

## Motion without Structure

Sébastien Roy\* and Ingemar J. Cox  
 NEC Research Institute  
 4 Independence Way  
 Princeton, NJ 08540, U.S.A.

E-mail: sebastien|ingemar@research.nj.nec.com

### Abstract

*We propose a new paradigm, motion without structure, for determining the ego-motion between two frames. It is best suited for cases where reliable feature point correspondence is difficult, or for cases where the expected camera motion is large. The problem is posed as a five-dimensional search over the space of possible motions during which the structural information present in the two views is neither implicitly or explicitly used or estimated.*

*To accomplish this search, a cost function is devised that measures the relative likelihood of each hypothesized motion. This cost function is invariant to the structure present in the scene. An analysis of the global scene statistics present in an image, together with the geometry of epipolar misalignment, suggests a measure based on the sum of squared differences between pixels in the first image and their corresponding epipolar line segments in the second image.*

*The measure relies on a simple statistical characteristic of neighboring image intensity levels. Specifically, that the variance of intensity differences between two arbitrary points in an image is a monotonically increasing symmetrical function of the distance between the two points. This assumption is almost always true, though the size of the neighborhood over which the monotonic dependency holds varies from image to image. This range determines the maximum permissible motion between two frames, which can be quite large.*

*Experiments with both outdoor scenes and an indoor calibrated sequence achieve very good accuracy (less than 1 pixel image displacement error) and robustness to noise.*

### 1. Introduction

Much work has been done on trying to recover camera motion (i.e. ego-motion) parameters from image pairs. In almost all cases, either optical flow or feature point correspondences are used as the initial measurements. In the first case, some inherent problems (aperture, large motions, etc.) related to optical flow computation, suggest that errors can never be lowered to a negligible level (see [2, 8, 10, 15]). Even methods using the intensity derivatives directly or normal flow (see [1, 6, 7, 12, 14, 15, 17]), suffer from high noise sensitivity. For feature-based methods, the reliable selection and tracking of meaningful feature points is generally very difficult, see [3, 11, 16, 17].

All prior methods of ego-motion implicitly or explicitly determine the structure present in the scene. For example, while feature based methods compute a motion estimate directly, the structure is implicitly available given the feature correspondences. Direct methods explicitly estimate both the ego-motion and structure, typically in an iterative fashion, refining first the motion, and then the structure estimates. Thus, good motion estimation appears to require good structure estimation (or at least point correspondence estimation). In contrast, we propose a paradigm that we call *motion without structure*. Under this paradigm, the recovery of ego-motion is independent of any structure or correspondence estimation. The benefit is that there are only five unknown motion parameters to be estimated. As such, we expect that the approach should be both robust and accurate. The experimental results support this.

The algorithm relies on statistically modeling the image behavior in the neighborhood of a point, as discussed in Section 2.1. This model is then used to estimate the likelihood of an assumed camera motion. In [4], we proposed using the difference between histograms computed along assumed correspondence epipolar lines as a likelihood function. This statistical measure is very effective in determining the rotational component of ego-motion, but is not always a

---

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

reliable measure of the likelihood of a translational motion. Consequently, we proposed in [5] a likelihood measure based on the sum of sums of squared differences between pixels in one image and their hypothesized corresponding line segments in the other image that is a reliable estimate of either the rotational or translational components of motion. This measure is detailed in Section 2.2.

Determining the true motion is then accomplished by searching for the maximum likelihood estimate over the space of translations and rotations. The search is straightforward since we show in Section 2.3 that the function to be minimized has only one minimum (which is the solution), provided the image is well behaved, i.e. the variance between neighboring intensity points increases monotonically and symmetrically with the distance between the points. In previous work [5], the sub-problems of finding rotation or translation when the other component of motion is known was shown to be solvable by locating the single global minimum. This paper extends these results and considers the full motion case when *both* rotation and translation must be simultaneously estimated. The effect of motion ambiguity (see in [13]) on the accuracy of motion estimation is also discussed.

Section 3 presents experimental results from a comprehensive evaluation based on real images of stereoscopic pairs and an indoor calibrated motion sequence.

