MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS 1963 A

AD-A179 409

Determining the Translation of a Rigidly Moving Surface, without Correspondence

John Aloimonos, Computer Science Department
Anup Basu, Department of Statistics
The University of Rochester
Rochester, NY 14627

D
ELI
APR

DTIC
ELECTE
APR 2 3 1987
S D
D

rochester

Department of Computer Science
University of Rochester
Rochester, New York 14627

87 4 21 266

Determining the Translation of a Rigidly
Moving Surface, without Correspondence

John Aloimonos, Computer Science Department
Anup Basu, Department of Statistics
The University of Rochester
Rochester, NY 14627

DTIC
SELECTED
APR 2 3 1987
D

Abstract

A method is presented for the recovery of the three-dimensional translation of a rigidly moving textured object. The novelty of the method consists of the fact that four cameras are used in order to avoid the solution of the correspondence problem. The method seems to be immune to small noise percentages and to have good behavior when the noise increases.

----

1. Introduction

An important problem in computer vision is to recover the three-dimensional motion of a moving object from its images. Up to now, there have been three approaches towards the solution of this problem:

1) The first assumes the dynamic image to be a three-dimensional function of two spatial arguments and a temporal argument. Then, if this function is locally well-behaved and its spatiotemporal gradients are computable, the image velocity or optical flow may be computed [1, 2, 3, 5, 8].

2) The second method for measuring image motion considers the cases where the motion is "large" and the previous technique is not applicable. In these instances the measurement technique relies upon isolating and tracking highlights or feature points in the image through time. In other words, operators are applied on both dynamic frames which output a set of points in both images, and then the correspondence problem between these two sets of points has to be solved (i.e., finding which points on both dynamic frames are due to the projection of the same world point) [9, 39, 40, 4].

In both the above approaches, after the optical flow field or the discrete displacements field (which can be sparse) are computed, then algorithms are constructed for the determination of the three-dimensional motion, based on the optic flow or discrete displacement values [6, 10, 11, 12, 13, 15, 16, 20, 21, 22, 23, 24, 25, 26, 28, 30, 31, 32, 33, 34, 35, 41, 42].

3) The three-dimensional motion parameters are computed directly from the spatial and temporal derivatives of the image intensity function. In other words, if f is the intensity function and (u, v) the optical flow at a point, then the equation $f_x u + f_y v + f_t = 0$ holds approximately. All the methods in this category are based on substitution of the optical flow values in terms of the three-dimensional motion parameters in the above equation, and there is very good work in this direction [36, 37, 17].

1

As the problem has been formulated over the years, one camera is used, and so the three-dimensional motion parameters that have to be computed, and can be computed, are five (two for the direction of translation and three for the rotation). In our approach, four cameras are used to recover the three translation parameters, instead of the direction only of the translation, and despite the fact that our theory assumes that the object in view is only translating, our results (i.e., the three-dimensional translation) are affected very little even if the object is moving with a small rotation, in addition to a translation.

2. Motivation and Previous Work

The basic motivation for this research is the fact that optical flow (or discrete displacement) fields produced from real images by existing techniques are corrupted by noise and are partially incorrect [7]. Most of the algorithms in the literature that use the retinal motion field to recover three-dimensional motion fail when the input (retinal motion) is noisy. Some algorithms work reasonably for images in a specific domain.

Some researchers [23, 31, 32, 41, 13, 33] developed sets of nonlinear equations with the three-dimensional motion parameters as unknowns, which are solved by iterations and initial guessing. These methods are very sensitive to noise, as it is reported in [23, 31, 13, 33]. On the other hand, other researchers [26, 42] developed methods that do not require the solution of nonlinear systems, but the solution of linear ones. Despite that, under the presence of noise, the results are not satisfactory [26, 42].

Bruss and Horn [12] presented a least-squares formalism that tried to compute the motion parameters by minimizing a measure of the difference between the input optic flow and the predicted one from the motion parameters. The method, in the general case, results in solving a system of nonlinear equations with all the inherent difficulties in such a task, and it seems to have good behavior with respect to noise only when the noise in the optical flow field has a particular distribution. Prazdny, Rieger, and Lawton presented methods based on the separation of the optical flow field in its translational and rotational components, under different assumptions [21, 22]. But difficulties are

2

reported with the approach of Prazdny in the present of noise [34], while the methods of Rieger and Lawton require the presence of occluding boundaries in the scene, something which cannot be guaranteed. Finally, Ullman in his pioneering work [6] presented a local analysis, but his approach seems to be sensitive to noise, because of its local nature.

Several other authors [20, 30] use the optical flow field and its first and second spatial derivatives at corresponding points to obtain the motion parameters. But these derivatives seem to be unreliable with noise, and there is no known algorithm which can determine them reasonably in real images. Others [10] follow an approach based partially on local interpretation of the flow field, but it can be proved [27] that any local interpretation of the flow field is unstable.

