Rigidity And Three-Dimensional Motion From Binocular Image Motion

Niels da Vitoria Lobo
Department of Computer Science,
University of Toronto, Toronto, Canada M5S 1A4
e-mail: niels@ai.toronto.edu

Abstract

We describe how using binocular motion, knowing the binocular disparities of two scene points and their projected velocities in the left and right images yields a test for rigidity. When the test is positive, the complete set of 3-D motion parameters can be obtained, if there is a unique solution. The assumptions used in the model are minimal and standard.

Keywords: Rigidity, motion parameters, binocular image motion

1 Introduction

The ability for a vision system to know if objects in its scene are rigid or not would be useful in many contexts. First, rigidity could help with object segmentation both at the level of separating moving objects from their background, and that of segmentation into parts, when there is jointed rigidity. Second, rigidity is useful as a means of classifying objects in scene recognition. Different degrees of non-rigidity exist (Goldgof *et al.*, 1988) and are generally associated with different physical materials. Third, at an earlier stage in visual processing, any shape from motion computation needs to know where the bounds are to the rigidity assumption that is often made in recovering shape. Much vision research to date has concentrated on analysing a rigid world, and the issues involved in analysing non-rigidity are not well-defined. However, without doubt, rigidity will prove useful as a property worth measuring.

The concept of rigidity is integrally tied to that of motion. Only by moving (and/or deforming) do surfaces provide glimpses of their rigidity properties. Hence, it is not too surprising that a test for rigidity is associated with an algorithm for computing the 3-D motion between an observer and a surface. The central issues in such an endeavour are how much measurement from the image is needed as input, whether or not such input would be available easily, what assumptions are needed for the scheme to work, how reliable and robust the results are in the presence of noise, etc. In this paper we address some of these issues.

We describe a scheme that given stereo disparity for 2 scene points appearing in the left and right image, and given the instantaneous image velocity measurements in the left and right images, will test if the points are moving in a rigid fashion relative to the stereoscopic observer. If so, the scheme will compute the rotation and translation parameters of motion. We will report on some experiments performed on synthetic data. Then, we will argue that the inputs needed by this system are in the realm of possibility, while noting that the assumptions made are typical and realistic.

2 Binocular Motion

In this paper we will use the 3-D motion notation of Longuet-Higgins and Prazdny (1980), and Bruss and Horn (1983). In work that examined the relationship between stereo imaging and 3-D motion Waxman and Duncan (1986) present the 3-D motion parameters for two cameras in a moving parallel stereo configuration with a fixed baseline separation b, as shown in Figure 1. They note that when relative motion between the object and observer is ascribed to the observer, the left and right cameras are in motion with respect to each other. So, suppose that relative to a scene point the instantaneous rigid body motion for the left camera is given by the translation (U, V, W) and the rotation (A, B, C). Then, for the right camera the motion parameters relative to the same scene point are $(U, V - bC, W + bB)$ for translation and (A, B, C) for rotation. That is, the translation component of the right camera is influenced by the left camera's rotation (see Waxman and Duncan, 1986). Similar relationships can be derived for arbitrary configurations between the two cameras.

Consider a scene point P at (X, Y, Z) in the left camera's 3-D coordinate system, projecting to p_l at (x_l, y_l) in the left image and to p_r at (x_r, y_r) in the right image. Suppose the image velocities resulting from relative motion of the scene point and the cameras are denoted by the quantities (u_l, v_l) for the left camera, and by (u_r, v_r) for the right camera. Then these quantities relate to the

3-D motion parameters as follows (Longuet-Higgins and Prazdny 1980, and using the appropriate terms for the right camera's translation, as above),

$$\begin{aligned} u_l &= \frac{-U+x_l W}{Z} - A x_l y_l + B(x_l^2+1) - C y_l \\ v_l &= \frac{-V+y_l W}{Z} - A(y_l^2+1) + B x_l y_l + C x_l \\ u_r &= \frac{-U+x_r(W+bB)}{Z} - A x_r y_r + B(x_r^2+1) - C y_r \\ v_r &= \frac{-(V-bC)+y_r(W+bB)}{Z} - A(y_r^2+1) + B x_r y_r + C x_r \end{aligned} \qquad (1)$$

