STATISTICAL MODELLING OF EPIPOLAR MISALIGNMENT

Ingemar J. Cox and Sébastien Roy

NEC Research Institute
4 Independence Way
Princeton, NJ 08540, U.S.A.
ingemar|sebastien@research.nj.nec.com

ABSTRACT

We investigate whether epipolar misalignment can be automatically detected and corrected without explicit knowledge of point correspondences. In this regard, the work is closely related to the problem of structure-and-motion from two frames. However, the motion estimation described here is independent of any estimation of the structure of the scene and consequently is expected to be significantly more robust than structure-and-motion algorithms in which the number of unknowns is proportional to the number of pixels in the image. Instead, it may be thought of as forming the basis of a motion-without-structure algorithm, i.e. the solution requires neither knowledge nor estimation of structure or associated properties such as correspondences or flow fields, in order to estimate motion. Of course, structure may be determined by subsequent processing. In particular, we present a method for recovering camera motion for the special cases of (1) known rotation and (2) known translation. The method does not require optical flow fields, feature point correspondences or intensity derivatives. Instead, it relies on a simple statistical characteristic of neighbouring image intensity levels. Specifically, that the variance in intensity between two arbitrary points in an image increases (approximately) monotonically with distance between the two points. Then, it is shown that a simple measure taken across the image can yield a very robust measure of the likelihood of an estimated motion. The likelihood measure allows motion estimation to be cast as an efficient search over the space of possible rotations or translations. The relation between image statistics (textures, etc.) and the accuracy of the estimated motion is discussed and experimental results on real images are presented.

1. INTRODUCTION

Much work has been done on trying to recover camera motion parameters from image pairs. In almost all cases, either optical flow or feature points correspondence are used as the initial measurements. In the first case, some inherent problems (aperture, large motions, etc.) related to optical flow computation, suggests that errors can never be lowered to a negligible level (see [1, 2, 3, 4]). Even methods using the intensity derivatives directly or normal flow, as in [11, 12, 8, 4, 5, 6, 7], suffer from high noise sensitivity. For feature-based methods, the reliable selection and tracking of meaningful feature points is generally very difficult, see [8, 9, 10].

All prior methods of egomotion 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 egomotion and structure, typically in an iterative fashion, refining first the motion and then the structure estimates, etc. Thus, good motion estimation appears to require good structure estimation (or at least point correspondence estimation). In contrast, we propose a paradigm that might be called motion-without-structure that allows the recovery of egomotion independently of any structure or correspondence estimation. The benefit of this is that there are only and exactly five unknown motion parameters to be estimated. As such, we expect that such an approach should be both robust and accurate. Initial experimental results support this.

The algorithm relies on statistically modelling the image behavior in the neighbourhood of a point, as discussed in Section 2. This model is then used to estimate the likelihood of an assumed camera motion. Determining the true motion is then accomplished by searching for the maximum likelihood estimate over the space of translations or rotations. The search is straightforward since we show in Section 3.1 that the function to minimize has only one minimum (which is the solution), provided the image is well behaved, i.e. the variance between neighboring intensity points increases monotonically with the distance between the points. Prior work by the authors [13] 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. However, epipolar histograms are not always a reliable measure of the likelihood of a translational motion. Section 4 presents experimental results from a comprehensive evaluation based on the JISCT stereo database [14].

PUBLISHED IN *INTERNATIONAL WORKSHOP ON STEREOSCOPIC AND THREE DIMENSIONAL IMAGING (IWS3DI'95), SANTORINI, GRECE, (P. 115-121)*

Sébastien Roy is visiting from Université de Montréal, Département d'informatique et de recherche opérationnelle, C.P. 6128, Succ. Centre-Ville, Montréal (Québec), Canada, H3C 3J7

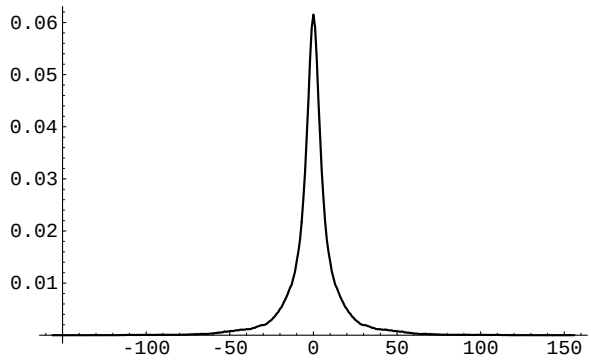

Figure 1: Intensity distribution for a chosen horizontal distance $\vec{\delta} = (4, 0)$.

