Direct Estimation of Rotation from Two Frames via Epipolar Search [*]

Sébastien Roy[**] and Ingemar J. Cox

NEC Research Institute
4 Independence Way
Princeton, NJ 08540, U.S.A.

Abstract. A direct method for estimating the rotational motion between two image frames is developed. The algorithm does not require knowledge of image correspondences, optical flow or scene structure and only assumes approximate knowledge of the translational motion. Spatial and temporal intensity gradients are avoided, resulting in an algorithm that is noise resistant. Moreover, the algorithm does not assume a particular projection model and is valid for both orthographic and perspective models.

It is based on a statistical measure of epipolar misalignment. Specifically, that (1) the intensity histograms of corresponding epipolar lines are invariant (ignoring occlusions) and, more importantly, that (2) the histograms of "almost corresponding" epipolar lines are similar. This latter property is a function of the spatial correlation present in the image and it is empirically demonstrated to be well behaved over a large class of scenes. These epipolar properties of histograms, i.e. that the difference between two histograms is a minimum when the two epipolar lines truly correspond and (approximately) increases monotonically with the degree of misalignment between two "epipolar" lines, allows the rotational motion to be estimated in a straightforward manner as a 3-dimensional "epipolar search".

Experimental results are presented on the SRI JISCT stereo database to empirically support the epipolar properties of intensity histograms. The calibrated NASA helicopter flight sequence is then analyzed to quantify the accuracy with which the rotations can be estimated. Experimental results indicate that very precise rotational estimates can be achieved.

1 Introduction

The problem of estimating the ego-motion and structure from two image frames has long been studied in computer vision. The solution to this problem has utility

---

[*] Published in 6th International Conference on Computer Analysis of Images and Patterns, Prague, September 1995, (p. 880-887)

[**] Visiting from the Université de Montréal, Département d'informatique et de recherche opérationnelle, C.P. 6128, Succ. Centre-Ville, Montréal, Québec, Canada, H3C 3J7

in many domains, especially for navigation and three-dimensional scene reconstruction. There are two distinct classes of structure-and-motion algorithms. The first is feature-based and assumes that there is a known number of feature correspondences between the two frames [7, 11, 12, 3]. While it has been shown that very few correspondences are needed in theory to solve the structure-and-motion problem, in practice, these algorithms are very sensitive to noise and many more correspondences are needed to stabilize the solution. Unfortunately, it is often the case that no feature correspondences are known a priori.

This paper is concerned with the class of *direct* methods of motion-and-structure estimation in which explicit feature correspondences are not required. By using intensity derivatives and exploiting the brightness-change constraint equation, solutions for motion-and-structure are derived directly. Negahdaripour and Horn [8] and Horn and Weldon [6] develop algorithms for the special cases of (1) known depth, (2) pure rotation and (3) pure translation or known rotation, (4) a planar world. Hanna [4] describes an iterative, multi-resolution approach that assumes piecewise planar (5×5) patches for the observed structure. Because the global minimization for ego-motion is non-linear, an initial estimate of the rotation and translation is required. Subsequently, Hanna [5] extended this work by combining stereo and motion analysis to the estimation of scene structure. However, this work is quite different from that proposed here, since the stereo pair introduces a third image.

This paper develops a direct method for determining the rotational ego-motion based on a search through the 3-dimensional rotational space. Such a search is possible only if there exists image properties such that each hypothesized ego-motion can be evaluated relative to one another. Section 2 derives two properties of intensity histograms computed along assumed epipolar lines based on the hypothesized three dimensional rotation. In Sect. 2.1, extensive experiments empirically show that the epipolar histogram properties are well behaved over the wide class of images contained in the SRI JISCT database. Generally, the statistical objective function has a clearly defined global minimum when the ego-motion parameters are correctly assigned and is approximately monotonically increasing with increasing error in the estimate of rotational motion.

Section 3 describes experimental results of applying the algorithm to the calibrated NASA helicopter flight sequence [10]. Quantitative comparison between the estimated and calibrated rotational motions are presented for the case of known translational motion and then for the case of small errors in translation.

