Borrower: U$C
Lending String: *CLA,CCH,CSA,CNO,CPS
Patron: Wolfe, Bill
Journal Title: IEEE transactions on pattern analysis and machine intelligence.
Volume: 13 Issue: 1
Month/Year: Jan 1991 Pages: 66 - 73
Article Author:
Article Title: Wolfe, W.J. / The perspective view of three points
ILL Number: 36710169

Call #: Q327 .I19
Location: Periodicals North - A LIB USE ONLY

ARIEL
Charge
Maxcost: $25 IFM
Shipping Address:
California State University Channel Isla
University Library - Interlibrary Loan
Camarillo, CA 93012-8599
Fax:
Ariel:
NOTICE: This material may be protected by copyright law (Title 17 U.S. Code).
Fig. 6. Errors of the selected estimates of the motion parameters as functions of the solid angle subtended by the object for NDATA = 12 and 1% noise. The solid angle is a decreasing function of the variable FOCAL. (See Section VI.) Averages are denoted by $\Delta$ and medians by $m$. Accuracy and selection levels are the same as in Fig. 4. ** shows selected percentage of data (with 100% = 10 on the scale). This percentage is below 10% for FOCAL = 3. By using higher levels for AEA and $\gamma_3$ at the larger FOCAL values, 25% of the data can be retained with only 5% increase of the errors. The wavy character of the curves around FOCAL = 0.5 is probably due to the data generation process in which data sets making $\gamma_3$ negative are discarded.

so that line of the equation system for $Q$ is

$$Y + e_Y I \hat{Q}(X + e_X) = Y^t \hat{Q} X + Y^t \hat{Q} e_X + e_Y \hat{Q} X + e_Y \hat{Q} e_X = 0.$$  

Neglecting the last term, we have approximately

$$Y^t \hat{Q} X = -Y^t \hat{Q} e_X - e_Y \hat{Q} X.$$  

The RHS of this equation can be thought of as errors of $b$ that result in a residual $r$. We have assumed the $e_X$ and $e_Y$ to be independent random variables with the same distribution. Let $\epsilon$ denote a random variable with this distribution. Then, the RHS above is a linear combination of four independent $\epsilon$ variables and we have

$${\text{var}}\{b\} = [(Y^t \hat{Q} X)^2 + (Y^t \hat{Q} _2)^2 + (Y^t X_1)^2 + (Y^t X_2)^2] \cdot {\text{var}}\{\epsilon\}.$$  

Define ERRFACTOR as the average of the expression in the square brackets over the NDATA equations. This gives the following relation between the estimate of ${\text{var}}\{b\}$ and the sought $${\text{var}}\{\epsilon\} = \text{ERRFACTOR} \cdot {\text{var}}\{\epsilon\}.$$  

Solution of the system for $Q$ by SVD gives $${\text{var}}\{b\}.$$  

The obtained values of $${\text{var}}\{\epsilon\}$$ have a fairly low accuracy and are not used as test variables to sort out inaccurate estimates.

REFERENCES


The Perspective View of Three Points

William J. Wolfe, Donald Mathis, Cheryl Weber Sklair, and Michael Magee

Abstract—In this paper we focus on the perspective view of three noncollinear points whose image-to-object correspondence is known. Such measurements are known to be ambiguous, resulting in as many as four possible solutions to the perspective three-point problem. Although there can be four solutions, it is quite often the case that there are only two solutions. We provide a geometric explanation of the camera-triangle configurations that cause one, two, three, or four solutions. The results also provide a justification for the common wisdom that there are "usually" two solutions.

Index Terms—Computer vision, perspective n-point problem, perspective projection, perspective view of a triangle.
I. INTRODUCTION

There have been many approaches to building model-based vision systems, typically involving a combination of the following [1]–[5]:

1) Extracting primitives (e.g., edges and corners from the image).
2) Assigning a correspondence between extracted primitives and object model features.
3) Using a camera model to infer the 3-D positions of the object that are consistent with the observed primitives.
4) Introducing domain constraints (e.g., viewpoint assumptions) and iterating until a stable solution is derived.

In this correspondence we are primarily interested in step 3, but in order to keep the context clear it is important to list the basic difficulties associated with each step:

1) Extracting primitives is unreliable and inaccurate (noisy and incomplete).
2) The combinatorics of possible assignments of image-derived primitives to object features can be explosive.
3) The mathematical relationships between the geometry of solid objects and their image projections are sometimes difficult to solve, even under the assumption of perfect data.
4) Domain constraints seriously limit the general applicability of a method.