2. 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 neighbourhood of a given point $\vec{p}$, in a single image A. We assume that the probability of a point $I_A(\vec{p}+\vec{\delta})$ having intensity a conditioned on a given point $\vec{p}$ with intensity b has a Normal distribution, assuming the distance between the two points is sufficiently small. Thus we have

$$P(I_A(\vec{p}+\vec{\delta}) = a \mid I_A(\vec{p}) = b) \quad = \quad G_{[b;\sigma^2(\vec{\delta})]}(a) \quad = \quad \frac{1}{\sqrt{2\pi\sigma^2(\vec{\delta})}} e^{-(a-b)^2/2\sigma^2(\vec{\delta})} \tag{1}$$

where $G_{[b;\sigma^2(\vec{\delta})]}(x)$ is a Gaussian distribution with mean b and variance $\sigma^2(\vec{\delta})$. The variance $\sigma^2(\vec{\delta})$ is a function of the distance $\|\vec{\delta}\|$. This property is intuitively related to the correlation present in a scene and is experimentally verified next.

For a given image, we can evaluate the parameters of the distributions, namely $\sigma^2(\vec{\delta})$, for all possible separations $\vec{\delta}$ within a selected neighbourhood. For a given $\vec{\delta}$, we wish to evaluate the distribution of the samples

$$s_i(\vec{\delta}) = I_A(\vec{p}_i + \vec{\delta}) - I_A(\vec{p}_i) \ , \quad 1 \leq i \leq n$$

taken over all $\vec{p}_i$ points in the image. Note that the mean of this sample is always 0. The variance $\sigma^2(\vec{\delta})$ is obtained from the samples as

$$\sigma^2(\vec{d}) = \frac{1}{n-1} \sum_n s_i(\vec{\delta})^2 = \frac{1}{n-1} \sum_n I_A(\vec{p}_i + \vec{\delta}) - I_A(\vec{p}_i) \tag{2}$$

where n is the number of samples taken.

In order to determine the validity of the Gaussian assumption. We calculated these statistics for a variety of images. Figure (1) shows the distribution of intensities a fixed horizontal distance, $\vec{\delta} = (4, 0)$, from an arbitrary image point. It is evident that the Gaussian model of Equation (1) is a good approximation to the experimental curve of Figure (1).

Once the variance is estimated for all $\vec{\delta}$ such that $\|\vec{\delta}\| \leq r_{max}$ where r_{max} is the maximum size of the neighbourhood, we have a useful global statistic that describes the local behavior of image intensities. This statistic is experimentally determined by directly measuring the distribution of intensity values in the neighbourhood of all pixels in an image. For the images shown in Figure 2, the variance of the distributions are shown in Figure 3 for a neighbourhood of 50 pixels around the reference point. The darker a point, the smaller the variance. The mean of the distributions is not shown here since it is always very close to the predicted value (the value of the reference pixel). Figure 3 indicates that the variance increases 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 3. Note, also, that while the relationship between variance and distance is monotonically increasing, it is not isometric, indicating that intensities are more correlated in certain directions, as expected. For example, the "Parking meter" of Figure 2A is clearly more correlated in the vertical direction and this is evident in Figure 3A in which the variance increases more slowly with distance in the vertical direction.

Our experimental observations indicate that most natural images are well behaved. Only images featuring highly correlated textures or that are highly non-stationary generally present badly-behaved variance functions (non-monotonic, multiple minima). By examining how well behaved the variance function is, it should be possible to measure how accurate the method is.

3. EVALUATING ALIGNMENT

We propose to determine the translation or rotation between two frames via an efficient search. If the rotation is known, then it is necessary to evaluate the likelihood of an assumed translation T, and vice versa for rotation R. For a given point