2 Epipolar properties of intensity histograms

The displacement due to camera motion of an image point P'_a in image A to a point P'_b in image B can be decomposed in two components. The first component $\mathbf{M}_{P_a}$ is the rotational part of the displacement while the second component $\mathbf{E}_{P_a}$ is the epipolar vector, or translational part of the displacement. The relation between those components is expressed as $P'_b = P'_a + \mathbf{M}_{P_a} + e\mathbf{E}_{P_a}$, $(0 \leq e \leq 1)$, where e is the disparity along the epipolar vector. The rotational part $\mathbf{M}_{P_a}$

is *independent* of depth while the translational displacement $\mathbf{E}_{P_a}$ shifts points along the epipolar line by amounts that are inversely proportional to depth, as illustrated in Fig. 1.

Property 1 (Epipolar Histogram Invariance)
If we assume (1) *the constant brightness constraint, i.e. the brightness of an imaged point is unchanged by the motion of the camera, and* (2) *that the number of occlusions is small, then it is clearly the case that the histograms of the intensities of two corresponding epipolar lines are identical since the two lines contain identical pixel intensities, only their position may be changed because of depth.*

Now consider the case in which the camera motion contains a small error, either on its rotational or translational component. As a consequence, the "epipolar" lines will be erroneous, but close to the true epipolar lines. Will the intensity histograms of two almost corresponding epipolar lines be identical?

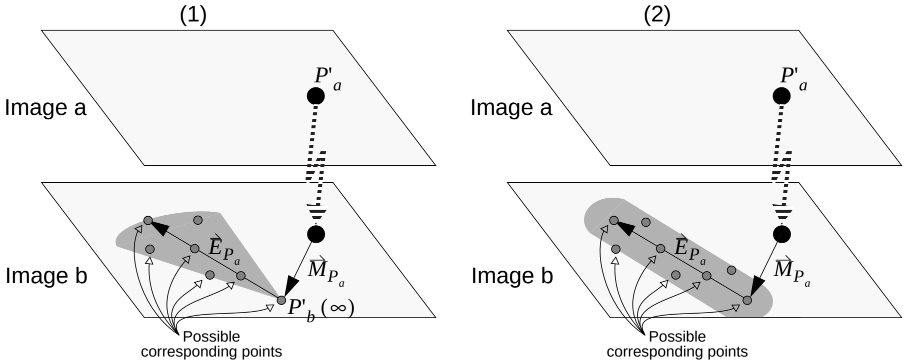

Fig. 1. Epipolar errors for inaccurate translation (1) and rotation (2). True corresponding points are in the neighbourhood (dark grey region) of the true epipolar lines.

Property 2 (Epipolar Histogram Correlation)
Assuming (1) *constant brightness constraint, and* (2) *that the number of occlusions is small, then the intensity histograms of two "pseudo-epipolar" lines that are spatially close to a pair of truly corresponding epipolar lines have similar (in a sum of squared errors sense) histograms. The difference between two pseudo-epipolar histograms is a minimum when the lines correspond to the true epipolar geometry and increases approximately monotonically with the size of the rotational error.*

Property 2 is valid for a large class of spatial imagery, despite the fact that it is straightforward to construct *artificial* images for which Property 2 does not hold. In [1] Property 2 is proved for the case in which intensity values are spatially Normally distributed, a condition that is true for many natural scenes.

Intuitively, Property 2 is a consequence of the high degree of spatial correlation [9] present in most images. As depicted in Fig. 1, small errors in the camera displacement cause a point P'_a in image A to be projected to a point which is spatially close to the true epipolar line $\mathbf{E}_{P_a}$. The smaller the error, the closer this point is to $\mathbf{E}_{P_a}$. Local image coherence then insures that the intensity value of an erroneous correspondence is close to the true intensity value that lies somewhere on the true epipolar line.

It can be shown that translational error generally creates less displacement from the true epipolar line then rotational error. In the limit case for points at infinity, the translational error has no effect whatsoever. That is, the rotational error is the dominant source of all point displacement and consequently of the sum of squared differences (SSD) between epipolar histograms. For translational error only, the SSD between epipolar histograms is much weaker and other statistical measures of the epipolar line must be used. In this paper, we therefore restrict ourselves to the case of unknown rotational motion and known (or approximately known) translation motion. It should be noted however, that the rotational motion so dominates the translational component that it may be possible to correct for the rotational error in the presence of an unknown translation.