Now, due to the parallel stereo configuration, a scene point will project onto the two images along a horizontal epipolar line. Thus, the left and right image vertical positions will be the same for a given scene point, i.e., $y_r = y_l$; the difference in horizontal positions is the binocular disparity δ at (x_l, y_l) in the left image, i.e., δ is given by $x_r - x_l$, which in turn is equal to $-b/Z$. Thus, $1/Z = -\delta/b$. So the four equations in (1) can be re-written as a function of disparity (b is assumed known). Because of the epipolar constraint, (i.e., $y_r = y_l$), the second and fourth equations in the set above are identical, thus providing us with three independent equations.

Next, suppose we have *two* scene points P_1 and P_2, for which we can perform stereo correspondence to obtain their image disparities δ_1 and δ_2, and whose instantaneous image velocities in the left and right images can be measured. Then, assuming P_1 and P_2 are in a rigid relationship together, P_1 and P_2 will each provide three equations similar to those in equations (1), thus giving six independent linear equations in the six unknown motion parameters. These equations can be solved to calculate the six parameters.

3 Motion Parameters and Rigidity

For the linear system derived above, of the form $\mathbf{A}x = b$, three outcomes can occur. There could be a unique solution, no solution or infinitely many solutions.

A unique solution can occur if and only if the coefficient matrix $\mathbf{A}$ is non-singular, and this would necessitate a rigid interpretation. On the other hand, a singular system could arise out of two situations, either the equations do not offer enough constraint to provide a unique solution *or* the equations are inconsistent. Lack of sufficient constraint is due to redundancy in the measurements, whereas inconsistency is a consequence of either the two scene points moving independently of each other (i.e., non-rigidly), or the measurements (the vector b which holds the image velocities) being noisy.

There is an easy way to distinguish between the cases of redundancy and inconsistency. Algebraically, this is done by checking whether the vector b lies in the Range of the coefficient matrix $\mathbf{A}$, or whether it lies in $\mathbf{A}$'s Null space (Stoer and Burlirsch 1980). This can be done by computing the pseudo-inverse (Moore-Penrose Inverse) and using it to project the b onto the the Null space. If the projection leads to a scalar value of zero, we would know that the system is not inconsistent, i.e., that rigid interpretations exist. Otherwise, the system would have to be interpreted as inconsistent, i.e., either noisy or non-rigid. Given the limited information being used in this scheme (namely measurements from only two points), it is unclear whether the scheme can have the capability of distinguishing between the two inconsistent situations.

4 Experimental Results

We ran experiments with synthetic data, in which we could control for different conditions, including noise in the image velocities and percent difference in the three-dimensional motion of the two points. For these experiments, instead of computing the projections using the pseudo-inverse in the manner described above, we just treated the system as a linear least-squares problem and examined the residuals. This is a rather course but encouraging demonstration that the approach espoused here is promising. When the residuals were high, the system was deemed inconsistent, while when they were low, as having a rigid interpretation. For all the experiments, the measure of error-of-fit used was the L_1 norm of the vector of residuals.

The experiments deal with two points moving in the scene with rotation and translation. In the first set of experiments, both points moved in the scene with the same motion parameters and the input measurements were noise-free. The resulting measures of residuals in this set of experiments were very small, on the order of round-off error.

In the next set of experiments we introduced gaussian noise centered around various means into the measured image velocities and recomputed the residuals. The error-of-fit rose with the mean of the gaussian noise.

Similarly, when we let the parameters of motion of the second point differ from the motion of the first point, the error-of-fit rose with the mean difference in the parameters. Thus a threshold would need to be selected to distinguish cases. At this point we have not come up with a principled way of threshold selection. Similar results were obtained when both noise and non-rigidity were introduced.