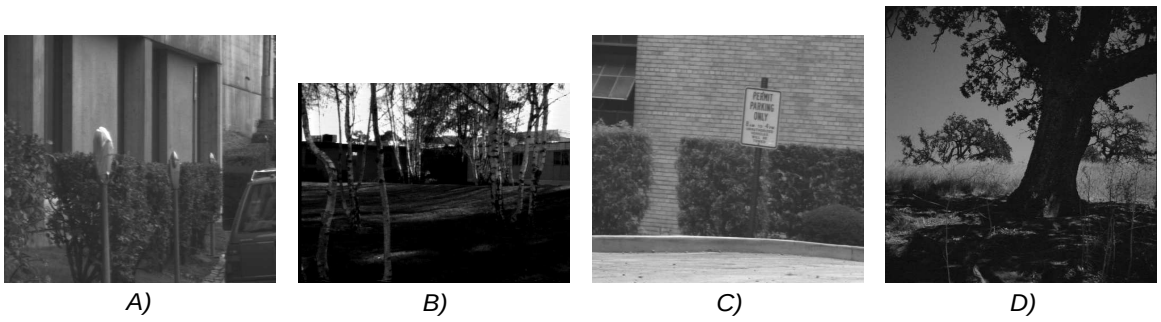

Figure 2: Four images from the JISCT database. A) `parking meter`, B) `birch`, C) `shrub`, D) `tree`.

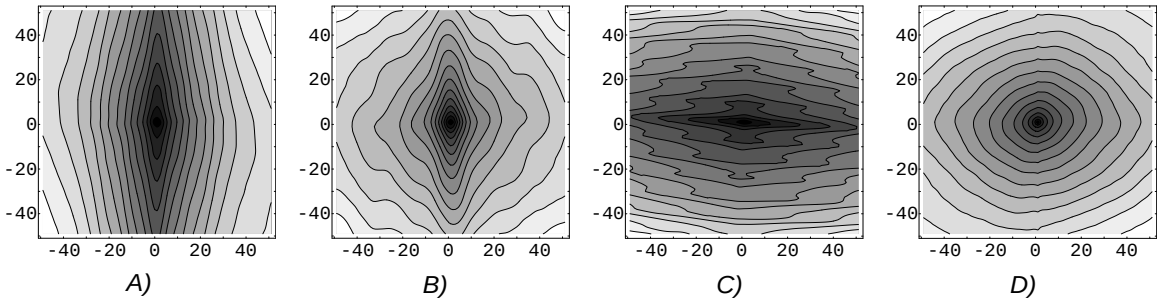

Figure 3: Variance functions $\sigma^2(\vec{\delta})$ for the images A) `parking meter`, B) `birch`, C) `shrub`, D) `tree`. Distances along the axis are in pixels. Darker points have smaller variance.

$I_A(\vec{p})$ in image A and a camera motion, we can compute the corresponding point $I_B(\vec{p}_\infty)$ (the *zero-disparity* point) in image B that has infinite depth, as well as the *focus of expansion* (FOE), see Figure 4. A known translation but unknown rotation implies that the FOE is known but the point $I_B(\vec{p}_\infty)$ has unknown location. Conversely, a known rotation but unknown translation implies that the corresponding point $I_B(\vec{p}_\infty)$ in image B is known but the location of the FOE is not. Since we do not know the real depth of point $I_A(\vec{p})$, we can only assume that the actual corresponding point $I_B(\vec{p}_z)$ is somewhere in the neighbourhood of point $I_B(\vec{p}_\infty)$, depending on the unknown depth z. In fact, it is always located on the line joining the true $I_B(\vec{p}_\infty)$ and the true focus of expansion. Since the points $I_A(\vec{p})$ and (the unknown) $I_B(\vec{p}_z)$ correspond, the variance function around $I_B(\vec{p}_z)$ should be identical to that of $I_A(\vec{p})$.

For the case of unknown translation, 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 candidate FOE. The value of r_{max} is chosen to reflect the maximum disparity expected. A candidate FOE provides a candidate translation and vice versa. If we select a number of sample intensity values u_i along the segment u and define the error measure e_u as

$$e_u = \sum_{i=1}^{n} (u_i - I_A(\vec{p}))^2 \tag{3}$$

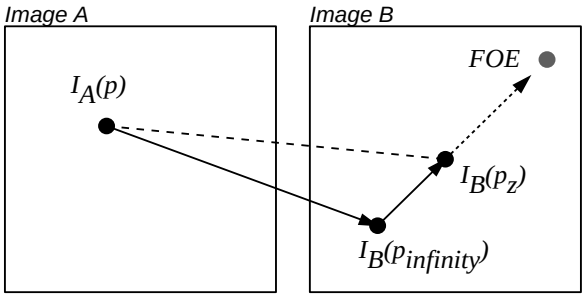

Figure 4: Basic geometry for known rotation. For a given $I_A(\vec{p})$, its unknown corresponding point $I_B(\vec{p}_z)$ is on the line joining $I_B(\vec{p}_\infty)$ and the FOE.