At this point it is worth noting that all the aforementioned methods assume an unrestricted motion (translation and rotation). In the case of restricted motion (only translation), a robust algorithm has been reported by Lawton [35], which was successfully applied to some real images. His method is based on a global sampling of an error measure that corresponds to the potential position of the focus of expansion (FOE); finally, a local search is required to determine the exact location of the minimum value. However, the method is time-consuming, and is likely to be very sensitive to small rotations. Also the inherent problems of correspondence, in the sense that there may be drop-ins or drop-outs in the two dynamic frames, is not taken into account. All in all, most of the methods presented up to now for the computation of three-dimensional motion depend on the value of flow or retinal displacements. Probably there is no algorithm until now that can compute retinal motion reasonably (for example, 10% accuracy) in real images.

Even if we had some way, however, to compute retinal motion in a reasonable (acceptable) fashion, i.e., with at most an error of 10%, for example, all the algorithms proposed to date that use retinal motion as input would still produce non-robust results. It seems that the reason for this is the fact that the motion constraint (i.e., the relation between three-dimensional motion and retinal displacements) is very sensitive to small perturbations. Table 1 shows how the error of motion parameters grows as the error in image point correspondence increases when 8-point correspondence is used, and Table 2

3

shows the same relationship when 20-point correspondence is used with 2.5% error on point correspondences based on a recent algorithm of great mathematical elegance. (Tables 1 and 2 are from [26].)

Table 1: Error of motion parameters for 8-point correspondence
for 2.5% error in point correspondence.

Error of E (essential parameters)	73.91%
Error of rotation parameters	38.70%
Error of translations	103.60%

Table 2: Error of motion parameters for 20-point correspondence
for 2.5% error in point correspondence.

Error of E (essential parameters)	19.49%
Error of rotation parameters	2.40%
Error of translations	29.66%

It is clear from the above tables that the sensitivity of the algorithm in [26] to small errors is very high. It is worth noting at this point that the algorithm in [26] is solving linear equations, but the sensitivity to error in point correspondences is not improved with respect to algorithms that solve non-linear equations. Finally, the third approach, which computes directly the motion parameters from the spatiotemporal derivatives of the image intensity function, gets rid of the correspondence problem and seems very promising. In [17, 36, 15], the behavior with respect to noise is not discussed. But extensive experiments [38] implementing the algorithms presented in [37] show that noise in the intensity function affects the computed three-dimensional motion parameters a great deal. We should also mention that the constraint $f_x u + f_y v + f_t = 0$ is a very gross approximation of the actual constraint under perspective projection [43]. So, despite the fact that no correspondences are used in this approach, the resulting algorithms seem to have the same sensitivity to small errors in the input as in the previous cases. This fact should not be surprising, because even if we avoid correspondences, the constraint between three-dimensional motion and retinal motion (regardless of whether the retinal

4

motion is expressed as optic flow or the spatiotemporal variation of the image intensity function) will be essentially the same when one camera is used (monocular observer, traditional approach). This constraint cannot change, since it relates three-dimensional motion to two-dimensional motion through projective geometry.

So, as the problem has been formulated (monocular observer), it seems to have a great deal of difficulty. This is again not surprising, and the same problem is encountered in many other problems in computer vision (shape from shading, structure from motion, stereo, etc.). There has recently been an approach to combine information from different sources in order to achieve uniqueness and robustness of low-level visual computations [44]. With regard to the three-dimensional motion parameters determination problem, why not combine motion information with some other kind of information? It is clear that in this case the constraints won't be the same, and there is some hope for robustness in the computed parameters. As this other kind of information that should be combined with motion, we choose stereo.

The need for combining stereo with motion has recently been appreciated by a number of researchers [14, 29, 45, 46]. Jenkin and Tsotsos [14] used stereo information for the computation of retinal motion, and they presented good results for natural images. Waxman et al. [29] presented a promising method for dynamic stereo, which is based on the comparison of image flow fields obtained from cameras in known relative motion, with passive ranging as goal. Whitman Richards [46] is combining stereo disparity with motion in order to recover correct three-dimensional configurations from two-dimensional images (othography-vergence). Finally, Huang and Blostein [45] presented a method for three-dimensional motion estimation that is based on stereo information. In their work, the static stereo problem as well as the three-dimensional matching problem have to be solved before the motion estimation problem. The emphasis is placed on the error analysis, since the amount of noise (in typical image resolutions) in the input of the motion estimation algorithm is very large.

