Video-Assisted Global Positioning in Terrain Navigation: Hough Transform Solution

Ha V. Le and Guna Seetharaman
Center for Advanced Computer Studies, Unive. of Louisiana at Lafayette, LA 70504-4330
hvle, guna@louisiana.edu
Harold Szu
Office of Naval Research, Code 313, Sensors & Surveil. Program, Arlington VA 22217-5660

Abstract

We present a rigorous geometric analysis of the computation of the global position of an airborne video camera using aerial images of known landmarks. This has also been known as the perspective-n-point (PnP) problem. Based on the analysis, a robust Hough transform-like method, facilitated by a class of CORDIC-structured computations, to find the solution for the problem, is shown feasible within the framework of terrain navigation. It empowers aerial surveillance systems to navigate effectively when the global position and inertial navigation sensors become faulty, inaccurate, or dysfunctional.

1. Introduction

The perspective-n-point problem is to determine the position of the camera from images of known landmarks, given the intrinsic parameters of the camera. In case there are three nonlinear landmarks, a set of three quadratic equations with three unknowns can be formed [1], that can produce as many as eight solutions. Fischler and Bolles [2] derive a biquadratic polynomial equation with one unknown, that has a maximum of four solutions. For the case of four landmarks, Rives et al. [1] find a unique solution if the landmarks are coplanar by finding the solutions using three landmarks, then verifying them with the fourth one. If the four landmarks are not coplanar, they derive a set of six quadratic equations with four unknowns. Horaud et al. [3] form a biquadratic polynomial equation in one unknown for non-coplanar landmarks. The equation can have as many as four solutions, but some of them can be eliminated. A more general approach to finding solutions for the perspective-n-point problem is to use the least-square techniques [4].

In terrain navigation, the global position of an airborne camera system is usually available through GPS. However, there are circumstances where the position given by a GPS is not accurate due to built-in errors, dysfunctional GPS, or noise caused by intentional R-F jamming. Information gathered from images of known landmarks then can be used to compensate the GPS errors in such cases [5] [6].

We now present our analysis for the problem through constructive geometry. The main result of the analysis is: for any two landmarks in an image, we can construct a unique toroid in the world space, that constitutes all possible positions of the camera. The toroid can be easily generated using a CORDIC [7] structured hardware, to register a vote at the appropriate cells of a three-dimensional array of accumulators. For example, in the case of four landmarks, six toroids could be generated, corresponding to the six quadratic equations derived by Rives et al. [1]. An acceptable solution would be identifiable at the cell with a vote of at least three and at most six. The geometric nature of the proposed approach helps understand the span of the subspace (3-dimensional world) to which the expected solution would be constrained. Thus, a finite number of accumulators are sufficient. Moreover, we use a multi-resolution recursive approach outlined in [8] for the implementation. It is envisioned that real world applications would employ recursive/adaptive Kalman predictors to estimate the GPS based on most recent positions over the flight path. Such information, if available, could accelerate the effective search for the hough-transform cells that instantiate the final solution.

2. The Analysis

2.1. Geometric Analysis

We assume that the operational nature of the imaging system follows geometric optics [9] and perspective pro-

jection. Our results are based on the classic theorem of angles on circles [10] which states: Given a circle, and a chord AB dividing the circle into two sectors. Then, the angle subtended by the arc AB at the center is exactly twice that of the angle subtended by the arc AB at any point on the circumference along the sector AB. Its corollary states that: The locus of the vertex C of all triangles with a fixed apex angle $\angle ACB$ and fixed end points A and B is a sector of a circle [10].

Theorem 1. Given an image of two landmarks and therefore the angle subtended by them at the pupil of the camera. Then, the unknown geoposition of the camera is constrained to a unique circle on the principal plane of imaging; and, it is also constrained onto the surface of a unique toroid in space when the principal plane can not be uniquely fixed.

Proof. We prove the above claim by constructing the circle and the toroid as described below.