2.1 Experimental verification of Property 2

We examined 23 stereo pairs from the SRI JISCT stereo database[3]. For different ranges of rotations (±20° around each axis) and known translation, a selected number of epipolar lines pairs were extracted according to the assumed ego-motion. The intensity histograms, $h_a(i)$ and $h_b(i)$, were then computed for each pair of pseudo-epipolar lines and the sum of squared distances, s_{ab}, calculated, i.e. $s_{ab} = \sum_{i=0}^{255} |h_a(i) - h_b(i)|$. This error was then averaged over all pairs of epipolar lines to arrive at a final similarity measure, S_{ab}, $S_{ab} = \frac{1}{N} \sum_{i=1}^{N} s_{ab}(i)$ where N is the number of epipolar line pairs extracted from the two images.

Figure 2 show the average sum of squared errors, S_{ab}, in the intensity histograms as a function of assumed rotational ego-motion about x, y and z axes[4], for two representative images pairs from the JISCT database. The curves show a very well defined minimum when the pseudo-epipolar lines have the same rotational component as the true epipolar lines, i.e. the rotation is correctly estimated. Moreover, the difference criterion, S_{ab}, is approximately monotonically increasing around a well defined minimum thus allowing in theory very rapid determination of the global minimum via gradient descent search.

---

[3] Although the JISCT database contains a total of 49 image pairs, many of these are repetitious, with only very small motion differences between pairs.

[4] The x and y axes are assumed to be oriented with the horizontal and vertical axes of the image plane and the z axes is aligned with the optical axis of the camera system.

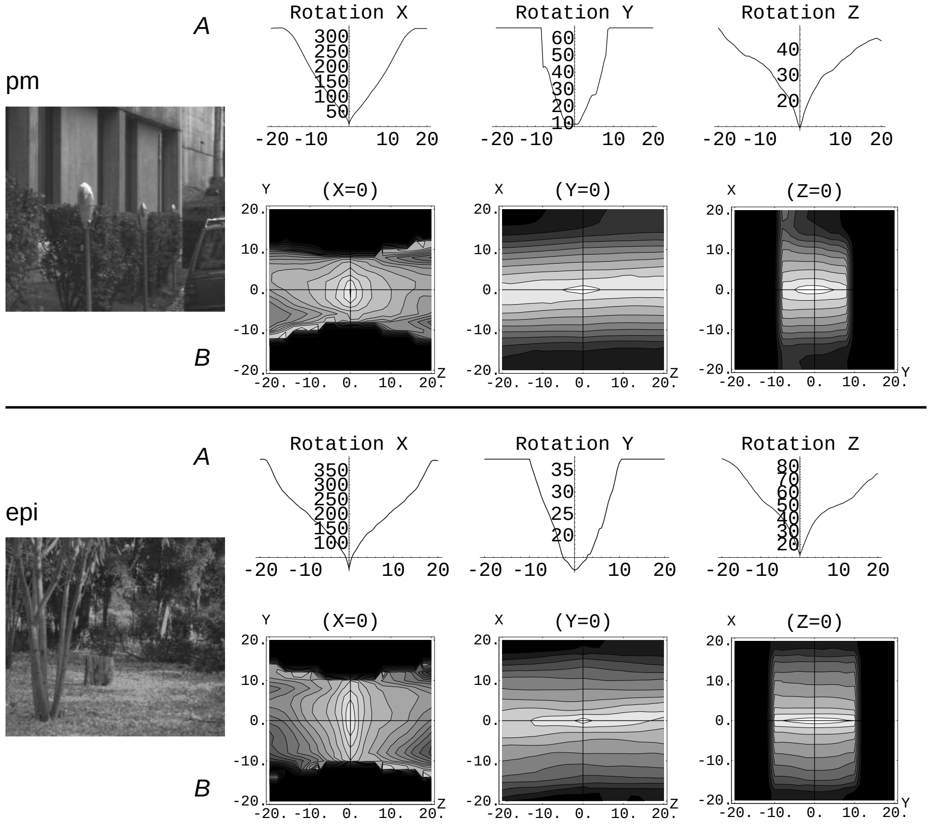

Fig. 2. Histogram mean square error, S_{ab}, as a function of rotation around x, y or z axis. A) Plots of individual axis. B) Plot of two axis of rotation. Minima should be at 0. All angles in degrees, spanning $\pm 20^\circ$.