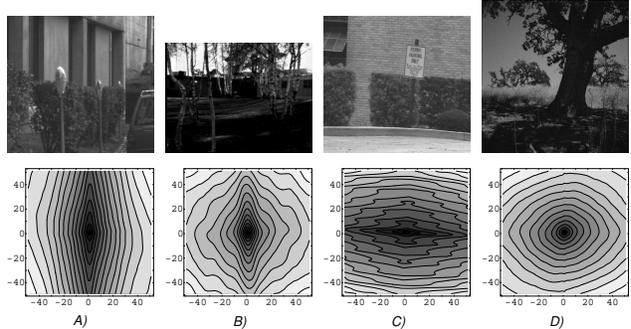
## 2. Motion Estimation as a 5-D search

Our goal is to determine the motion between two frames by a search over the space of possible rotations and translations. The number of parameters to be estimated are three for rotation and two for translation. Only two translational components are needed because the magnitude of the translation cannot be estimated, only its direction (due to the depth-scale ambiguity). The translation is thus assumed to have unit magnitude, and the estimation of translation reduces to determination of the direction of translation on the surface of a unit sphere.<sup>1</sup>

In order for such a search to be possible, a cost function is needed that evaluates the likelihood of an assumed motion. Essential characteristics of such a cost function are (1) invariance to structure in the scene, (2) a well-defined global minimum at the correct motion estimate, and (3) no local minima in the neighborhood of the correct motion.

In Section 2.2, we describe one such structure-invariant cost function, based on a simple statistical model of local intensity variation (see Section 2.1), that possesses these desired properties.

<sup>1</sup>Consequently, in the experimental section, the translational error is recorded in degrees over the unit sphere.



**Figure 1. JISCT image database. The four images A) *Parking meter*, B) *Birch*, C) *Shrub*, D) *Tree* are shown on top of their variance functions  $\sigma^2(\vec{\delta})$ . Distances along the axis are in pixels. Darker points have smaller variance.**

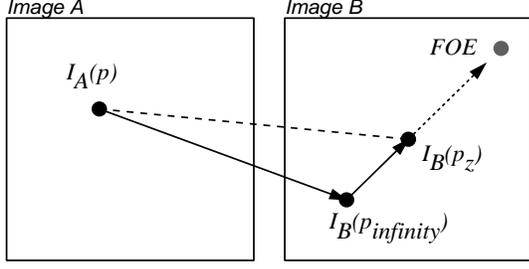
### 2.1. A statistical model of image intensities

A simple statistical model is used to represent image behavior around a point. Consider the intensity distribution in the neighborhood of a given image point  $\vec{p}$ . We are interested in the probability of differences in intensity between point  $\vec{p} + \vec{\delta}$  and  $\vec{p}$ , conditioned on the displacement  $\vec{\delta}$  between the two points.

This property is intuitively related to the correlation present in a scene. For a given image, we can evaluate the parameters of the distributions, namely  $\sigma^2(\vec{\delta})$ , for all possible displacements  $\vec{\delta}$ .

Example of these variance functions are shown in Figure 1 for a neighborhood of 50 pixels. The mean of the distributions is not shown here since it is always very close to 0. The variance functions increase approximately monotonically with distance, with a single minimum centered at  $\vec{\delta} = (0, 0)$ . This property is exploited to derive the likelihood measure in Section 2.2. Note that while the relationship between variance and distance is monotonically increasing, it is not always symmetrical, indicating that intensities are more correlated in certain directions. It is straightforward to find a mapping between two monotonically increasing functions to restore symmetry. This mapping will be applied to correct pixel value differences in the cost function.

Our experimental observations indicate that most natural images are well-behaved. We define a *well-behaved* image as one that possesses a monotonically increasing variance function. Only images that contain repetitive textures or those that are highly non-stationary, generally present badly-behaved (i.e. non-monotonic) variance functions. By examining how well-behaved the variance function is, it should be possible to measure how accurate the method is expected to perform.



**Figure 2. Basic geometry for known rotation. For a given  $I_A(\vec{p})$ , its unknown corresponding point  $I_B(\vec{p}_z)$  is on the line joining  $I_B(\vec{p}_\infty)$  and the FOE.**

## 2.2. A Depth-invariant cost function

We wish to evaluate the likelihood of a motion, composed of a rotational and a translational component, to be the true motion of the camera. As shown in Figure 2, for a given point  $I_A(\vec{p})$  in image A and a camera motion, we can compute the matching point  $I_B(\vec{p}_z)$  (the *zero-disparity* point) in image B that corresponds to infinite depth, as well as the *focus of expansion* (FOE). The point  $I_B(\vec{p}_\infty)$  is related to the rotational component of the motion while the FOE is related to the translational component.