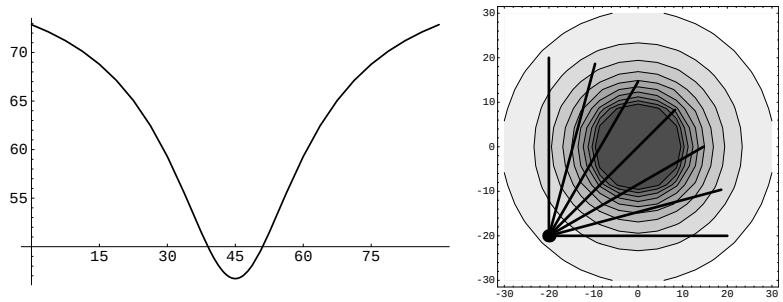

Figure 5: Analytic error function for a typical variance function.

then e_u will be a minimum when the segment u contains $I_B(\vec{p}_z)$, and thus points towards the FOE. An example of the error function e_u is shown in Fig. 5. For a typical variance function defined as

$$\sigma^2(\vec{\delta}) = \frac{2\|\delta\|^2}{\|\delta\|^2 + 50}$$

an analytic error curve is computed for segments over an interval of 0° to 90° and shows a single minimum at 45°, the angle at which the line segment is correctly oriented towards the true FOE. This minimum exists and is unique when the variance function of the images is well behaved. Section 3.1 discusses this point in detail. We can now use this property to estimate if a candidate FOE is good. If we select a number of points $I_A(\vec{p}_i)$ and compute the sum of the individual line segment error measures e_{q_i} where q_i is the segment starting at $I_A(\vec{p}_i)$ and pointing toward the candidate FOE, we expect all these error measures to be simultaneously a minimum if this candidate FOE is indeed the true FOE. We thus use the sum of the individual line segment error measures as a global estimate of the likelihood of the FOE. In the case of well behaved images (see below) we expect only one minimum and can do a simple search for the exact FOE based on gradient descent.

It is easy to change this method to estimate rotation by fixing the FOE (known translation) and selecting candidate points $I_B(\vec{p}_\infty)$ associated with candidate rotations.

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 3 is observed when a segment u contains $I_B(\vec{p}_z)$ and joins the true zero-disparity point and the true FOE. We define a *well behaved* image as one that possesses a monotonically increasing variance function. Since by definition this function always has a global minimum at $(0,0)$, this condition is enough to insure that the likelihood function possesses a unique minimum. This is demonstrated next.

Consider a segment u in the neighbourhood of $\vec{p}_z$, starting at $\vec{p}_\infty$, and containing n sample intensities as depicted in Figure 6. Then from the distribution property we can say 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 $\vec{d}_{u_i}$ 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 3, the error measure e_u is a random variable defined as

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

with an expectation value defined as

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

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 as segment u. The error measure e_v is defined as the random variable

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

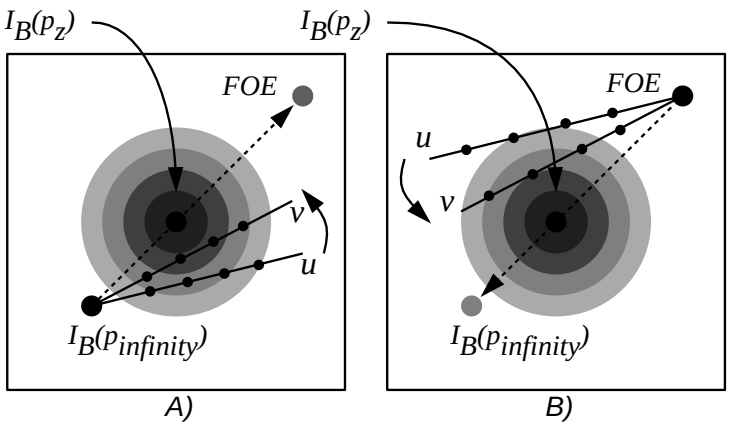

Figure 6: Error function for two segments u and v. When v is closer to $\vec{p}_z$ then u, its expectation is smaller for a well behaved variance function. A) Unknown translation. B) Unknown rotation.

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 then $E(e_u)$. First, it is straightforward to see that

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

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 increasing with $\|\vec{d}\|$ from $\vec{p}_z$. From these two observations, we can immediately conclude that

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

It then follows that

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

which shows that as we get closer to the segment containing $I_B(\vec{p}_z)$, the expected error value gets smaller until it reach a minimum when the candidate FOE correspond to the true FOE. As long as the variance function is monotonic, this minimum is guaranteed to exist and is unique.