Note that the vertical axes for the three component rotations have different scales. The effect is most pronounced for x rotations and least significant for y rotations. This is caused by the small displacement of epipolar line induced by the ambiguity of y-rotation and x-translation.

Experiments suggest that this epipolar property of intensity histograms is quite robust to noise as well as being effective for both small and large image displacement. Only 5 of the 23 images examined did not behave as expected and these exhibited very strong deviations from the constant image brightness (CIB) assumption or contained uncommon periodic textures. However, the sensitivity to the CIB constraint is no more so than other direct and indirect methods. Moreover, when the CIB assumption was not valid, this could be corrected for many times, via a simple prepocessing step [2] to significantly extend the applicability of the algorithm. As with other techniques, the algorithm is sensitive to images with a very small field of view since, under this condition, it becomes very difficult to differentiate between small rotations and translations. Finally, images that contained regular textures exhibited many strong local minima as well as the expected strong global minimum which make gradient descent search difficult. However, these images are not common and rarely occur in nature.

3 Epipolar search

The results of Sect. 2.1 indicate that the rotational components of the camera motion can be determined by a 3D search for the minimum of the average sum of squared differences, S_{ab}, provided that the initial rotational estimates are within $\pm 20°$ of their true values, which provides for large initial uncertainty.

To evaluate this approach, we estimated the three component rotations between every 9th frame of the calibrated NASA motion sequence. Each rotational component was searched in a range of $\pm 20°$ with a precision of $\pm 1°$. The solution obtained was then refined by an additional search of $\pm 3°$ with a precision of $\pm 0.15°$. The precision of the calibrated rotation on (x, y, z) is $(\pm 0.12°, \pm 0.7°, \pm 0.24°)$. Figure 3a shows the true and estimated rotations, when the translational error is zero. Very good quantitative agreement is achieved for all three axes of rotation. Figure 3b shows the true and estimated rotations, when the translational error is approximately $5°$. Very good quantitative agreement is achieved for all three axes of rotation. To demonstrate how well-behaved the histogram similarity function is, Fig. 3c shows the two dimensional contour plots around the selected minima for the pair of frames 36 and 45.

4 Conclusion

"Epipolar search", a new paradigm for determining the egomotion between two frames, was proposed based on statistical properties calculated along "corresponding" epipolar lines. One such statistical property is the sum of squared differences between intensity histograms calculated along epipolar lines. It was empirically shown that this epipolar histogram measure is a minimum when the epipolar geometry is known exactly and increases monotonically with error in the rotational components of egomotion. Intuitively, this is a consequence of the high degree of spatial correlation present in most images and can be theoretically proven for certain statistical classes of scenes. The method does not assume any prior knowledge of the structure of the scene and is applicable to both orthographic and projective imaging.

This property of epipolar histograms allow the rotational motion to be estimated in a straightforward manner as a 3-dimensional "epipolar search". In theory, a 5-dimensional search should be possible in which the two translational components of motion are also estimated. However, the effect is weaker for translational motions since points at infinity (background points) are only affected by rotational motion. Nevertheless, we believe that a full motion-and-structure

