NISTIR 4764

Computing Depth From Temporal Cross-Correlation: A Comparison of Two Methods of Computation

JOHN A. HORST

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Unmanned Systems Group
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

U.S. DEPARTMENT OF COMMERCE
Rockwell A. Schnabel, Acting Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

NIST

NISTIR 4764

Computing Depth From Temporal Cross-Correlation: A Comparison of Two Methods of Computation

JOHN A. HORST

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Unmanned Systems Group
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

February 1992

U.S. DEPARTMENT OF COMMERCE
Rockwell A. Schnabel, Acting Secretary

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

Computing Depth from Temporal Cross-Correlation: A Comparison of Two Methods of Computation

John Albert Horst
National Institute of Standards and Technology (NIST)
Bldg 220, Rm B-124, Gaithersburg, MD 20899, USA

Abstract

Calculating the distance from camera to feature point by temporal cross-correlation of pixels in a flat image plane is better done using simple geometric relations than using time differentials. Both these methods of computation will be presented and contrasted. The effect of measurement error and other inaccuracies on the determination of range is also described.

1 Introduction

An important problem in machine vision is to determine the distance from camera to feature by appropriately interpreting a two dimensional image or sequence of images of that feature [1,2,3,4,5]. Several solutions have succeeded well by exploiting a sequence of images obtained through relative motion of camera and feature point [6,7,8,9,10,11,12,13]. Another method of determining distance, called temporal cross-correlation [14], which exploits known motion shows promise. This is similar in principle to the Reichardt model of human motion sensors in which one calculates the time required for a feature (e.g., a point source of light) to move between two small detectors which are separated by a known and fixed distance [15].

The problem might best be introduced by examining figure 1 and the definitions within it.

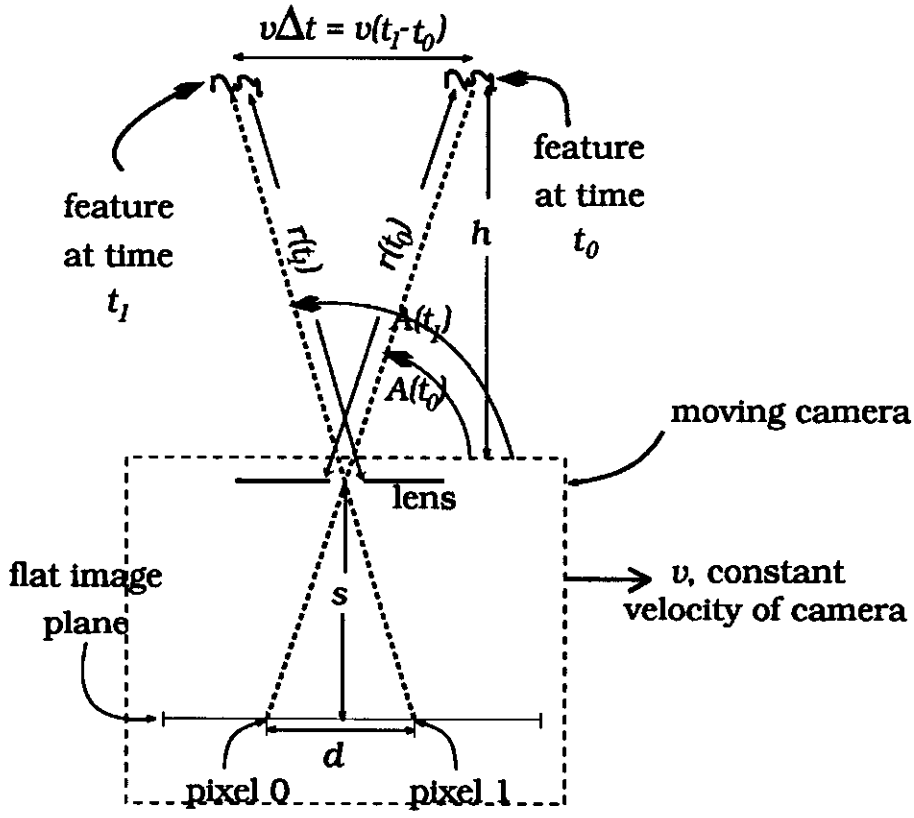

Fig 1

The goal is to determine the perpendicular depth, h, to a small (small enough to be nearly flat), yet, finite sized feature. We assume a simple case where a camera is moving perpendicular to the optical axis and there is only one component of camera motion. It is assumed that the velocity of the camera is known and constant over the time interval t_0 to t_1. Figure 1 indicates that pixels 0 and 1 are symmetrically located in the image plane on either side of the optical axis. We define d as the center-to-center distance between pixels. The actual width of each of the two pixels will, typically, be less than d. If one can determine the time, Δt, required for the feature to be detected by the pixel (picture element) at position 1 on the flat image plane after it is detected by the pixel at position 0, then h can be found. Two methods of obtaining h will be given: a simple geometric method and a time differential method.