Since we do not know the real depth  $z$  of point  $I_A(\vec{p})$ , we can only assume that the actual corresponding point  $I_B(\vec{p}_z)$  is somewhere in the neighborhood of point  $I_B(\vec{p}_\infty)$ . In fact, it is always located on the line joining the true  $I_B(\vec{p}_\infty)$  and the true focus of expansion.

For a given camera motion, a line segment,  $u$ , of length  $r_{max}$  is selected starting at the zero-disparity point  $I_B(\vec{p}_\infty)$  and oriented toward the FOE. The value of  $r_{max}$  is chosen to reflect the maximum disparity expected. After selecting a number of sample intensity values  $u_i$  along the segment  $u$ , we define the error measure  $e_u$  as

$$e_u = \sum_{i=1}^n (u_i - I_B(\vec{p}_z))^2 = \sum_{i=1}^n (u_i - I_A(\vec{p}))^2 \quad (1)$$

which will be a minimum when the segment  $u$  contains  $I_B(\vec{p}_z)$ . Equation 1 can assume that  $I_B(\vec{p}_z) = I_A(\vec{p})$  since these points correspond and therefore should have the same intensity value. To get a global estimate of the likelihood of a motion, we select a number of points  $I_A(\vec{p}_i)$  and compute the sum

$$S = \sum e_{q_i}$$

of the individual line segment errors  $e_{q_i}$  corresponding to each of these points.

The next section will show how this cost function satisfies the requirement enumerated in Section 2. It is

expected that for well-behaved images, this cost function will exhibit a single minimum at the true camera motion and that a simple search based on gradient descent will be sufficient to find it.

## 2.3. Convergence and smoothness properties

In order to successfully search over the motion space, the cost function must have a well-defined global minimum and few, if any, local minima. Section 2.3.1 shows that for a known rotation, the translational search space features only a single global minimum, assuming monotonic and symmetrical image intensity variances. The converse is also demonstrated, that is searching for rotation with known translation.

The preceding discussion assumed that either the translation or rotation was already known. In practice, both must be estimated. We do not have a proof of convergence for this situation and have proceeded with an experimental investigation to determine the utility of the cost function under these circumstances.

A second condition for successful search, is that the region of convergence should be large, to allow easy selection of an initial search point. This region (and the general smoothness of the function) should be derivable from the local image intensity statistics. Qualitatively, it is clear that large and frequent intensity variations do not allow a wide region of convergence (because of ambiguities) while low frequency variations allow for much larger motions.

### 2.3.1. Existence of a single minimum

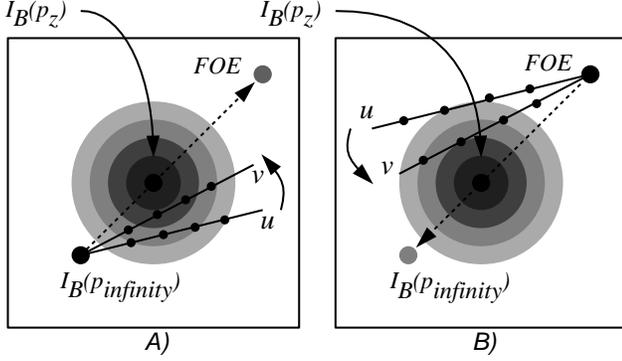
In this section we show that for well-behaved images, a single minimum of the error measure  $e_u$  of Equation 1 is observed when a segment  $u$  contains  $I_B(\vec{p}_z)$  and joins the true zero-disparity point and the true FOE. Since by definition a well-behaved variance function always features a global minimum at  $(0, 0)$ , this condition is enough to ensure that the likelihood function possesses a unique minimum. This is demonstrated next.

Consider a segment  $u$  in the neighborhood of  $\vec{p}_z$ , starting at  $\vec{p}_\infty$ , and containing  $n$  sample intensities as depicted in Figure 3A. Then we can assume that each sample behaves like a random variable  $u_i$  with distribution

$$f(u_i) = G_{[I_A(\vec{p}); \sigma^2(\vec{d}_{u_i})]}(u_i)$$

where  $G_{[\mu; \sigma^2]}$  is an arbitrary probability distribution and  $\vec{d}_{u_i}$  is the distance  $(x, y)$  from sample  $u_i$  to position  $\vec{p}_z$ , the unknown location of the corresponding point to  $I_A(\vec{p})$ . From Equation 1, the error measure  $e_u$  is a random variable defined as

$$e_u = \sum_{i=1}^n (u_i - I_A(\vec{p}))^2$$



**Figure 3. Error function for two segments  $u$  and  $v$ . When  $v$  is closer to  $\vec{p}_z$  than  $u$ , its expectation is smaller for a well behaved variance function. A) Unknown translation. B) Unknown rotation.**

with an expectation value defined as

$$E(e_u) = E\left(\sum_{i=1}^n (u_i - I_A(\vec{p}))^2\right) = \sum_{i=1}^n \sigma^2(\vec{d}_{u_i})$$