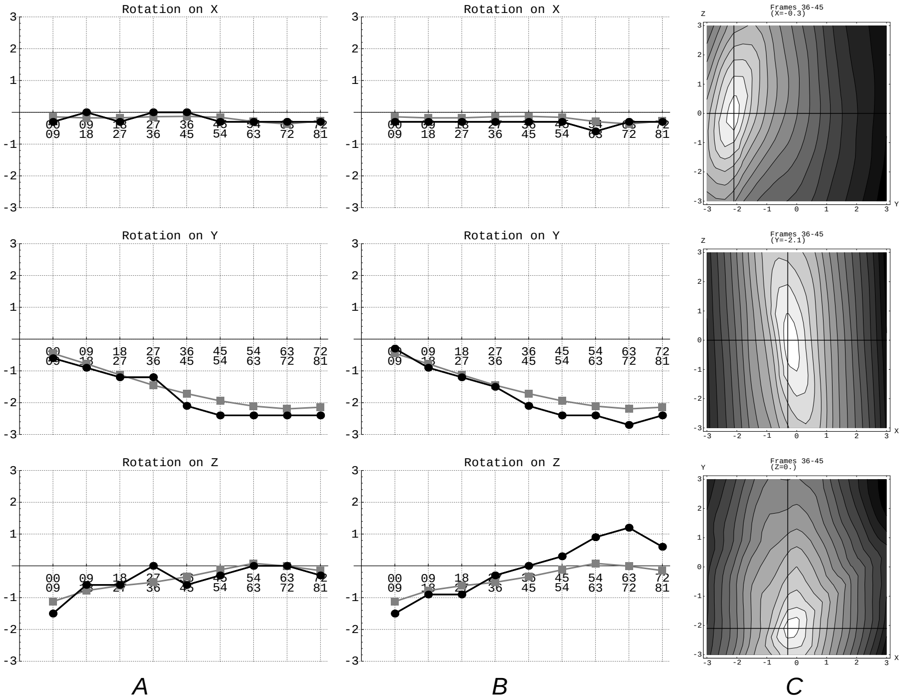

Fig. 3. A)True (squares) and estimated (circles) rotations for the NASA flight sequence, for a known translational motion. B)True (squares) and estimated (circles) rotations for the NASA flight sequence for a translational motion error of 5°. C)Mean square error as a function of rotation in $(x, y), (y, z), (x, y)$, for a translational motion error of 5°. All angles in degrees, spanning ±3°.

algorithm can be designed based on the paradigm of epipolar search and work is progressing in this direction, including the derivation of other statistical properties of epipolar misalignment.

Experimental results showed good agreement between the estimated and calibrated rotations of the NASA helicopter sequence when the translational motion was either known or erroneous within 5° accuracy. Further, the experimental results also suggest that epipolar search is quite robust to both noise and the relative magnitude of displacement between image frames. Preliminary work suggests that the translational motion can be estimated using other statistical measures along epipolar lines via an independent epipolar search. The proposed 5-dimensional search then reduces to an iterative refinement scheme in which the rotational motion is estimated by a 3-dimensional epipolar search and the translational motion by a 2-dimensional search.

The statistical properties of images calculated along epipolar lines appears to be an interesting research direction. Clearly, intensity histograms are just one of many possible statistical measures and it is importatnt to develop a mathematical framework in which they can be developed and evaluated. Finally, it is also apparent that the technique may be applied to the problem of tracking and stereo camera calibration both to estimate the initial geometry of the two cameras and to automatically refine that estimate over time should the geometry drift due to temperature, vibration or other environmental factors.

Acknowledgments

The authors thank Phillip N. Smith of NASA Ames Research Center for providing the flight data of Sect. 3. Also thanks to Lance Williams for many fruitful discussions regarding this work.

References

1. I. J. Cox and S. Roy. Probabilistic modelling of epipolar misalignment. In *Int. Workshop on Stereoscopic and Three Dimensional Imaging*, Santorin, Greece, September 1995.
2. I. J. Cox, S. Roy, and S. L. Hingorani. Dynamic histogram warping of images pairs for constant image brightness. In *IEEE Int, Conf. on Image Processing*, 1995.
3. O. Faugeras. *Three Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.
4. K. J. Hanna. Direct multi-resolution estimation of ego-motion and structure from motion. In *IEEE Workshop on Visual Motion*, pages 156–162, 1991.
5. K. J. Hanna and N. E. Okamoto. Combining stereo and motion analysis for direct estimation of scene structure. In *3rd Int. Conf. on Computer Vision*, pages 357–365, 1993.
6. B. K. P. Horn and E. J. Weldon Jr. Direct methods for recovering motion. *Int. J. of Computer Vision*, 2:51–76, 1988.
7. H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.
8. S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *IEEE PAMI*, 9(1):168–176, 1987.
9. W. K. Pratt. *Digital Image Processing*. John Wiley & Sons, 1978.
10. P. N. Smith, B. Sridhar, and B. Hussein. Vision-based range estimation using helicopter flight data. In *IEEE Proc. on Computer Vision and Pattern Recognition*, pages 202–208, 1992.
11. R. Y. Tsai. Multiframe image point matching and 3-d surface reconstruction. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 5(2):159–174, 1983.
12. J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from two perspective views: Algorithms, error analysis, and error estimation. *IEEE Trans. on PAMI*, 11(5):451–476, 1989.

This article was processed using the LaTeX macro package with LLNCS style