So a natural question arises: is it possible to recover three-dimensional motion from images without having to go through the very difficult correspondence problem? And if such a thing is possible, how immune to noise will the algorithm be? In this paper, we prove that if we combine stereo and

5

motion in some sense and we avoid any static or dynamic correspondence by using four cameras, then we can compute the three-dimensional translation of a moving object. At this point, it is worth noting recent results by Kanatani [18, 19] that deal with finding the three-dimensional motion of planar contours in small motion, without point correspondences. These methods seem to suffer from numerical errors a great deal, but they have a great mathematical elegance. Our experiments show that the computation is very reliable even in the presence of noise, or even when the object in view is not only translating but also rotating with a small rotation. Table 3 shows the average error in the computed translational parameters as the noise in the images increases, using the method developed in this paper, where the noise was randomly generated.

Table 3: Error of Translation Parameters
vs. Noise in Images

Average Error in Images	Approximate Average Error in Translation Parameters
1%	negligible
5%	negligible
10%	5%
20%	5%
30%	6%
50%	8%
67%	15%
75%	20%
90%	unreliable

Later in the paper we will formally define the meaning of noise and measure of the error in the computed parameters.

The organization of this paper is as follows. The next section introduces the reader to some technical prerequisites. Section 4 describes the geometric model and the developed constraints. Section 5 describes the algorithms, and Section 6

presents experiments and the effect of noise in the computation of three-dimensional translation. Finally, Section 7 concludes the work and discusses future research.

3. Technical Prerequisites

Consider a coordinate system $OXYZ$ fixed with respect to the camera, where O is the nodal point of the eye and the image plane is perpendicular to the Z-axis, that is, pointing along the optical axis. Let us represent points on the image plane with small letters (x, y) and points in the world with capital leters (X, Y, Z). Let a point $P = (X, Y, Z)$ in the world have perspective image (x_1, y_1), where $x_1 = fX_1/Z_1$ and $y_1 = fY_1/Z_1$. If the point P moves to $P' = (X_2, Y_2, Z_2)$ with

$$X_2 = X_1 + \Delta X$$

$$Y_2 = Y_1 + \Delta Y$$

$$Z_2 = Z_1 + \Delta Z$$

and P' has the perspective image (x_2, y_2), then it can be easily shown that

$$x_2 - x_1 = \frac{f\Delta X - x_1 \Delta Z}{Z_1 + \Delta Z}$$

$$y_2 - y_1 = \frac{f\Delta Y - y_1 \Delta Z}{Z_1 + \Delta Z}$$

The above equations relate the retinal motion of an image point with the three-dimensional motion of the corresponding world point. We now proceed with the description of the imaging system.

4. The Model

Let $OXYZ$ be a cartesian coordinate system, fixed with the Z-axis pointing along the optical axis, and consider the image plane Im_1 perpendicular to the Z-axis at a point $(0, 0, f)$ (focal length $= f$). This is obviously the model of a camera. The geometry of the system induces a natural cartesian coordinate system on the image plane with the center at the intersection of the Z-axis with the image plane, and the x- and y-axes parallel to the X and Y ones. Furthermore, consider

7

three more cameras with image planes Im_2, Im_3, and Im_4 with nodal points $(dx, 0, 0)$, $(dx, dy, 0)$, and $(0, dy, 0)$, respectively, such that any world point has the same depth with respect to any of the cameras (see Figure 1).
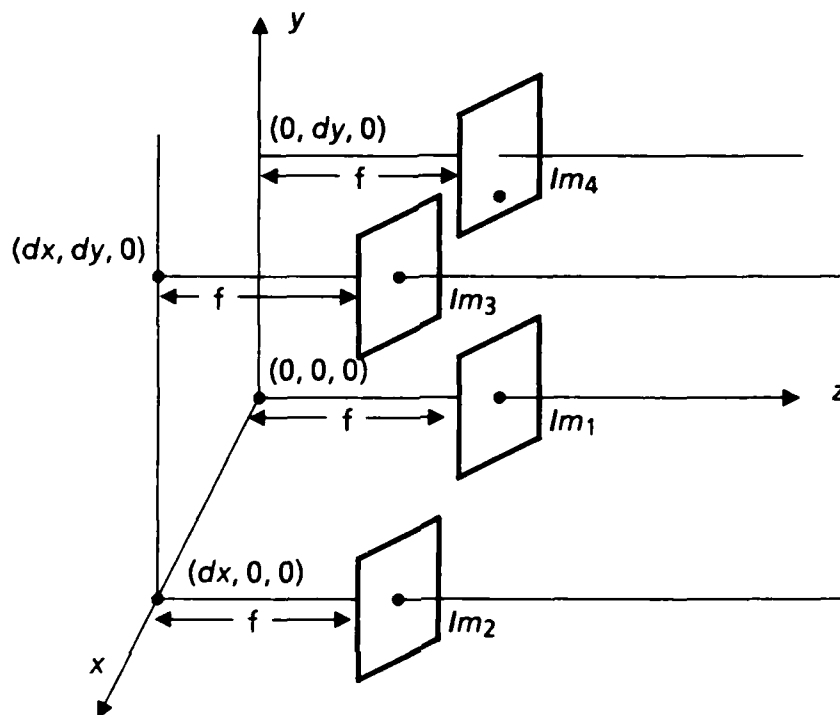