2 Geometric Method

A geometric method for finding the depth, h, falls easily out of the definitions in figure 1 by noting the two similar triangles.

$$\frac{h}{v\Delta t} = \frac{s}{d} \Rightarrow h = \frac{sv\Delta t}{d} \tag{1}$$

Δt is the only unknown in this expression and is obtained by the method of temporal correlation as discussed in section 4.

3 Time Differential Method

A time differential method for finding the depth, h, can also be derived using the definitions in figure 1.

$A(t)$ is defined as the angle (at time t) between the camera velocity vector, v, and the range vector, $r(t)$, from lens to feature. h is the perpendicular distance from camera to feature and is constant with respect to time (see figure 1). Let $t_0 = -t_1$ in figure 1. This implies (assuming a constant camera velocity) that $A(t = 0) = \pi/2$. As a result of these definitions,

$$\cot(A(t)) = -\frac{vt}{h} \Rightarrow \frac{d}{dt}\left(-\frac{vt}{h}\right) = \frac{d}{dt}\cot(A(t)) \Rightarrow \frac{v}{h} = \frac{dA(t)}{dt}\csc^2(A(t)) \qquad (2)$$

From simple geometry,

$$\csc(A(t)) = \frac{r(t)}{h} \qquad (3)$$

If we substitute equation 3 into equation 2, we get,

$$\frac{v}{h} = \frac{dA(t)}{dt}\frac{r(t)}{h}\csc(A(t)) \text{ or } r(t) = \frac{v\sin(A(t))}{dA(t)/dt} \qquad (4)$$

The accuracy of equation 4 for calculating $r(t)$ is dependent on whether one can accurately evaluate $dA(t)/dt$, which, among other things, is dependent on the shape of $A(t)$. From equation 2,

$$A(t) = \cot^{-1}(-vt/h) \qquad (5)$$

Referring back to figure 1, we see that $t_0 = -\Delta t/2$ and $t_1 = \Delta t/2$. Then we can define $\Delta A(0) \triangleq A(t_1) - A(t_0)$. From equation 5 and figure 2 it can be seen that $A(t)$ is not a linear function of t. In fact,

$$\left.\frac{dA(t)}{dt}\right|_{t=0} > \frac{\Delta A(0)}{\Delta t} \qquad (6)$$

as can be seen from figure 2.

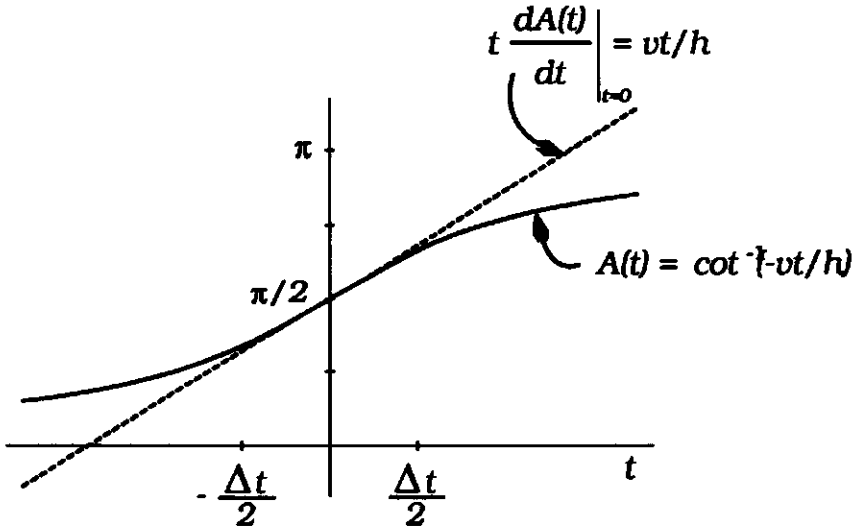

Fig 2

Using equation 4 and since the derivative is only approximated by $\Delta A(t)/\Delta t$,

