An Algorithm for Estimation of Camera Motion and Object Depth by Making Position Dependent Use of Imaged Points

Naoto KURASAWA Oichi ATODA

Graduate School of Bio-Applications and Systems Engineering,
Tokyo University of Agriculture and Technology
Nakacho 2-24-16,Koganei-shi,Tokyo,184
E-mail:kurasawa@atlab.base.tuat.ac.jp,atoda@cc.tuat.ac.jp
JAPAN

Working Group IC WG V/III

Key Words: Depth Estimation,Camera Motion

Abstract

Images of objects in the central part of a picture are insensitive to approaching or receding motion of the camera while all images are equally sensitive to lateral shift. Clued by this simple fact, motion parallax is directly resolved into six components corresponding to translation and rotation of the camera in order to estimate relative depth of imaged objects.

1 Introduction

A human is able to perceive three dimensional world by one eye through his motion. But probably he can not guess how distant a small LED hung alone in the dark is. Images of many objects move very consistently on his retina to let him know how they are arranged in the space.

The integral motion of images can be considered as the superposition of depth-dependent and depth-independent translations, depth-dependent radial or so called zooming motion, and depth-independent rotation. The latter two are also dependent upon location in the image plane, which is a clue of direct decomposition of motion parallax into its components in an artificial vision system Then relative three-dimensional positions of objects are readily estimated through unknown camera motion in six degree of freedom without factorization or inversion.

2 Theory

2.1 Assumptions and notations

Suppose that a camera arrives at the origin of XYZ coordinate system and is directed toward $+Z$ after a small translation $(-\delta X, -\delta Y, -\delta Z)$ accompanied by small yaw, pitch and roll whose magnitude are $-\xi/b, -\eta/b$ and $-\phi$ radians respectively where b denotes the length between lens and image plane of the camera. Suppose that a point at (X, Y, Z) is imaged by that camera. Offset or parallax $(\delta x, \delta y)$ of the image at $(x, y) = (bX/Z, bY/Z)$ in the image plane caused by that motion can be written as follows when the six translating and rotational motion components are small enough for the parallax to be approximated as the sum of their independent contributions.

$$(\delta x, \delta y) = b(\frac{\delta X}{Z}, \frac{\delta Y}{Z}) - (\frac{\delta Z}{Z})(x, y) + (\xi, \eta) + \phi(-y, x) \quad (1)$$

where the first and the second terms come from $(-\delta X, -\delta Y)$ and $-\delta Z$ movement of the camera respectively while the rests are attributed to yaw, pitch, and roll.

2.2 Estimation of offset epipolar line

If the parallax $(\delta x_k, \delta y_k)$ observed for each points $(X, Y, Z) = (X_k, Y_k, Z_k)$ can be decomposed into above four terms, estimation of Z_k will be very much easier. Note that significance of the four terms is not uniform over the image plane but the less (x_k, y_k) deviates from the center of image plane, the more insignificant the second and the fourth are. Especially, if the image resides at the center, they will be zero. Our idea of direct decomposition starts from this simplest fact.

In a practical scene, enormous number of points are imaged. Even in small central part of the image plane many image points will be observed. For such points, as long as $\sqrt{{x_k}^2 + {y_k}^2} \leq \varepsilon$, only the first and the third terms are substantial and the rests can be neglected or dealt as small perturbations as follows.

$$(\delta x_k, \delta y_k) = b(\frac{\delta X}{Z_k}, \frac{\delta Y}{Z_k}) + (\xi, \eta) + noise \quad (2)$$

The average m_x, m_y, the variances c_{xx}, c_{yy}, and the

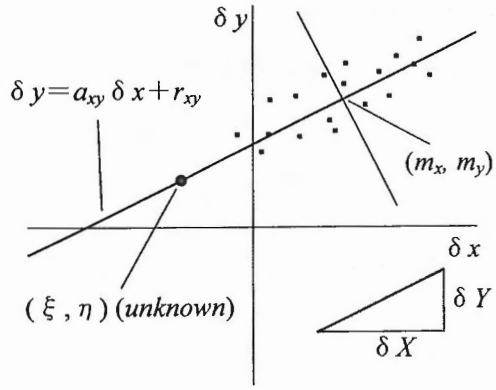


Figure 1: Offset epipolar line

covariance c_{xy} of $\delta x_k, \delta y_k$ will be readily estimated. The estimation of $a_{xy} = \delta Y/\delta X$ will be given from the eigenvector corresponding to the larger eigenvalue of the covariance matrix as follows if $c_{xx} \geq c_{yy}$.

$$a_{xy} = \frac{2c_{xy}}{(c_{xx} - c_{yy} + \sqrt{(c_{xx} - c_{yy})^2 + 4c_{xy}^2}}$$

If $c_{xx} < c_{yy}$, x and y shall be interchanged.