Figure 1: The Imaging (Four-Eye) System

On each one of the image planes a coordinate system is defined exactly as it was done for Im_1. From now on, coordinates of three-dimensional points will be denoted with X, Y, Z, while coordinates of points in each of the images will be denoted by (x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), respectively. Coordinates of image points in the second dynamic frame (i.e., projections of three-dimensional points after the motion) will be denoted by the same symbols as before the motion, but primed (i.e., (x_1', y_1'), etc.). Consider a set $A = \{(X_i, Y_i, Z_i): i = 1, ..., n\}$ of points in the world, which translates rigidly along the vector $(\Delta X, \Delta Y, \Delta Z)$ to form a new set $A' = \{(X_i', Y_i', Z_i'): i = 1, ..., n\}$, where $X_i' = X_i + \Delta X$, $Y_i' = Y_i + \Delta Y$, $Z_i' = Z_i + \Delta Z$, $i = 1, ..., n$. From the projections of the sets A and A' on the four cameras we wish to recover the quantities ΔX, ΔY, ΔZ without using any static or dynamic correspondence.

8

Let the projections of the set A on the four image planes be $\{(x_{1i}, y_{1i}), i = 1, ..., n\}$, $\{(x_{2i}, y_{2i}), i = 1, ..., n\}$, $\{(x_{3i}, y_{3i}), i = 1, ..., n\}$, and $\{(x_{4i}, y_{4i}), i = 1, ..., n\}$, respectively, and the projections of the set A' be $\{(x_{1i}', y_{1i}'), i = 1, ..., n\}$, $\{(x_{2i}', y_{2i}'), i = 1, ..., n\}$, $\{(x_{3i}', y_{3i}'), i = 1, ..., n\}$, and $\{(x_{4i}', y_{4i}'), i = 1, ..., n\}$, respectively. To simplify things for the reader, consider the imaging system as shown in Figure 2.
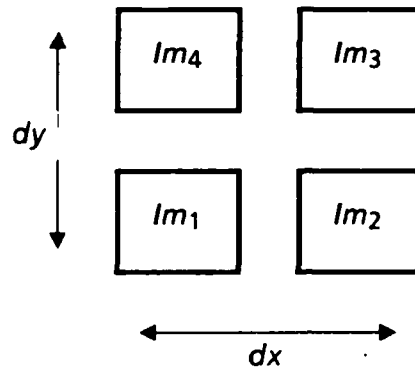

Figure 2: Orthographic Projection of the System on the Plane YZ

We proceed with the following propositions.

4.1 Proposition 1: Using the aforementioned nomenclature the quantity

$$\sum_{i=1}^{n} \frac{1}{Z_i}$$

is directly computable from the projection of the points of the set A on Im_1 and Im_2.

Proof: Consider a point $(X_i, Y_i, Z_i) \in A$ and its projections $A_1 = (x_{1i}, y_{1i})$, $A_2 = (x_{2i}, y_{2i})$ on Im_1 and Im_2 respectively (i.e., A_1 and A_2 are corresponding). Then

$$x_{1i} = f \frac{X_i}{Z_i} \quad \cdots \tag{4.1.1}$$

and

$$x_{2i} = \frac{f(X_i - dx)}{Z_i} \quad \cdots \tag{4.1.2}$$

9

From 4.1.1 and 4.1.2, we get

$$x_{1i} - x_{2i} = f \frac{dx}{Z_i}$$

or

$$\frac{1}{Z_i} = \frac{x_{1i} - x_{2i}}{fdx} \quad \ldots \tag{4.1.3}$$

Therefore,

$$\sum_{i=1}^{n} \frac{1}{Z_i} = \frac{1}{fdx} \left\{ \sum_{i=1}^{n} x_{1i} - \sum_{i=1}^{n} x_{2i} \right\} \quad \ldots \tag{4.1.4}$$

Equation (4.1.4) proves proposition 1.

4.2 Corollary: The quantity

$$\sum_{i=1}^{n} \frac{1}{Z_i'}$$

is also directly computable, from the projections of the set A' on Im_1 and Im_2.

4.3 Proposition 2: Using the aforementioned nomenclature, the quantity

$$\sum_{i=1}^{n} \frac{y_{1i}}{Z_i}$$

is directly computable from the projections of A on Im_1 and Im_2.