Suppose we now take a second segment  $v$  starting also at  $\vec{p}_\infty$ , but closer to the point  $\vec{p}_z$ . A set of samples  $v_i$  is chosen with the same sampling<sup>1</sup> as segment  $u$ . The error measure  $e_v$  is defined as the random variable

$$e_v = \sum_{i=1}^n (v_i - I_A(\vec{p}))^2$$

which has an expected value

$$E(e_v) = \sum_{i=1}^n \sigma^2(\vec{d}_{v_i})$$

where  $\vec{d}_{v_i}$  is the distance  $(x, y)$  from sample  $v_i$  to position  $\vec{p}_z$ . We now wish to show that the expectation of  $e_v$  is always smaller than  $E(e_u)$ . First, it is straightforward to see that

$$\|\vec{d}_{v_i}\| < \|\vec{d}_{u_i}\|, \quad \forall i$$

since  $v$  is a rotated version of  $u$  toward  $\vec{p}_z$ , except for the special pathological case where  $\vec{p}_z = \vec{p}_\infty$ . Second, the variance function  $\sigma^2(\vec{d})$  is assumed to be monotonically and symmetrically increasing with  $\|\vec{d}\|$  from  $\vec{p}_z$ . From these two observations, we can immediately conclude that

$$\sigma^2(\vec{d}_{v_i}) < \sigma^2(\vec{d}_{u_i}), \quad \forall i$$

It then follows that

$$E(e_v) = \sum_{i=1}^n \sigma^2(\vec{d}_{v_i}) < \sum_{i=1}^n \sigma^2(\vec{d}_{u_i}) = E(e_u)$$

<sup>1</sup>The case of different sampling and different lengths of  $u$  and  $v$  can also be handled in a more elaborate proof.

which shows that as we get closer to the segment containing  $I_B(\vec{p}_z)$ , the expected error value gets smaller until it reaches a minimum when the candidate FOE correspond to the true FOE. As long as the variance function is monotonic and symmetrical, this minimum is guaranteed to exist and is unique. Since this is true for any epipolar line segment, it is also true for the sum of these segments in global cost function. The same procedure is applied for rotation estimation, just by exchanging the role of the FOE and the zero-disparity point (see Figure 3B).

### 3. Experiments and Results

Results of the motion without structure method are shown here for different kinds of real images pairs and for a calibrated motion sequence. The image pairs are taken from the SRI JISCT stereo database which provide partial ground truth since the motion between frames is a horizontal translation. However, we do not exploit this knowledge during the estimation procedure, and only use it to qualitatively compare the estimated and expected motions.

Most of the motions estimated here have a small forward (or backward) component. Our experiments show that large forward translation is much easier to estimate than lateral (i.e. sideways) motion. This is caused by the infamous rotation-translation ambiguity stating that a lateral translation (i.e. little or no forward component) combined with a small camera field of view is hardly distinguishable from a rotation. Inversely, forward translation is not much affected by this ambiguity and therefore is easier to estimate.

#### 3.1. Searching the solution space

A direct search of the motion space is performed by approximating the gradient and following steepest descent. The algorithm usually needs around 60 to 100 iterations to converge to the solution. Much improvement could be made to this search method, since no emphasis has yet been put on speed.

In all experiments conducted, we took care to select realistic initial estimates, i.e. as far as possible from the solution while taking into account the convergence constraint derived from the image texture. It is important to note that in most practical situations of motion tracking, the motion parameters from the previous frame can be used as an initial estimate for the next frame, taking advantage of the fact that motion tends to be similar and thus allowing faster convergence.

For all the experiments presented, only about 4% of the points of the images are arbitrarily selected for likelihood estimation. The typical running time is between 30 seconds and 10 minutes on a 150 MHz Silicon Graphics workstation. The execution time can be reduced by



**Figure 4.** The *Pentagon* image pair. The solution is superimposed over the images as a grid of selected points with their corresponding epipolar segments. The epipolar line segments are approximately horizontal, indicating good alignment.