The line in $(\delta x, \delta y)$-plane shown in Fig. 1 and given by

$$\begin{aligned} \delta y &= a_{xy}\delta x + m_x - a_{xy}m_y \\ &= a_{xy}\delta x + r_{xy} \end{aligned} \tag{3}$$

will be referred to as the offset epipolar line since it is offset by r_{xy} which reflects some but not all fractions of yaw and pitch. In (x, y)-plane the direction along $y = a_{xy}x$ is called the epipolar direction as well.

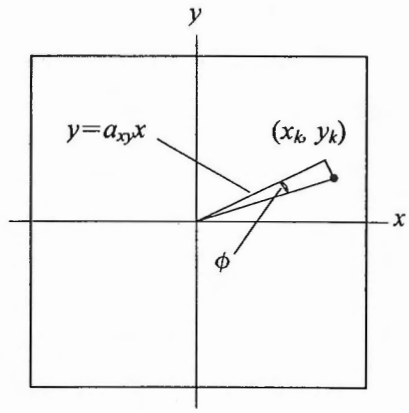
Depth Z_k in three dimensional XYZ space can be expressed as

$$\frac{b}{Z_k} = \frac{\delta x_k + a_{xy}\delta y_k - \omega}{\delta X + a_{xy}\delta Y} \tag{4}$$

but two parameters $\delta X + a_{xy}\delta Y$ and $\omega = \xi + a_{xy}\eta$ remain unknown. To estimate absolute depths of imaged points up to this stage, a couple of points whose depths are known must be referred.

2.3 Extinction of roll components

For an image located near the perimeter of the image plane the second and the fourth terms in $(\delta x_k, \delta y_k)$ will not be so small since at least either of x_k and y_k are large enough. Among them, the fourth is a nuisance for depth estimation contrary to the third which retains depth information. For an image in epipolar direction, geometrical representation of the second term is collinear to that of the first term, then the sum of the former three terms resides on the offset epipolar line while the fourth is perpendicular to the offset epipolar



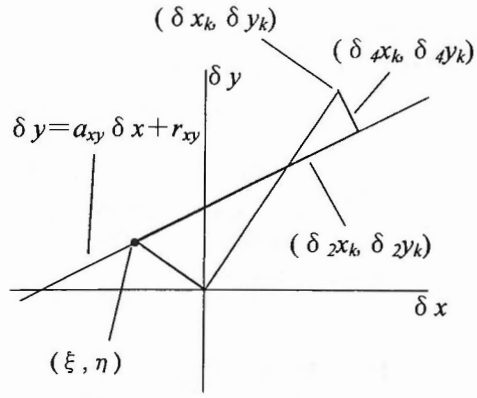


Figure 2: Detection of roll components

line as shown in Fig. 2. This fact enables ϕ to be estimated from the statistics of several points (x_k, y_k)-s approximately in epipolar direction as

$$\phi = \frac{\Sigma l_k \sqrt{1 + a_{xy}^2} x_k}{\Sigma(1 + a_{xy}^2)x_k^2} \tag{5}$$

where

$$l_k = \frac{a_{xy}\delta x_k - \delta y_k + r_{xy}}{\sqrt{1 + a_{xy}^2}}$$

Then from all $(\delta x_k, \delta y_k)$-s the fourth terms $(\delta_4 x_k, \delta_4 y_k) = \phi(-y_k, x_k)$ are subtracted. Resultant parallax without the roll component is denoted by $(\delta_{123} x_k, \delta_{123} y_k)$.

2.4 Extraction of Z-translation component

Now the second term of $(\delta_{123} x_k, \delta_{123} y_k)$ is extracted unless (x_k, y_k) is in the epipolar direction since its direction is the same as (x_k, y_k) while the sum $(\delta_{13} x_k, \delta_{13} y_k)$ of the rest terms is on the offset epipolar line as shown in Fig. 3. The vector from $(\delta_{123} x_k, \delta_{123} y_k)$ extending along $y - \delta_{123} y_k = (y_k/x_k)(x - \delta_{123} x_k)$ and reaching the offset epipolar line is no other than the second term $(\delta_2 x_k, \delta_2 y_k)$ which is formulated as follows.