Proof: We have

$$\sum_{i=1}^{n} \frac{y_{1i}}{Z_i} = \sum_{i=1}^{n} \frac{y_{1i}}{fdx} (x_{1i} - x_{2i}) \quad [\textit{from (4.1.3)}]$$

or

$$\sum_{i=1}^{n} \frac{y_{1i}}{Z_i} = \frac{1}{fdx} \left(\sum_{i=1}^{n} y_{1i} x_{1i} - \sum_{i=1}^{n} y_{1i} x_{2i} \right)$$

But corresponding points in Im_1 and Im_2 have the same y coordinates, so

$$\sum_{i=1}^{n} \frac{y_{1i}}{Z_i} = \frac{1}{fdx} \left(\sum_{i=1}^{n} y_{1i} x_{1i} - \sum_{i=1}^{n} y_{2i} x_{2i} \right) \quad \ldots \tag{4.3.1}$$

10

The equation 4.3.1 proves proposition 2.

4.4 Proposition 3: Using the aforementioned nomenclature, the quantity

$$\sum_{i=1}^{n} \frac{x_{1i}}{Z_i}$$

is directly computable, from the projections of the set A on Im_1 and Im_4.

Proof: Similar to (4.1.3), we can derive

$$\frac{1}{Z_i} = \frac{1}{fdy}(y_{1i} - y_{4i}) \dots \tag{4.4.1}$$

Using (4.4.1), we get

$$\sum_{i=1}^{n} \frac{x_{1i}}{Z_i} = \frac{1}{fdy}\left\{\sum_{i=1}^{n} x_{1i}(y_{1i} - y_{4i})\right\} = \frac{1}{fdy}\left\{\sum_{i=1}^{n} x_{1i}y_{1i} - \sum_{i=1}^{n} x_{4i}y_{4i}\right\}, \dots \tag{4.4.2}$$

(since corresponding points in Im_1 and Im_4 have the same x coordinates). Equation (4.4.2) proves proposition 3.

5. Recovering Three-Dimensional Translation Without Correspondence

Consider the projections of the sets A and A' on Im_1. Furthermore, consider a point (x_{1i}, y_{1i}) and its dynamic corresponding one (x_{1i}', y_{1i}'). (Note that we do not consider point correspondence, i.e., we do not worry for the moment where the position of (x_{1i}', y_{1i}') is.) From Section 3 we have:

$$x_{1i}' - x_{1i} = \frac{f\Delta Y - y_{1i}\Delta Z}{Z_i'} \dots \tag{5.1}$$

$$y_{1i}' - y_{1i} = \frac{f\Delta X - x_{1i}\Delta Z}{Z_i'} \dots \tag{5.2}$$

If we write Equation (5.1) for all the pairs of corresponding points and we sum up these equations, we get

$$\sum x_{1i}' - \sum x_{1i} = f\Delta X \sum \frac{1}{Z_i'} - \Delta Z \sum \frac{x_{1i}}{Z_i'} \dots \tag{5.3}$$

Assuming that the motion in depth is small with respect to the depth equation,

11

(5.3) can be approximated by:

$$\sum x_{1i}' - \sum x_{1i} = \Delta X \left(f \sum \frac{1}{Z_i'} \right) - \Delta Z \left(\sum \frac{x_{1i}}{Z_i} \right) \dots \tag{5.4}$$

Similarly, with Equation (5.2) we obtain

$$\sum y_{1i}' - \sum y_{1i} = \Delta Y \left(f \sum \frac{1}{Z_i'} \right) - \Delta Z \left(\sum \frac{y_{1i}}{Z_i} \right) \dots \tag{5.5}$$

If we apply the same procedure for the projections of the sets A and A' on Im_2 we get two more equations. One of them is the same as (5.5), and the other is:

$$\sum x_{2i}' - \sum x_{2i} = \Delta X \left(f \sum \frac{1}{Z_i'} \right) - \Delta Z \left(\sum \frac{x_{2i}}{Z_i} \right) \dots \tag{5.6}$$

Equations (5.4) through (5.6) constitute a linear system in the unknowns ΔX, ΔY, ΔZ, which always has a unique solution, given by:

$$\widehat{\Delta Z} = \frac{(\sum x_{2i}' - \sum x_{1i}') - (\sum x_{2i} - \sum x_{1i})}{(\sum \frac{x_{1i}}{Z_i} - \sum \frac{x_{2i}}{Z_i})} \dots \tag{5.7}$$

$$\widehat{\Delta X} = \frac{\{\sum x_{1i}' - \sum x_{1i} + \widehat{\Delta Z}(\sum \frac{x_{1i}}{Z_i})\}}{(f \sum \frac{1}{Z_i'})} \dots \tag{5.8}$$

and

$$\widehat{\Delta Y} = \frac{\{\sum y_{1i}' - \sum y_{1i} + \widehat{\Delta Z}(\sum \frac{y_{1i}}{Z_i})\}}{(f \sum \frac{1}{Z_i'})} \dots \tag{5.9}$$

Note that the denominators in the expressions (5.7) through (5.9) are always different from zero (for dx, dy non-zero).