**Figure 5.** The *Tree*. The estimated motion is superimposed over the images as a grid of points and their corresponding epipolar segments. The motion is approximately horizontal.

selecting a smaller number of points, at the expense of less accuracy in the motion estimate.

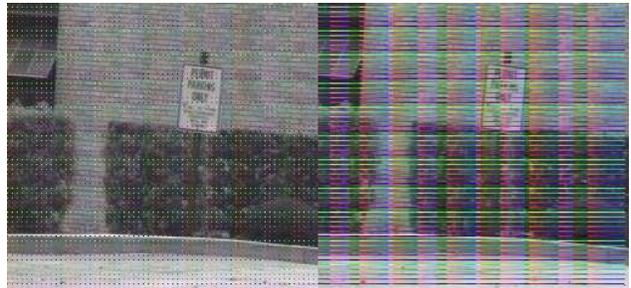
### 3.2. JISCT image pairs

The **Pentagon** image pair has a very well-behaved local intensity statistic. The image pair is very well aligned so that the motion between frames is purely due to horizontal translation. However, the magnitude of the translation is small, on average less than 2 pixels, which would usually make accurate estimation of the translation difficult.

The results are illustrated in Figure 4. The initial translation was  $35^\circ$  from the correct translation on the unit sphere, while the initial rotation was set to  $10^\circ$  around an arbitrary axis. The rotation obtained is  $0.17^\circ$ , corresponding to a maximum of 0.4 pixels error anywhere in the image. The true rotation is  $0^\circ$ . The translation obtained is  $(-0.994, -0.102, 0.035)$ , which correspond to a  $6^\circ$  error. While at first sight this appears large, we note that this is well within the accuracy of other two-frame algorithms [9, 15] and that, within the image, this error correspond to a maximum displacement of 0.3 pixel. The expected translation is  $(-1, 0, 0)$ .

The results for the **Tree** image pair, which also exhibits a pure horizontal translation, are illustrated in Figure 5.

The initial motion estimate is a translation oriented  $35^\circ$  from the horizontal on the unit sphere and the initial rotation estimate is  $5^\circ$ , corresponding to an image displacement of up to 12 pixels. The estimated translation is  $(0.996, -0.0765, 0.0485)$ , which is  $5.4^\circ$  from the true horizontal motion of  $(-1, 0, 0)$ . The estimated rotation is  $0.4^\circ$  which is believed to accurately reflect a slight vergence effect of the camera that can be manually observed.



**Figure 6.** The *Shrub*. The recovered motion (approximately horizontal translation), superimposed on the right image.

The third example, the **Shrub**, also features only a horizontal translation. However, this type of imagery is usually difficult to analyze because of the ambiguous textures presented by the brick wall and the bushes. The results are illustrated in Figure 6. The initial motion estimate has a translation at  $35^\circ$  from the horizontal and a rotation of  $5^\circ$ . The estimated translation is  $(-0.9992, 0.0369, 0.00836)$  which correspond to a  $2.2^\circ$  error, for an expected translation of  $(-1, 0, 0)$ . The estimated rotation is  $0.1^\circ$  which correspond to an image displacement of maximum 0.2 pixels.

### 3.3. The PUMA sequence

The motion without structure algorithm was tested on a Puma calibrated motion sequence, courtesy of the University of Massachusetts and shown in Figure 7. The rotation between each frame is approximately  $4^\circ$  around an axis parallel to the optical axis of the camera and located at  $(0.909, 0.416, -0.005)$  feet from the optical center of the camera.

We performed the motion analysis using only two successive frames at a time. The initial estimates for



Figure 7. The *Puma* image sequence, frames 1,4,7,10,13. The camera is at the end of a *Puma* robot arm rotating around its elbow in increments of  $4^\circ$ .