The mathematics of image projections depends directly on the camera model used, for which the following are the simplest assumptions:

1) Orthographic assumption: parallel projection.
2) Weak Perspective assumption: orthographic projection followed by a scaling factor.
3) Perspective assumption: a pinhole camera.

In general, the orthographic assumption is the easiest to work with, but it is also the least accurate. The weak perspective assumption closely approximates perspective, but degenerates with increasing depth of field. The perspective assumption is the most realistic, but it is the most difficult to solve in closed form. Because of these tradeoffs the computer vision literature is filled with a seemingly endless series
of variations on camera models. Our focus in this correspondence is on the perspective assumption.

Among the many image-derived primitives that might be used to suggest 3-D locations of a rigid object are the following:

1) Three noncollinear points. The three points can be any three noncollinear points on the rigid object. They are usually geometrically significant features, such as the vertices of a polyhedron, but this need not be the case in general. Throughout this correspondence we refer to the three points as a "triangle", but it is assumed that only the vertices of the triangles are actually detected in the image [6]–[9].

2) Vertex-pair. This consists of a line segment (spine) between two corners detected in the image. It uses only the angle measurements at a single corner; although it is assumed that the detected corners correspond to object vertices, the spine does not have to correspond to an edge of the object [10], [11].

3) Four coplanar points. It is assumed that no three of the points are collinear, and most typically we will assume that the points form a rectangle [12]–[16].

The use of three points is not very popular, with the notable exception of the work described in [9], primarily because of the fourfold ambiguity that can arise even when the image-to-object correspondence is known. The geometric nature of these ambiguities is the main focus of this correspondence. Fischer and Bolles [6] provided the foundational analysis of this problem (what they call the "perspective three-point problem") by setting up the constraining equations and demonstrating that there cannot be more than four solutions (they also give an example where there are exactly four solutions). We add to their work by providing a systematic analysis of the geometric configurations that are associated with ambiguity sets of one, two, three, or four solutions. Furthermore, we provide a justification for the commonly held wisdom, as commented in [9], that the solution set "usually" consists of two configurations. For a review of the general perspective $n$-point problem, see [20].

II. TRIANGLE PROBLEM

The triangle problem (or perspective three-point problem) can be defined as follows: From a single perspective view of the vertices of a known triangle, determine all the possible camera-triangle configurations. (The basic geometry of the problem is shown in Fig. 1.)

It is assumed throughout this paper that the correspondence between images of points $(A', B', C')$ and the vertices of the triangle $(A, B, C)$ is known. The perspective view of the triangle consists only of image plane coordinates that we assume to be of infinite resolution. Furthermore, there is no use of any other features such as brightness of the points, the shape of the points, the presence of edges, surface shading, etc.

III. "ROTATED-LEG" SOLUTIONS

Consider the plane that is perpendicular to the plane of the triangle and contains one of the altitudes of the triangle (Fig. 2). It does not matter which of the three altitudes is chosen at this point. When the center of perspective (CP) lies in this plane, it is easy to see that a rotation of the triangle about its base (as defined by the chosen altitude) will create a second position of the triangle that produces an identical projection. The geometry of this situation is that the circle formed by the moving vertex intersects the line of perspective in two places. Two special exceptions must be noted: 1) When the triangle is relatively close to the camera, this rotation may place the triangle partly behind the camera to produce a physically unrealizable configuration. 2) There are cases when the circle formed by the moving vertex is tangent to the line of perspective, in which case a second solution is not produced.

If we now consider all three planes that can be identified in this manner, defined in the same way by each of the three altitudes, we can generate situations with fourfold ambiguities. Note that first because the altitudes of the triangle intersect at a single point (the orthocenter of the triangle), the three planes intersect along a single line $L$ (Fig. 3). $L$ is perpendicular to the triangle and intersects it at the orthocenter. If the CP lies on $L$, there will be a fourfold ambiguous measurement. That is, the triangle can be rotated about each of its sides in turn to create three more positions of the triangle that would produce exactly the same perspective measurement.

IV. THE CANONICAL VIEW

There are six parameters in general, three positional and three rotational, that define a camera-triangle configuration. When determining the existence of multiple solutions, however, it is only the position of the CP relative to the triangle that matters. We only restrict the pointing of the camera to be such that the triangle is completely in front of the camera. For example, when discussing the "rotated-leg" solutions above, there was no need to mention any specific pointing angles. Thus, whenever it is necessary to specify the pointing angles we take a "canonical view" defined by the assumption that the camera is pointing at the apparent center of one of the sides of the triangle (Fig. 4). This will be useful later in the correspondence when we report our empirical results.

V. THE "ROTATION CIRCLE"

We have already shown how to produce multiple solutions under very special assumptions about the position of the CP relative to the triangle, but our goal in this correspondence is to attain more general results. Toward this end we will introduce a particular geometric parameterization of the problem.