We now proceed with the experimentations.

12

6. Experiments, the Effect of Noise and Practical Considerations

First of all we must admit that we were expecting a small error in the computed parameters due to the approximations done in the development of Equations (5.4) through (5.6) (i.e., $\Sigma(x_{1i}/Z_i') \simeq \Sigma(x_{1i}/Z_i)$, $\Sigma(y_{1i}/Z_i') \simeq \Sigma(y_{1i}/Z_i)$, and $\Sigma(x_{2i}/Z_i') \simeq \Sigma(x_{2i}/Z_i)$), but experiments showed that when the motion in depth is small with respect to the depth, this error is negligible.

In our experiments we considered a set of three-dimensional points, we projected them on each of the four frames, and then we gave the three-dimensional points a rigid translation and we projected them again on the four frames. Discretization effects, when the three-dimensional translation is not small, hardly affect the results. Our experiments with noise indicate that the method seems to be immune to small percentages. When we say that a frame has $a\%$ noise, we mean that if the frame contains n points then $an/100$ of the points are randomly generated using a random number generator. Note that in all we have eight frames, four before the motion and four after the motion. And the noise we added was not necessarily of the same amount in all these different frames; so when we talk about a noise of $a\%$ we mean that the average noise present in all the frames is $a\%$, and on the other hand when we say that we have an error of $\beta\%$ in the translation, we mean that

$$\beta = 100 \times \frac{1}{3} \times \left(\frac{|\Delta X - \widehat{\Delta X}|}{\Delta X} + \frac{|\Delta Y - \widehat{\Delta Y}|}{\Delta Y} + \frac{|\Delta Z - \widehat{\Delta Z}|}{\Delta Z} \right)$$

where $(\Delta X, \Delta Y, \Delta Z)$ the actual translation (with $\Delta X \cdot \Delta Y \cdot \Delta Z \neq 0$) and $(\widehat{\Delta X}, \widehat{\Delta Y}, \widehat{\Delta Z})$ the computed ones.

Furthermore, if the set of three-dimensional points is not only translating but is also experiencing small rotations (less than 20°), around an axis passing through the center of gravity of the points, then the computed three-dimensional translation is hardly affected (error less than 5%).

In a practical situation (real images), operators have first to be applied on all eight frames (four before the motion and four after the motion) that will produce points of interest [47, 48, 1, 4, 9] in all images, and then the theory developed in this paper is applied to these points. But any method that will produce points of interest from the intensity images is bound to have errors due to the noise in the

images and the unpredictability of the natural scenes. So, the number of points will not be the same in the four frames, neither before nor after the motion. But despite the fact that our theory is built on the assumption that the number of points is the same in all frames, our experiments show that even if the number of points in the different frames is not the same (at most a difference of 5%), the results are hardly affected.

At this point, we should mention that the equations used in the experiments are modified so that they can capture the difficulties from the different number of points in the various frames. In particular, we do the following. Equations (5.4), (5.5), and (5.6) are not affected if both sides are divided by the number of points (which is supposed to be the same number). For example, Equation (5.4) becomes:

$$\frac{\sum x_{1i}'}{n} - \frac{\sum x_{1i}}{n} = \Delta x \frac{1}{n} f \sum \frac{1}{Z_i'} = \frac{\Delta Z}{n} \sum \frac{x_{1i}}{Z_i}$$

If the number of points in the first frame before the motion is n_1 and after the motion n_1', then the above equation is written as:

$$\frac{\sum x_{1i}'}{n_1'} - \frac{\sum x_{1i}}{n_1} = \Delta x \frac{1}{n'} f \sum \frac{1}{Z_i'} = \frac{\Delta Z}{n_1} \sum \frac{x_{1i}}{Z_i}$$

The same procedure is applied to the rest of Equations (5.5) and (5.6), as well as for the computation of the quantities $\Sigma(1/Z_i)$, $\Sigma(1/Z_i')$, $\Sigma(x_{1i}/Z_i)$, and $\Sigma(y_{1i}/Z_i)$. Clearly, this is an approximation, which seems very robust from extensive experimentations.

The table presented in Section 2, showing the error in the translation vs. noise in the images, has been produced by running 100 simulations for each noise case and then averaging and taking the ceiling of the computed errors. Finally, it is worth mentioning that our experiments indicate that discretization effects hardly affect the result, provided that the retinal motion is large enough (at least five pixels).