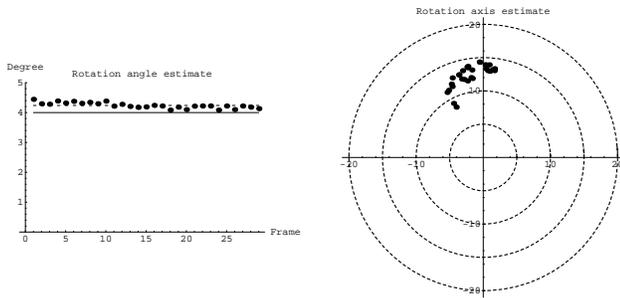


Figure 8. *Puma* sequence. On the left, the magnitude of rotation is shown along with the average angle (dashed line) and true calibration angle (solid line). On the right, the axis of rotation on a *flattened* unit sphere, shown with  $(0^\circ, 0^\circ)$  as the true axis of rotation.

the motion are always at least  $5^\circ$  off around an arbitrary axis for rotation, and at least  $35^\circ$  off for direction of translation. The rotation angle and axis estimates are shown in Figure 8. The rotation axis is estimated with an average of  $13^\circ$  error, while the rotation angle is estimated with an accuracy of  $0.2^\circ$ , which corresponds to a maximum image displacement of around 0.5 pixels. The results for translation are illustrated in Figure 9. When compared with calibration data, it appears that the estimated translations (thick line) are accurate and well within the calibration accuracy.

Since the calibration information is only available for the first 15 frames, the missing information was extrapolated whenever possible without affecting the reliability of the calibration. The fact that this motion analysis method does not require any *a priori* information like feature point correspondence while providing excellent accuracy confirms the usefulness and convenience of the “motion without structure” approach.

### 3.4. Noise sensitivity

The evaluation function for any hypothesized motion does not rely on image gradients, and consists of accumulating large amount of intensity difference information. We therefore expect this measure to be very robust to noise.

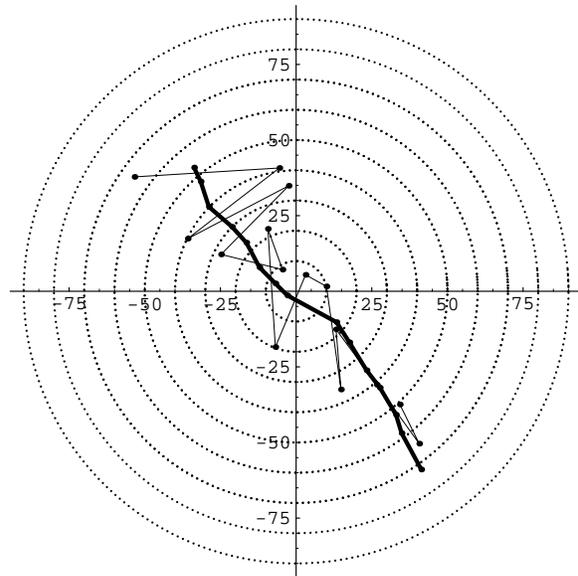


Figure 9. *Puma* sequence. Estimated (thick line) and calibrated (thin line) translations shown on the *flattened* unit sphere.

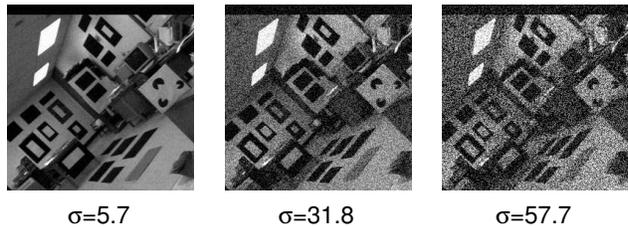
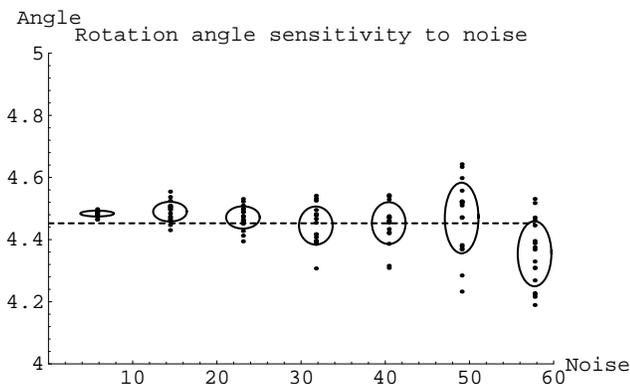


Figure 10. Image degraded by uniform noise. “ $\sigma$ ” is the standard deviation of the noise.