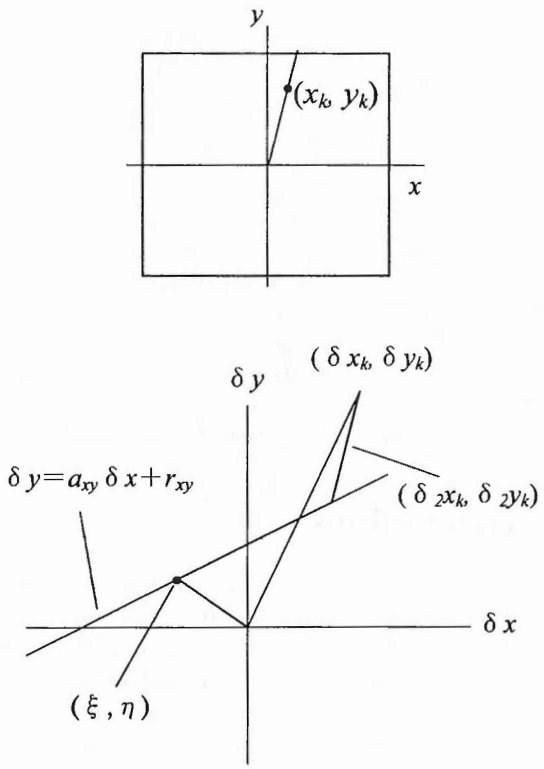


Figure 3: Extraction of Z-translation component

$$(\delta_2 x_k, \delta_2 y_k) = (\frac{x_k(a_{xy}\delta_{123}x_k - \delta_{123}y_k + r_{xy})}{a_{xy}x_k - y_k},$$

$$\frac{y_k(a_{xy}\delta_{123}x_k - \delta_{123}y_k + r_{xy})}{a_{xy}x_k - y_k})$$

The sum of remaining terms are expressed as

$$(\delta_{13}x_k, \delta_{13}y_k) = (\frac{x_k\delta_{123}y_k - y_k\delta_{123}x_k - x_k r_{xy}}{a_{xy}x_k - y_k},$$

$$\frac{a_{xy}x_k\delta_{123}y_k - a_{xy}y_k\delta_{123}x_k - y_k r_{xy}}{a_{xy}x_k - y_k}).$$

2.5 Unification of lateral and Z components

The magnitude of the second term is

$$\frac{(a_{xy}\delta_{123}x_k - \delta_{123}y_k + r_{xy})\sqrt{x_k^2 + y_k^2}}{a_{xy}x_k - y_k} = \sqrt{x_k^2 + y_k^2}\delta_2 v_k$$

where

$$\delta_2 v_k = \frac{a_{xy}\delta_{123}x_k - \delta_{123}y_k + r_{xy}}{a_{xy}x_k - y_k}$$

which includes depth information as

$$\frac{1}{Z_k} = \frac{\delta_2 v_k}{\delta Z}. \tag{6}$$

On the other hand, the first term also contains the depth information as

$$\frac{1}{Z_k} = \frac{\delta_{13}w_k - \omega}{b(\delta X + a_{xy}\delta Y)} \tag{4'}$$

where

$$\delta_{13}w_k = \delta_{13}x_k + a_{xy}\delta_{13}y_k$$

$$= \frac{(1 + a_{xy}^2)(x_k\delta_{123}y_k - y_k\delta_{123}x_k) - (x_k + a_{xy}y_k)r_{xy}}{a_{xy}x_k - y_k}$$

which must coincide with the former though the latter contains two unknowns $\delta X + a_{xy}\delta Y$ and ω while in the former only δZ is unknown. In practical estimation process, $\delta_2 v_k$-s and $\delta_{13}w_k$-s are sampled over many imaged points for stability in regions off-centered and out of the epipolar direction. Then the averages m_v, m_w, variances c_{vv}, c_{ww}, and covariance c_{vw} are statistically estimated to obtain the following relations concerning to unknowns.

$$a_{vw} = \frac{2c_{vw}}{c_{vv} - c_{ww} + \sqrt{(c_{vv} - c_{ww})^2 + 4c_{vw}^2}},$$

$$\omega = m_v - a_{vw}m_w$$

$$\frac{\delta X + a_{xy}\delta Y}{\delta Z} = \frac{a_{vw}}{b}$$

Yaw and pitch are derived from ω and $r_{xy} = \eta - a_{xy}\xi$ as

$$\xi = \frac{\omega - a_{xy}r_{xy}}{1 + a_{xy}^2}, \eta = \frac{a_{xy}\omega + r_{xy}}{1 + a_{xy}^2}.$$

Finally, from

$$Z_k = \frac{(b\delta X - x_k\delta Z) + a_{xyz}(b\delta Y - y_k\delta Z)}{\delta_{12}x_k + a_{xyz}\delta_{12}y_k}$$

where

$$(\delta_{12}x_k, \delta_{12}y_k) = (\delta x_k, \delta y_k) - \phi(-y_k, x_k) - (\xi, \eta)$$

$$= \frac{1}{Z_k}(b\delta X - x_k\delta Z, b\delta Y - y_k\delta Z)$$

and

$$a_{xyz} = \frac{b\delta Y - y_k\delta Z}{b\delta X - x_k\delta Z}$$

the following estimation of the depth Z_k-s yields for all images in the image plane.

$$Z_k = \delta Z((1+a_{xy}^2)(x_k^2+y_k^2)+a_{vw}(a_{vw}-2x_k-2a_{xy}y_k))$$

$$/(\delta_{12}x_k(a_{vw}-(1+a_{xy}^2)x_k)+\delta_{12}y_k(a_{vw}a_{xy}-(1+a_{xy}^2)y_k)) \tag{7}$$

where δZ is an inherently unknown scale parameter common to all k's.