$$r(t) \approx \frac{v\Delta t \sin(A(t))}{\Delta A(t)} \tag{7}$$

Referring to figure 1 again and letting $t = 0$,

$$r(0) = h \approx \frac{v\Delta t}{\Delta A(0)} \tag{8}$$

Equation 8 says that the depth, h, to the feature is approximated by the constant velocity of the camera times the time, Δt, it takes the light from the feature to traverse through $\Delta A(0)$ radians divided by $\Delta A(0)$. In addition, the expression, $v\Delta t/\Delta A(0)$, in equation 8 will always be greater than the correct value for the depth, h. This can be seen from the shape of $A(t)$ as depicted in figure 2.

4 Temporal Correlation

It has been shown [14] that Δt (in equations 1 and 8) can be found by doing cross-correlation of time sequences of intensity output of the pixels at positions 0 and 1 on the image plane (see figure 1). Each of these time sequences have N components. In this sense, Δt is equal to the shifting required to match or "fit" one time sequence with the other. In mathematical terms, if x_n is the finite series of discrete time samples of the analog output of the pixel at position 0, and y_n is the same for the pixel at position 1, then we want to find the $p_{\min} \in \{0,1,...\}$ that minimizes

$$\sum_{n=0}^{N-1} \left(x_n - y_{n+p}\right)^2 \tag{9}$$

over all positive integers p. If the sampling period is T, then $\Delta t = p_{\min} T$.

5 Error in the Distance to a Feature

An advantage to equation 1 ($h = sv\Delta t/d$) is that there might be a gain in the precision (and, perhaps, accuracy) of h (the

perpendicular distance to a feature) by making d bigger without sacrificing the exactness of the expression. It is argued below that, under certain conditions, there is some (possibly slight) gain in precision by making d bigger.

Equation 1 is only theoretically exact. Alas, the true values, d and s, may not perfectly match the published specifications of the manufacturers of the detector and camera. In addition, v is a measured quantity that will have its own errors, dependent upon the precision and accuracy of the velocity measuring equipment. Δt results from the cross correlation operation and is also a measured quantity with its own sources of error (*e.g.* detector noise, amplifier noise, poor focus to the feature, quantization, occlusion, and lack of flatness in the feature).

We will now define these errors and analyze their effect on h. For the remainder of this section, consider that h, s, d, v, Δt are random variables of mean μ_h, μ_s, μ_d, μ_v, $\mu_{\Delta t}$ and variance σ_h^2, σ_s^2, σ_d^2, σ_v^2, $\sigma_{\Delta t}^2$.

In addition, it can be expected that the errors in s and d will be fixed with respect to time and are therefore semi-correctable (*i.e.* calibrate the optical system by placing an object at a known distance from the lens and making a sufficient number of measurements of h in order to determine the correction factor). In this case, we can assume that

$$\sigma_d^2 = \sigma_s^2 = 0 \tag{10}$$

implying that s and d are deterministic. We will further assume that the random variables, v and Δt, are independent since their errors are caused by independent measurement systems. From equation 1 we get an expression for σ_h^2, the error in depth,

$$\sigma_h^2 = \frac{s^2}{d^2}\left(\sigma_v^2\sigma_{\Delta t}^2 + \sigma_v^2\mu_{\Delta t}^2 + \sigma_{\Delta t}^2\mu_v^2\right) \tag{11}$$

Now we will show that an increase in d will cause a decrease in σ_h^2. From figure 1 we see that, if we increase d to d', $\mu_{\Delta t}$ will increase to, say, $\mu_{\Delta t'}$. In addition,

$$\frac{\mu_{\Delta t}}{d} = \frac{\mu_{\Delta t'}}{d'} \tag{12}$$

This gives a new expression for $\sigma_h'^2$, the new error in depth,

$$\sigma_h'^2 = \frac{s^2}{d'^2}\left(\sigma_v^2\sigma_{\Delta t}^2 + \sigma_v^2\mu_{\Delta t'}^2 + \sigma_{\Delta t}^2\mu_v^2\right) \tag{13}$$

In equation 13, it is assumed that $\sigma_{\Delta t}^2 \approx \sigma_{\Delta t'}^2$, which is to say that the variance of the random variable Δt will be effectively independent of the size of its mean.