As a simple test for noise sensitivity, we degraded the first two images of the *Puma* sequence using uniform noise in the range  $\pm 10$  up to  $\pm 100$ , which corresponds to standard deviations ranging from 5.7 to 57.7 (see Figure 10). We computed the motion between the two frames 17 times at selected noise levels and observed the distribution of rotation angles recovered. In Figure 11, these angles are shown along with ellipses whose heights are the standard deviations of rotation angles at particular noise levels. These standard deviations range from 0.01 to 0.1 degree. The relationship between the image noise level and the observed rotation angle error is approximately linear, implying that image noise has to double to result in doubling the error on the estimated rotation angle.

These results clearly indicate that the algorithm is very resistant to noise<sup>2</sup>.

<sup>2</sup>This is for uncorrelated noise. For correlated noise (e.g. a



**Figure 11. Rotation angles obtained for different pixel noise levels. The height of a an ellipse gives the standard deviation of angles for a particular noise level.**

## 4. Conclusion

We presented a new paradigm to find the full motion between two frames. We refer to the approach as “motion without structure” because it does not require or compute any information related to the structure of the scene. The motion analysis problem is posed as a search in the space of possible motions and a likelihood measure is developed that evaluates a hypothesized motion based on the sum of sum of squared differences between points in one image and their corresponding epipolar segments in the other.

This likelihood function was shown to exhibit exactly one global minimum for the cases of either known rotation or known translation, provided the images are well-behaved, i.e. that the variance of intensity difference between two points is a monotonically increasing function of their distance apart. In the full motion case, a unique global minimum also exist, but may be subject to the well known ambiguity between rotational and translational motion.

Experimental results suggest that the method is applicable to a wide range of images while achieving very good accuracy and presenting strong robustness to noise. Large frame-to-frame motions can be handled and are only limited by the characteristics of the local intensity variation present in the image.

We believe that the paradigm of motion without structure can provide a robust and accurate algorithm to estimate the ego-motion between two frames. Moreover, we hope that it will prove superior to feature-based and direct or indirect methods of motion-and-structure estimation since neither optical flow, intensity derivatives or feature correspondence are needed.

---

single camera with a dirty lens), the effect on accuracy is likely to be larger.

## References

- [1] Y. Aloimonos and Z. Duric. Estimating the heading direction using normal flow. *Int. J. Computer Vision*, 13(1):33–56, 1994.
- [2] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *Int. J. Computer Vision*, 2(1):43–77, 1994.
- [3] I. J. Cox. A review of statistical data association techniques for motion correspondence. *Int. J. Computer Vision*, 10(1):53–66, 1993.
- [4] I. J. Cox and S. Roy. Direct estimation of rotation from two frames via epipolar search. In *6th Int. conf. on Computer Analysis of Images and Patterns*, 1995.
- [5] I. J. Cox and S. Roy. Statistical modelling of epipolar misalignment. In *International Workshop on Stereoscopic and Three-Dimensional Imaging*, 1995.
- [6] C. Fermuller. Global 3-d motion estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 415–421, 1993.
- [7] B. K. P. Horn and E. J. Weldon, Jr. Direct methods for recovering motion. *Int. J. Computer Vision*, 2:51–76, 1988.
- [8] K. Horn and B. Schunck. Determining optical flow. *Artificial intelligence*, 17:185–203, 1981.
- [9] R. Hummel and V. Sundareshwaran. Motion parameter estimation from global flow field data. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(5):459–476, 1993.
- [10] A. D. Jepson and D. J. Heeger. A fast subspace algorithm for recovering rigid motion. In *Proc. IEEE Workshop on Visual Motion*, pages 124–131, Princeton, NJ, 1991.
- [11] R. Kumar and A. R. Hanson. Robust estimation of camera location and orientation from noisy data having outliers. In *Proc. Workshop on Interpretation of 3D Scenes*, pages 52–60, Austin, TX, USA, 1989.
- [12] S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 9(1):168–176, 1987.
- [13] J. Oliensis. Rigorous bounds for two-frame structure from motion. Technical Report 95-155, NEC Research Institute, Princeton, NJ, 1993.
- [14] D. Sinclair, A. Blake, and D. Murray. Robust estimation of egomotion from normal flow. *Int. J. Computer Vision*, 13(1):57–69, 1994.
- [15] V. Sundareshwaran. Egomotion from global flow field data. In *Proc. IEEE Workshop on Visual Motion*, pages 140–145, Princeton, NJ, 1991.
- [16] C. Tomasi. Pictures and trails: a new framework for the computation of shape and motion from perspective image sequences. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 913–918, 1994.
- [17] C. Tomasi and J. Shi. Direction of heading from image deformations. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 422–427, 1993.