One geometric representation of this problem is shown in Fig. 5. Note that the vertices of the triangle are labeled $A$, $B$, $C$ while the image of the vertices are labeled $A'$, $B'$, $C'$. With the CP fixed, we can envision the search for solutions as the process of sliding the side $AB$ along the lines of perspective CPA and CPB, while rotating about the side $AB$ until $C$ projects to the measured $C'$. The triangle is rotated about side $AB$, the locus of the moving vertex $(C')$ is a circle that we will refer to as a "rotation circle." In general, a rotation circle projects to an ellipse on the image plane. In the special case when the plane of a rotation circle intersects the CP, as shown in Fig. 5, the ellipse degenerates to a line segment. It is easy to see that in this case any $C'$ that lies on this line segment will correspond to two solutions that have $AB$ in the same position. It should also be noted that this situation corresponds to the "rotated-leg" case mentioned above that produces two solutions from rotating the triangle about one of its sides.
the triangle and that the image of the side is perfectly horizontal with respect to the image plane.

Fig. 5. The locus of the vertex $C$ as side $AB$ is rotated forms a circle in 3-D, a "rotation circle," which projects to an ellipse on the image plane.

VI. THE "SWEPT" ROTATION CIRCLE

A general way to detect multiple solutions is to look for intersecting ellipses, projections of rotation circles, on the image plane. For example, Fig. 7 shows how the intersection of two ellipses corresponds to a two-solution case. If the side $AB$, along with its associated rotation circle, is "swept" along the constraint defined by angle $\theta$, a sequence of corresponding ellipses can be projected into the image (Fig. 8). The two-dimensional surface generated by sweeping the rotation circle in this manner is a complex surface, but happens to be a generalized cylinder and is very similar to the "cuspical caustic" shown in [17, photograph 2b]. Now, a line of perspective can be thought of as emanating from the CP, intersecting the image plane at $C'$, and then continuing into space until it intersects the surface. Thus, multiple solutions result from a single perspective line intersecting the surface at multiple points. (This must be interpreted carefully since the surface is self-intersecting.)

VII. SIMULATION RESULTS

We simulated the canonical view of many triangles in this manner, an example of which is shown in Fig. 9. The projection of the ellipses onto the image plane provides an "inside-out" view of the surface swept out by the rotation circles. With this geometric guidance, no means sufficient in and of itself, a numerical solution method was used to count the number of solutions corresponding to various locations of $C'$ within the image for a given $A'$ and $B'$. Fig. 9 shows a typical result. The two largest ellipses in the image plane correspond to having vertex $A$ and vertex $B$, respectively, as close as possible to the camera. Intuitively, the two-solution region results from a conjunction of left-handed and right-handed triangle positions. That is, the two-solution regions can usually be described as resulting from one position of the triangle where $A$ is closer to CP than $B$ and the triangle is leaning either forward or back, and another position of the triangle where $B$ is closer than $A$ and the triangle is leaning in the other direction.

The one-solution regions demonstrate that the triangle can lean enough to be out of reach of the other-handed possibility. The three-solution case occurs at the boundary between the four- and two-solution regions. The no-solution regions imply that the physical constraints defined by the measurement of $A'$ and $B'$ and the size of the actual triangle make it impossible for $C'$ to appear at those regions. The cusp-like four-solution region is very complex, resulting
Fig. 6. When the $CP$ lies in the plane of the rotation circle, the projection degenerates to a line segment. When $C'$ lies on this line segment, there must be two consistent triangle positions, the "rotated leg" solutions corresponding to $C_1$ and $C_2$. 

Fig. 7. The intersection of ellipses, generated by projecting rotation circles, indicates multiple solutions.

Fig. 8. As the side $AB$ is swept along the constraint defined by $\theta$-$AB$, the associated rotation circles project to a family of ellipses.

partially from the self-intersecting nature of the swept surface. Although the typical left- and right-handed pairs of solutions are relatively far apart in 3-D, the "four-solution" region indicates that other pairs of solutions can be arbitrarily close in 3-D.
Fig. 9. (a) A typical simulation of the projected ellipses for an acute triangle shows (b) regions of 0, 1, 2, 3, and 4 solutions.

Fig. 10. Similar results for (a) right, (b) obtuse, (c) equilateral, and (d) acute triangles.

Fig. 10 shows the results for several triangles: acute, right, obtuse and equilateral. Although these results demonstrate that a given measurement can have a two-, three-, or fourfold ambiguity, the larger fraction of the image plane is occupied by the combined one- and two-solutions regions. This observation supports the commonly held wisdom that it is "usually" true that there is, at worst, a twofold ambiguity [9].