Using equation 12, we find that the second terms on the right hand side of equations 11 and 13 are equal. Therefore, the first and third terms on the right hand side of equations 11 and 13, $s^2\left(\sigma_v^2\sigma_{\Delta t}^2 + \sigma_{\Delta t}^2\mu_v^2\right)/d^2$ and $s^2\left(\sigma_v^2\sigma_{\Delta t}^2 + \sigma_{\Delta t}^2\mu_v^2\right)/d'^2$, respectively, prove our claim (i.e., if $d < d'$, $\sigma_h^2 > \sigma_{h'}^2$).

To give a realistic example, let s = 60 mm, d = 0.5 mm, μ_v = 1 m/s, and μ_h = 30 cm (corresponds to a thin lens system with focal length equal to 50 mm). These values imply that $\mu_{\Delta t}$ = 2.5 ms. Let σ_v = 0.05 m/s and $\sigma_{\Delta t}$ = 50 μs. If we increase d by a factor of ten, d' = 5 mm (with no change in the **size** of each pixel) and $\mu_{\Delta t'}$ = 25 ms. Finally, using equations 11 and 13, we see that, indeed, the error in depth is reduced slightly, since σ_h' = 1.5 cm < σ_h = 1.6 cm. Note, however, that a rather large change in d produces a relatively small change in depth accuracy, since $(\sigma_h - \sigma_h')/\mu_h < 1\%$.

It should also be noted that an increase in Δt might cause greater error in camera velocity (i.e., an increase in σ_v^2). Therefore, an increase in d may very well not decrease the error in depth. To demonstrate this, using the values given in the above example, say that σ_v = 0.05 m/s increases to σ_v' = 0.055 m/s. With only this slight change, we see that (even with the tenfold increase in d) σ_h' = 1.65 cm > σ_h = 1.62 cm, i.e. the error in depth increases.

There is a further cost for choosing d large. If the scene has lots of occlusion, the pixel at position 0 may see the feature while the pixel at position 1 does not, if, at that perspective, the feature is occluded by some other feature. This tradeoff between accuracy and

occlusion would have to be considered for each individual application.

In conclusion, if the scene has minimal occlusion and one expects no increase in the camera velocity error, one might choose d to be large as possible to get maximum accuracy. However, in general, one can more easily maintain the required constant camera velocity over a shorter time period which argues for d small. Otherwise, one would choose pixels that are adjacent or nearly adjacent (which is equivalent to choosing d small).

6 Conclusion

All else being equal, it has been demonstrated that equation 1 (the geometric method) is a more accurate way to calculate depth, h, than is equation 8 (the time differential method), assuming a flat image plane (which will be true for the vast majority of available semiconductor photo detector arrays). Both methods require that the camera velocity, v, be constant over the time, Δt. Δt can be a relatively long time for distant features. Nonetheless, for many applications, the variation of v during the time, Δt, will be negligible.

It is argued in section 5 that, under certain conditions (minimal occlusion and no change in camera velocity error), there is some gain in precision by making d (in equation 1) bigger. However, there is a loss using the time differential method of equation 8 in making d larger (increasing the field-of-view) as can be seen by examining figure 2.

7 Acknowledgements

The author gratefully recognizes the technical help received from James Albus, Martin Herman, Tsai Hong Hong, Herbert Lau, and Richard Quintero. Many thanks to Ann Hargett and Shirley Jack for help preparing the manuscript.

8 References

[1] E. W. Kent and M. O. Shneier, Eyes for automatons, "The real problem of three-dimensional vision is in acquiring the range image", *IEEE Spectrum*, 1986, 23, 3, pp 37-45.

[2] R. A. Jarvis, "A perspective on range finding techniques for computer vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5, 2, pp 122-139.

[3] B. K. P. Horn, Obtaining shape from shading information, in *The Psychology of Computer Vision*, P. H. Winston Ed., McGraw-Hill, pp 115-155.

[4] B. Julesz, "Binocular depth perception without familiarity cues," *Science*, 145, pp 356-362.

[5] R. Bajcsy and Lieberman, "Texture gradient as a depth cue," *Computer Graphics Image Processing*, 5, pp 52-67.