Finally, we have begun to consider some of the issues involved in grouping rigid pairs of points into larger collections. For instance consider the case of jointed motion that can be illustrated by imagining the hands on a clock-face. If the clock moves as the hands rotate, then consider how the set of three points made up of the two end-points and the common hinged point move. The pair of end points for each hand of the clock will be seen as rigid, and a naive grouping algorithm may apply a transitive law (to the common hinged point) to wrongly group all three points into one rigid object. To avoid this, the rigidity test might have to be applied to different scales of distances (since it would be too complex to apply it to all pairings of points).

5 Discussion and Conclusion

We have described how using binocular motion, knowing the binocular disparities of two scene points and their projected velocities in the left and right images yields a test for rigidity. When the test is positive, the complete set of 3-D motion parameters is obtained. The assumptions used in the model are minimal and standard. What is unclear, though, is whether the inputs can be easily computed and what demands the system makes on their accuracy. While it is true that reliable estimates of binocular disparity are now within reach (Jepson and Jenkin 1989), the same is untrue for image velocity. The *normal* component of image velocity is calculable to a high degree of accuracy (Fleet 1990), but not so the *complete* image velocity which our scheme needs.

In fact, attempts to compute complete image velocity from normal velocity estimates, need to assume rigidity of the underlying scene primitives. But this is obviously a cyclic problem. Hence, it may be that this dilema will need an iterative interleaved computation of rigidity and complete image velocity, both influencing each other. Lastly, issues of reliability, confidence, and robustness remain. We are pursuing some of the above issues.

6 Acknowledgements

The author is grateful to Gregory Dudek, David Fleet, Tom Fairgrieve and David Wilkes for helpful discussions. This work was supported by the Natural Sciences and Engineering Research Council and through the Information Technology Research Center, a Province of Ontario Center of Excellence.

References

Bruss A.R. and Horn B.K.P. (1983), "Passive Navigation," *CVGIP*, **21**, 3-20.

Fleet D.J. (1990), *Ph.D. Thesis*, Dept. of Comp. Science, Univ. of Toronto. To appear.

Goldgof D.B., Hua L., Huang T.S. (1988), "Motion Analysis of Nonrigid Surfaces," *Proc. Computer Vision and Pattern Recog.*, Ann Arbor, Michigan, 375-380.

Jepson A.D. and Jenkin M.R.M. (1989), "The Fast Computation of Disparity from Phase Differences," *Proc. Computer Vision and Pattern Recog.*, San Diego, California, 398-403.

Longuet-Higgins H.C. and Prazdny K. (1980), "The interpretation of a moving retinal image," *Proc. Roy. Soc. Lond.*, **B 208**, 385-397.

Stoer J. and Bulirsch R. (1980), *Introduction to Numerical Analysis*, Springer-Verlag, London.

Waxman A.M. and Duncan J.H. (1986), "Binocular Image Flows: Steps Towards Stereo-Motion Fusion," *IEEE P.A.M.I.*, **8**, No. 6, 715-729.

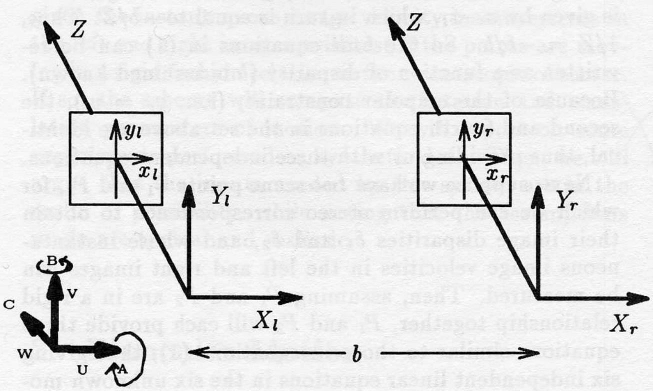

Figure 1: A parallel stereo configuration (left and right imaging systems) capable of moving as a rigid object. The images are produced under perspective projection, at a focal distance of unity. Baseline separation is b.