The same procedure is applied for rotation estimation, just reversing the FOE and the zero-disparity point.

4. EXPERIMENTAL RESULTS

An number of experiments where conducted on natural images, for different ranges of camera translation and rotation. For translation estimation, Figure 7 shows the error functions obtained for the images of Figure 2. The likelihood is shown for various angles[1] ($\pm 45°$) around an arbitrary translation which, in this case, is pure horizontal displacement. In the four cases, the minimum should be located in the center at (0°,0°). At this point, we observe an irregularity which is an artifact of bi-cubic intensity interpolation. The error in the location of the likelihood minimium is between 1 and 3 quantization units, corresponding to 2.25 to 6.75 degrees of accuracy. These results compare favorably with other methods [4, 8] which give a FOE localization error of around 9 degrees. Moreover, it is believed that these results would be improved if a finer quantization search had been performed.

For rotation estimation, Fig. 8 shows the error functions for a range of $\pm 45°$ around three differents axis (X, Y, and Z). Here, the minimum should and is observed to be at 0°. It should be noted that for large rotation around the X or Y axis, the likelihood function becomes noisy because of the small overlap between the images.

For all these results, around 4% of the points of the images are randomly selected to yield between 2500 and 3000 (9700 points for the `tree` image) line segments for likelihood estimation. Up to 25 samples are taken along each segment and used

[1]Since translation is only known up to a scale factor, it is represented as a unit vector on a sphere, which can be characterized by two angles. A purely horizontal motion is represented by $(0°, 0°)$, a purely vertical motion is denoted by $(0°, 90°)$ and a motion along the optical axis by $(90°, 0°)$.

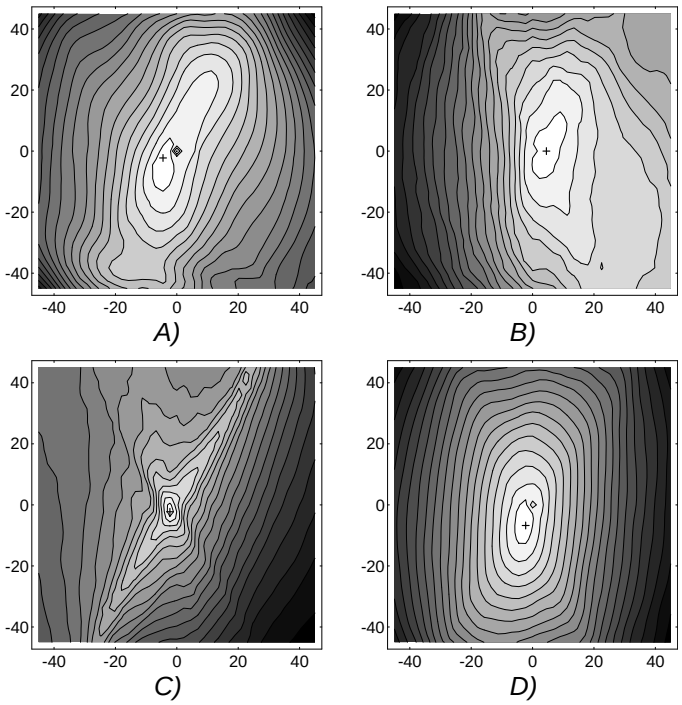

Figure 7: Translation error functions for images A) `parking meter`, B) `birch`, C) `shrub`, D) `tree`. The position of the FOE should be at (0°,0°). Ligher points show smaller error. The axis represents rotation in degrees from reference translation $(1, 0, 0)$, see footnote. The cross denotes the observed minimum of the likelihood function.