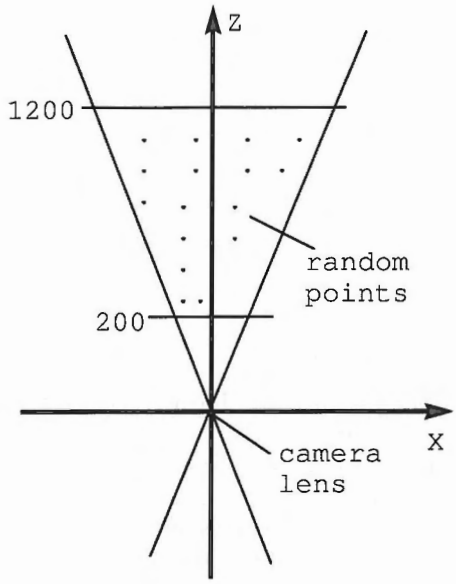


Figure 4: Arrangement for Simulation Experiments

3 Corrections for larger camera motion

In case that camera translation is not so small but $(-\Delta X, -\Delta Y, -\Delta Z)$ which makes parallax exceed, say 10% of the size of image plane. Then, Parallax becomes

$$(\Delta x, \Delta y) = b(\frac{X+\Delta X}{Z+\Delta Z}, \frac{Y+\Delta Y}{Z+\Delta Z}) - b(\frac{X}{Z}, \frac{Y}{Z})$$

$$= b(\frac{\Delta X}{Z}, \frac{\Delta Y}{Z}) - b\frac{\Delta Z}{Z}(\frac{X+\Delta X}{Z+\Delta Z}, \frac{Y+\Delta Y}{Z+\Delta Z})$$

But $(b(X+\Delta X)/(Z+\Delta Z), b(Y+\Delta Y)/(Z+\Delta Z))$ is no other than the position of the image before the camera motion, which should be substituted instead of (x_k, y_k) in previous discussions. Considering simultaneous camera motion as an ordered sequence of small roll and translation followed by small yaw plus pitch, parallax is decomposed in the same manner to obtain the depth from the camera just before yaw and pitch. Oblique depth seen from the final camera position should be compensated by multiplying $(1 + (x_k\xi + y_k\eta)/b^2 + (\xi^2+\eta^2)/2b^2)$.

4 Simulation experiments

Since the estimation algorithm are grounded on some approximations, verification by simulation studies is of much importance. In simulation experiment, 1000 points are randomly distributed within 1000 by 1000 by 1000 space 200 apart from the origin in Z direction as shown in Fig. 4. Images are taken by a simulated camera with unity length of b and image plane sized to 0.8 by 0.8. An example result is shown in Table. 1 for which the camera was translated by (40, 20, 30) and given 3, 0,and 5 degrees of yaw, pitch and roll respectively. 21 points were adopted for estimation of offset epipolar line, 14 for roll angle, and 18 for Z motion. The result proves that the algorithm is valid.

Table 1: Result of experiment

	ΔX	ΔY	ΔZ
estimated value	35.50	17.02	28.84
true value	40	20	30

5 Discussions

The size of the central area in which images are extracted for estimation of the offset epipolar line denoted by ε in Chap. 2 and angular margin in selecting images for roll estimation should be determined by experimental tradeoff since making them smaller improves approximation but decreases noise tolerance owing to statistics.

To make the algorithm more robust, self-checking should be incorporated. Information for self-checking can be derived from those parameters shown below.

(1) The ratio of two eigenvalues $\lambda_{xy1}/\lambda_{xy2}$ of covariance matrix of δx_k and δy_k which represents the validity of offset epipolar line.

(2) Consistency of sign in x_k, y_k and $\delta_2 x_k, \delta_2 y_k$. All the $\delta_2 x_k/x_k$ and $\delta_2 y_k/y_k$ must have the same sign according to whether the camera approaches or recedes.

(3) correlation coefficient between $\delta_2 v_k$ and $\delta_{13} w_k$ which must be as large as 1.

If the validity conditions are not met, warning may be issued and estimation should be discarded to wait for the new camera motion.

A convergent algorithm can be introduced to adjust the epipolar direction precisely in order to attain more correct estimation.

6 Conclusion

An algorithm without inversion is presented for estimation of relative depth of imaged objects based on unknown camera motion in six degree of freedom. In the algorithm parallax information is unevenly availed according to the sensitivity to motion components. Though it is grounded on rather crude approximation, simulation experiment demonstrated that the algorithm works well.