[6] H. C. Longuet-Higgins and K. Prazdny, "The interpretation af a moving retinal image," *Proceedings of the Royal Society of London B*, 208, pp 385-397.

[7] Joachim Heel, "Direct dynamic motion vision", *Proceedings of the IEEE Conference on Robotics and Automation*, 1990, pp 1142-1147.

[8] H. Harlyn Baker, "Scene structure from a moving camera," in *AI and the Eye*, Wiley, 1990, pp 229-260.

[9] Michael K. Brown and David Lee, "From a sequence of images to a depth map", *Ninth IASTED International Symposium on Robotics and Automation*, Santa Barbara, May 27-29, 1987, pp 6-11.

[10] M. L. Braunstein, *Depth perception through motion*, Academic Press, New York, 1976.

[11] Larry Matthies and Takeo Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," in *International Journal of Computer Vision*, 3, 1989, pp 209-236.

[12] David Vernon and Massimo Tistarelli, "Using camera motion to estimate range for robotic parts manipulation," *IEEE Transactions on Robotics and Automation*, Vol. 6, No. 5, October 1990, pp. 509-521.

[13] Larry Matthies and Takeo Kanade, "Kalman filter-based algorithms for estimating depth from image sequences," in *International Journal of Computer Vision*, 3, 1989, pp 209-236.

[14] James S. Albus and Tsai Hong Hong, "Motion, depth, and image flow", *Proceedings of the IEEE Conference on Robotics and Automation*, 1990, pp 1161-1170.

[15] Jan P. H. van Santen and George Sperling, "Elaborated Reichardt detectors", *Journal of the Optical Society of America A*, Vol. 2, No. 2, February 1985, pp 300-321.

NIST-114A
(REV. 3-89)

U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY

BIBLIOGRAPHIC DATA SHEET

1. PUBLICATION OR REPORT NUMBER
NISTIR 4764

2. PERFORMING ORGANIZATION REPORT NUMBER

3. PUBLICATION DATE
FEBRUARY 1992

4. TITLE AND SUBTITLE
Computing Depth from Temporal Cross-Correlation: A Comparison of Two Methods of Computation

5. AUTHOR(S)
John A. Horst

6. PERFORMING ORGANIZATION (IF JOINT OR OTHER THAN NIST, SEE INSTRUCTIONS)
U.S. DEPARTMENT OF COMMERCE
NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
GAITHERSBURG, MD 20899

7. CONTRACT/GRANT NUMBER

8. TYPE OF REPORT AND PERIOD COVERED

9. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (STREET, CITY, STATE, ZIP)

10. SUPPLEMENTARY NOTES

[ ] DOCUMENT DESCRIBES A COMPUTER PROGRAM; SF-185, FIPS SOFTWARE SUMMARY, IS ATTACHED.

11. ABSTRACT (A 200-WORD OR LESS FACTUAL SUMMARY OF MOST SIGNIFICANT INFORMATION. IF DOCUMENT INCLUDES A SIGNIFICANT BIBLIOGRAPHY OR LITERATURE SURVEY, MENTION IT HERE.)

Calculating the distance from camera to feature point by temporal cross-correlation of pixels in a flat image plane is better done using simple geometric relations than using time differentials. Both methods of computation will be presented and contrasted. The effect of measurement error and other inaccuracies on the determination of range will also be covered.

12. KEY WORDS (6 TO 12 ENTRIES; ALPHABETICAL ORDER; CAPITALIZE ONLY PROPER NAMES; AND SEPARATE KEY WORDS BY SEMICOLONS)

distance to feature; flat image plane; geometric method; machine vision; measurement error; motion vision; occlusion; Reichardt model; temporal cross-correlation, time differential method

13. AVAILABILITY

[X] UNLIMITED

[ ] FOR OFFICIAL DISTRIBUTION. DO NOT RELEASE TO NATIONAL TECHNICAL INFORMATION SERVICE (NTIS).

[ ] ORDER FROM SUPERINTENDENT OF DOCUMENTS, U.S. GOVERNMENT PRINTING OFFICE, WASHINGTON, DC 20402.

[X] ORDER FROM NATIONAL TECHNICAL INFORMATION SERVICE (NTIS), SPRINGFIELD, VA 22161.

14. NUMBER OF PRINTED PAGES
13

15. PRICE
A02

ELECTRONIC FORM