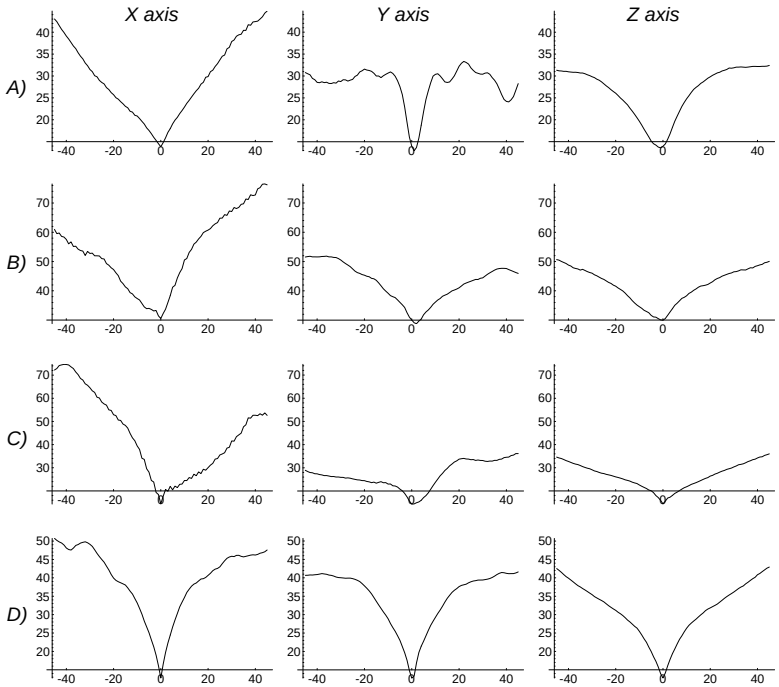

Figure 8: Rotation error functions for images A) `parking meter`, B) `birch`, C) `shrub`, D) `tree`. Rotation likelihood on a range of ±45° around the X, Y, and Z axis are presented. The true rotations are located at 0°.

in Equation 2 to compute the likelihood. For most images, only a few hundred points are needed to generate useful results that can be used to quickly find a good estimate. By increasing the number of points used, the accuracy of the estimation is also increased.

5. CONCLUSION

We described a new method to find either the translational or rotational motion between two frames assuming the other component of motion is known. The problem is posed as a search for the most likely motion, and as such, a likelihood measure is required to evaluate each candidate motion. A likelihood measure was derived based on the sum of squared distance (SSD) between a point $\vec{p}$ in image A and a series of points in image B that lies on a line joining $\vec{p}$ with the candidate FOE. It was shown that this likelihood function has a clearly defined minimum which is easily located by gradient descent provided the two images are well-behaved, i.e. that the variance in intensity between two points monotonically increases with their distance apart.

Experimental results on the SRI JISCT stereo database support the monotonic variance assumption in almost all cases. The likelihood function was also shown to be well behaved with a clearly defined global minimum over all translations/rotations. Translation estimates were within 2.25 to 6.72 degrees of their true values and rotation estimates were correct within the limits of the quantization error, indicating that very accurate estimation may be possible. A significant portion of the translational error is expected to be due to the coarse quantization, 2.25°, of the search. More work is, however, needed to evaluate the accuracy of the method over a wide class of scenes.

Currently, we have restricted the search to either a three dimensional search for rotation or a two dimensional search for translation. In principle, a full five dimensional search for all components of motion is possible. Preliminary results from such a process suggest that the translational estimation can be decoupled from the rotational estimation and that very accurate motion estimates can be determined by an iterative process that first estimates rotation, then uses this estimate in the determination of the translation, etc. Only a small number of iterations appears necessary. We believe that the paradigm of motion-without-structure can provide a robust and accurate algorithm to estimate the ego-motion between two frames. It is our hope that this paradigm will prove superior to feature-based and indirect and direct methods of shape-and-motion estimation since neither optical flow, intensity derivatives or feature correspondence are needed.

6. REFERENCES

[1] K. Horn and B. Schunck. Determining optical flow. *Artificial intelligence*, 17:185–203, 1981.

[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] 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.

[4] V. Sundareswaran. Egomotion from global flow field data. In *Proc. IEEE Workshop on Visual Motion*, pages 140–145, Princeton, NJ, 1991.

[5] C. Fermuller. Global 3-d motion estimation. In *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pages 415–421, 1993.

[6] D. Sinclair, A. Blake, and D. Murray. Robust estimation of egomotion from normal flow. *Int. J. Computer Vision*, 13(1):57–69, 1994.

[7] Y. Aloimonos and Z. Duric. Estimating the heading direction using normal flow. *Int. J. Computer Vision*, 13(1):33–56, 1994.

[8] 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.

[9] I. J. Cox. A review of statistical data association techniques for motion correspondence. *Int. J. Computer Vision*, 10(1):53–66, 1993.

[10] 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.

[11] B. K. P. Horn and E. J. Weldon, Jr. Direct methods for recovering motion. *Int. J. Computer Vision*, 2:51–76, 1988.

[12] S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *IEEE PAMI*, 9(1):168–176, 1987.

[13] 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.

[14] R. C. Bolles, H. H. Baker, and M. J. Hannah. The JISCT stereo evaluation. In *Proc. of DARPA Image Understanding Workshop*, pages 263–274, 1993.