) it can be seen that if $d\alpha < 0$ and $|d\alpha|$ increases, then $dG < 0$ and the estimated G[v] is $< G'[v]$ and deceases (It should be noticed that $G[v] > 0$), hence the position of a point in space will be estimated lower and nearer than it really is.

Experiment was done with tilt angle changes slightly (from 29.63 ° to 30.63 °). Table 2 show the results the above experiment ($d\alpha$ is the error measurement of tilt angle α), which agree with the theoretical analysis. From Table 2 we can find that the y and z coordinate value decrease when tilt angle increases, but the estimate of the distance between two point is quite robust.

Conclusions

In this paper we has developed further our method called Height From Motion (HFM). Though some constraint is imposed on motion to force each moving point to lie on a plane, it is still very general in real applications. Under this constraint, motion and structure of objects can be determined without difficulty. We emphasize that using HFM method, the complexity of correspondence problem can be reduced. Experiments of real scene images and error analysis have demonstrated the efficiency and robustness of the algorithm. This method would be useful in mobile robot navigation, such as environment building by vision, obstacle detection, navigation in a complicated environment, moving targets detection and tracking, and etc., Recently, several correspondenceless algorithms based on the similar principles have been investigated and developed in our lab [6]. We will further exploit the advantages of this method to build a practical vision system used in our lab.

Acknowledgements
The authors would like to thank reviewers for helpful comments.

References

[1] Basu, A., aloimonos, J., "A robust algorithm for determining the translation of a rigidly moving surface without correspondence, for robotics applications," Proc. IJCAI, pp. 815-818, 1987.

[2] Lawton, D.T., "Processing Translational motion sequences, "CVGIP vol. 22, pp 116-144, 1983.

[3] Horn, B.K.P., Weldon, Jr.E.J., "Computationally efficient methods for recovering translational motion," Proc. 1st ICCV, pp.2-11, 1987.

[4] Sandini, G., Tistarrelli, M., "Active tracking strategy for monocular depth inference over multiple frames, " IEEE Trans. on Pattern Anal. and Machine Intell., vol.12, No.1, 1990.

[5] Zhu, Z.G., Lin, X.Y., "HEM: a robust and efficient method for detecting structure and constrained motion, " Pacific Rim International Conf. on Artificial Intelligence, Naogoya, Japan, 1990.

[6] Zhu, Z.G., Lin X.Y., "Realtime algorithms for obstacle avoidance by using reprojection transformation", IAPR Workshop on Machine Vision Application, Tokyo, Japan, Nov. 28-30, 1990.