Notice also that the exact geometry of these regions changes in a
nonlinear fashion as the camera is rotated away from the canonical view, but the topology of the regions remains invariant. That is, although the shape of the regions changes, there always exists a homeomorphism that maps the image in the canonical view to any other view preserving the number and connectivity of the regions.

VIII. CONCLUSIONS

This work demonstrates that the fourfold ambiguity inherent to the perspective view of a triangle cannot, in general, be ignored. This work has avoided the other ambiguities that exist when the correspondence between image points and triangle vertices is not known. The six permutations of the three vertices (i.e., six assignments of the three vertices to the three image points) combined with the fourfold ambiguity discussed above can produce 24 possible camera-triangle configurations consistent with a single perspective view [7].

Many researchers have resorted to using four or more points and other features such as vertex-pairs and edge/surface elements to remove the ambiguity inherent to three points. References [6] and [20] discuss the general perspective n-point problem, and [18] discusses the triangle problem in context with vertex-pairs and rectangles. Reference [19] uses the Gaussian sphere to model the image and considers the introduction of heuristics or other domain knowledge to constrain the 3-D interpretation of measured features such as the three angles of a triangle.

Our work does not address the issue of computational efficiency, and it is not clear how these results can be put into practice, but it does provide theoretical insight into the nature of this geometric ambiguity.

ACKNOWLEDGMENT

The authors would like to thank J. Tietz, P. Van Atta, and J. Dietzler of Martin Marietta and K. Jones and J. Thomas of the University of Colorado for their support and helpful comments concerning the visualization of the triangle problem. Finally, we would like to thank the reviewers at PAMI for their perceptive comments.

REFERENCES

Trinocular Stereo Vision for Robotics
Nicholas Ayache and Francis Lustman

Abstract—We present an original approach for building a three-dimensional description of the environment of a robot using three cameras. The main advantages of trinocular versus binocular stereo are simplicity, reliability, and accuracy. We believe that these advantages now make trinocular stereo vision of practical use for many robotics applications. The technique has been successfully applied to several indoor and outdoor scenes. Experimental results are presented and discussed.

Index Terms—Computer vision, edge segments, mobile robots, stereo vision, 3-D maps, trinocular.

Manuscript received August 25, 1989; revised August 8, 1990. Recommended for acceptance by R. J. Woodham. This work was supported in part by ESPRIT Project P940.

The authors are with INRIA, BP 105, 78153 Le Chesnay Cedex, France. IEEE Log Number 9040366.

I. INTRODUCTION

Stereo vision is a technique for building a three dimensional description of a scene observed from several viewpoints. It is considered passive if no additional lighting of the scene, for instance by a laser beam, is required. So defined, passive stereo vision happens to be very attractive for many applications in robotics, including 3-D object recognition and localization as well as 3-D navigation of mobile robots.

Most of the research on passive stereo vision has been devoted to binocular vision for which two cameras are observing the same scene from two slightly different viewpoints. As soon as two image points are matched, i.e., identified as corresponding to the same physical point, it is possible to compute the three-dimensional coordinates of this physical point.

Unfortunately the matching problem is difficult. This is mainly because the geometric constraints of binocular stereo are not sufficient to impose a unique solution: several heuristic constraints must be added to compute a plausible matching solution.

Using a third camera increases the geometric constraints, and reduces the influence of heuristics in stereo-matching. Presently, following Yachida [1], [2], an increasing number of studies are devoted to trinocular vision. A review of some of these techniques can be found in [3] which includes most of the following publications [4]–[10].

For a discussion of both geometric and heuristic constraints used in binocular stereo vision and for a review of research on this topic, one can refer to [11], for example. A nonexhaustive list of publications on the subject is given by the following references [12]–[22]. Last but not least, the work of [23] on binocular stereo vision pioneered the work on trinocular stereo vision presented here, and the following references were kindly suggested by one of the reviewers [24]–[26].

The correspondence is organized as follows: first we make explicit what is needed to constrain the stereo matching problem. This includes geometry of trinocular stereo vision, representation of images, calibration, rectification, and spatial reconstruction. Then, we detail the matching algorithm and the validation procedure. Finally experimental results are presented and discussed. We conclude by a summary and future research.

II. GEOMETRY OF TRINOCULAR STEREO VISION

Fig. 1 illustrates the geometric constraints of trinocular stereo vision. Camera i (i = 1, 2, or 3) is represented by its optical center Ci and its image plane Pi. Given a scene point P, its image i by camera i is given by the intersection of the line PCi with the plane Pi. This is the classical pinhole model. Points I1, I2, and I3 form a triplet of homologous image points.