Finally, the appendix contains pictures from our experiments. Every picture shows four frames before the motion and the same four frames after the motion. The object that is imaged consists of connected points. The noise points are randomly generated and are not connected. The pictures in the first dynamic

frame (before the motion) are with green color, and the ones after the motion are with yellow. The noise sources are three: (1) the randomly put points; (2) discretization; and (3) rotation. The noise percentage written captures only the randomly generated points. (When we write that the noise level is, for example, 10%, we only mean that 10% of all the number of points in a frame is randomly generated; this noise percentage does not include rotation and discretization. Furthermore, when we write that the noise percentage is 10%, we mean that the average noise in all eight frames is 10%, since the noise in every frame is not the same, the maximum difference between any two frames being 5%.) The actual parameters and the computed ones (as well as the error in the computed parameters as it is defined previously) are shown in the pictures.

7. Conclusion and Future Directions

We have proposed a method for recovering the three-dimensional translation of a rigidly moving object. The method seems to be very robust against noise as well as small perturbations of the retinal points due to small rotations of the object. We have showed that combination of stereo and motion is a promising way of approaching the motion determination problem, as it has already been appreciated by Huang and Blostein. But we have also showed that at least for the case of translation, we can face the problem without having to go through the intermediate stage of the computation of point correspondences, neither static nor dynamic. Due to the special arrangement of cameras, we are not able at this point to recover the rotation parameter in the case where the object is translating and rotating, but as we have already stated, the method is immune to small rotations. We are currently working on addressing the general problem (translation and rotation) without correspondence, as well as making a theoretical analysis of the error of the translation parameters that are computed from our algorithm.

8. References

1. Davis, Z. Wu and H. Sun, "Contour based motion estimation," *CVGIP 23*, 313-326, 1983.

2. Haralick, R. and J. Lee, "The facet approach to optical flow," *Proc., Image Understanding Workshop*, Arlington, VA, June 1983.

3. Horn, B.K.P and B.G. Schunck, "Determining optical flow," *Artificial Intelligence 17*, 185-204, 1981.

4. Nagel, H., "Displacement vectors derived from second order intensity variations in image sequences," *CVGIP 21*, 85-117, 1983.

5. Ullman, S. and E. Hildreth, "The measurement of visual motion," in O.J. Braddick and A.C. Sleigh (Eds). *Physical and Biological Processing of Images (Proc., Int. Symp. Rank Prize Funds, London).* Springer-Verlag, September 1982 (pp. 154-176).

6. Ullman, S., "The interpretation of visual motion," Ph.D. thesis, 1977.

7. Ullman, S., "Analysis of visual motion by biological and computer systems," *IEEE Computer 14*, 8, 57-69, 1981.

8. Wohn, K. and A. Waxman, "Contour evolution, neighborhood deformation and local image flow: curved surfaces in motion," CAR-TR-134, Computer Vision Lab, U. Maryland, 1984.

9. Bandopadhay, A., "A multiple channel model for the perception of optical flow," *Proc., IEEE Workshop on Computer Vision: Representation and Control*, 78-82, 1984.

10. Adiv, G., "Determining three-dimensional motion and structure from optical flow generated from several moving objects," COINS-TR 84-07, U. Mass., Amherst, July 1984.

11. Bandopadhay, A. and J. Aloimonos, "Perception of rigid motion from spatio-temporal derivatives of optic flow," TR 157, Computer Science Dept., U. Rochester, 1985.

12. Bruss, A. and B.K.P. Horn, "Passive navigation," *CVGIP 21*, 3-20, 1983.

13. Fang, J.A. and T.S. Huang, "Solving three-dimensional small rotation motion equations: uniqueness, algorithms, and numerical results," *CVGIP 26*, 183-206, 1984.

14. Jenkin, M. and J. Tsotsos, "Applying temporal constraints to the dynamic stereo problem," Computer Science Dept., U. Toronto; to appear, *CVGIP*, 1986.

15. Kanade, T., "Camera motion from image differentials," *Proc., Annual Meeting, Optical Society of America*, Lake Tahoe, March 1985.

16. Aloimonos, J. and C. Brown, "The relationship between optic flow and surface orientation," *Proc., 7th ICPR*, Montreal, Canada, 1984.

17. Huang, T.S., "Three-dimensional motion analysis by direct matching," *Proc., Topical Meeting on Machine Vision*, Incline Village, Nevada, 1985.

18. Kanatani, K., "Detecting the motion of a planar surface by line and surface integrals," *CVGIP 29*, 13-22, 1985.

19. Kanatani, K., "Tracing planar surface motion from a projection without knowing the correspondence," *CVGIP 29*, 1-12, 1985.

20. Longuet-Higgins, C. and K. Prazdny, "The interpretation of a moving retinal image," *Proc., Royal Society of London B208*, 385-397, 1980.

21. Prazdny, K., "Determining the instantaneous direction of motion from optical flow generated by curvilinearly moving observer," *CVGIP 17*, 94-97, 1981.

22. Rieger, J. and D. Lawton, "Determining the instantaneous axis of translation from optic flow generated by arbitrary sensor motion," COINS-TR 83-1, U. Mass., Amherst, January 1983.

23. Roach, J. and J.K. Aggarwal, "Determining the movement of objects from a sequence of images," *IEEE Trans. PAMI 2*, 6, 554-562, November 1980.

24. Tsai, R.Y., T.S. Huang, W. Zhu, "Estimating three-dimensional motion parameters of a rigid planar patch II: singular value decomposition," *IEEE Trans. ASSP 30*, 525-534, August 1982.

25. Tsai, R.Y. and T.S. Huang, "Estimating three-dimensional motion parameters of a rigid planar patch III: finite point correspondences and the three view problem," *Proc., IEEE Conf. ASSP*, Paris, May 1982.

26. Tsai, R.Y. and T.S. Huang, "Uniqueness and estimation of three-dimensional motion parameters of rigid objects," in S. Ullman and W. Richards (Eds). *Image Understanding 1984*. New Jersey: Ablex Publishing Co., 1984 (p. 135).

27. Ullman, S., "Computational studies in the interpretation of structure and motion: summary and extension," AI Memo 706, MIT AI Lab, March 1983.

28. Yen, B. and T.S. Huang, "Determining three-dimensional motion and structure of rigid objects containing lines using the spherical projection," in T.S. Huang (Ed). *Image Sequence Processing and Dynamic Scene Analysis*. 1983.

29. Waxman, A.M. and S.S. Sinha, "Dynamic stereo: passive ranging to moving objects from relative image flows," CAR-TR 74, Computer Vision Lab, U. Maryland, July 1984.

30. Waxman, A.H. and S. Ullman, "Surface structure and three-dimensional motion from image flow: a kinematic approach," CAR-TR 24, Center for Automation Research, U. Maryland, 1983.

31. Prazdny, K., "Egomotion and relative depth map from optical flow," *Biol. Cybernetics 36*, 1980.

32. Nagel, H., "On the derivation of three-dimensional rigid point configurations from image sequences," *Proc., PRIP*, Dallas, TX, 1981.

33. Fang, J.Q. and T.S. Huang, "Estimating three-dimensional movement of rigid objects: experimental results," *Proc., 8th IJCAI*, Karlsruhe, West Germany, 1983.

34. Jerian, C. and R. Jain, "Determining motion parameters for scenes with translation and rotation," *Proc., Workshop on Motion*, Toronto, Canada, 1983.

35. Lawton, D., "Motion analysis via local translational processing," *Proc., Workshop in Computer Vision*, Rindge, NH, 1982.

36. Negahdaripour, S. and B.K.P. Horn, "Determining three-dimensional motion of planar objects from image brightness patterns," *Proc., 9th IJCAI*, 898-901, Los Angeles, CA, August 1985.

37. Aloimonos, J. and C. Brown, "Direct processing of curvilinear sensor motion from a sequence of perspective images," *Proc., IEEE Workshop on Computer Vision: Representation and Control*, Annapolis, MD, 1984.

38. Sarachan, B., "Experiments in rotational egomotion calculation," TR 152, Computer Science Dept., U. Rochester, February 1985.

39. Barnard, S.T. and W.B. Thompson, "Disparity analysis of images," *IEEE Trans. PAMI 2*, 333-340, 1980.

40. Prager, J.M. and M.A. Arbib, "Computing the optic flow: the MATCH algorithm and prediction," *CVGIP 21*, 271-304, 1983.

41. Nagel, H.-H. and B. Neumann, "On three-dimensional reconstruction from and two perspective views," *Proc., 7th IJCAI*, Vancouver, Canada, 1981.

42. Longuet-Higgins, H.C., "A computer algorithm for reconstructing a scene from two projections," *Nature 293*, 1981.

43. Schunck, B., personal communication, 1984.

44. Aloimonos, J., "Low-level visual computations," Ph.D. thesis, Computer Science Dept., U. Rochester, in preparation.

45. Huang, T.S. and S.D. Blostein, "Robust algorithms for motion estimation based on two sequential stereo image pairs," *Proc., CVPR*, 518-523, 1985.

46. Richards, W., "Structure from stereo and motion," *JOSA A 2*, 2, February 1985.

47. Kitchen, L. and A. Rosenfeld, "Gray-level corner detection," TR 887, Computer Science Dept., U. Maryland, 1980.

48. Moravec, H.P., "Towards automatic visual obstacle avoidance," *Proc., 5th IJCAI*, 1977.

END

5-81

DTIC