Let the points A and B be the two landmarks observed by the camera C from an unknown geolocation. Then, the plane ACB is the principal plane as shown in Figure 1. Let θ be the angle $\angle ACB$, which can be computed from the video image. Details of the imaging model will be shown in Section 2.2. Also, let $\mathbf{n}_P$ be the normal of the principal plane ACB. And, let L be the length of the line segment AB. From the geolocations of A and B compute the midpoint M on AB. Draw the bisector of AB, the line perpendicular to AB passing through M. Locate the center O on the bisector at a distance h from M; and, compute the radius R as:

$$h = \frac{L}{2}\frac{1}{\tan\theta}; \quad \text{and} \quad R = \sqrt{\left(h^2 + \frac{L^2}{4}\right)} \qquad (1)$$

Also, $\mathbf{X}_M = 0.5\ (\mathbf{X}_A + \mathbf{X}_B)$. The direction cosines $\mathbf{l}$ of the line segment AB, and $\mathbf{p}$ that of its bisector are:

$$\mathbf{l} = \frac{\mathbf{X}_B - \mathbf{X}_A}{|\mathbf{X}_B - \mathbf{X}_A|} \quad \text{and} \quad \mathbf{p} = \mathbf{n}_P \times \mathbf{l} \qquad (2)$$

Then, trace the locus of all feasible positions of the camera C, by tracing $\mathbf{X}_C(\alpha), \left(\theta - \frac{\pi}{2}\right) \leq \alpha \leq \left(\frac{3\pi}{2} - \theta\right)$:

$$\mathbf{X}(\alpha) = \mathbf{X}_M + h\mathbf{p} + (R\cos\alpha)\mathbf{l} + (R\sin\alpha)\mathbf{p} \qquad (3)$$

Then, based on the Theorem of Thales[10] we prove that the circle is unique.

When the principal plane is not explicitly known, any plane can be used to construct the circular arc first, which is then rotated about the axis AB to create the toroidal surface. For example, consider the unit vector $\mathbf{e}_3 = \frac{1}{\sqrt{3}}(1,1,1)$. If $\mathbf{l} \neq \mathbf{e}_3$ then let $\mathbf{n}_B = \mathbf{e}_3$; else, let $\mathbf{n}_B = \frac{1}{\sqrt{3}}(-1,1,1)$. Then compute, $\mathbf{n}_P = \mathbf{n}_B \times \mathbf{l}, \mathbf{p} = \mathbf{n}_P \times \mathbf{l}$, and $\mathbf{q} = \mathbf{l} \times \mathbf{p}$. Then, for any $(\alpha, \beta), -\pi < \beta \leq \pi$ and $\left(\theta - \frac{\pi}{2}\right) \leq \alpha \leq \left(\frac{3\pi}{2} - \theta\right)$, compute $\mathbf{X}(\alpha, \beta)$ on the surface of the toroid as:

$$\mathbf{X}(\alpha, \beta) = (V_\alpha \cos\beta)\mathbf{p} + (V_\alpha \sin\beta)\mathbf{q} + U_\alpha \mathbf{l} + \mathbf{X}_M \qquad (4)$$

where

$$U_\alpha = R\cos\alpha \quad \text{and} \quad V_\alpha = h + R\sin\alpha \qquad (5)$$

This completes our proof.

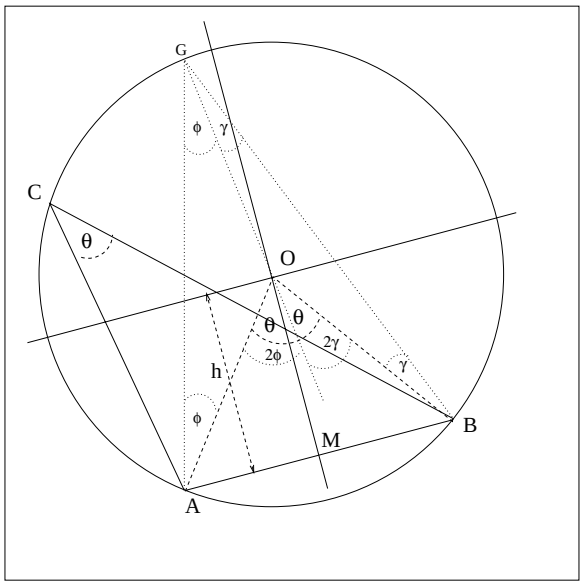

Figure 1: *Two landmarks A and B are imaged from a camera at an unknown geolocation C. The angle θ is measured from the image. Then, the circle is constructed from the value θ and the locations of A and B is unique.*

2.2. Imaging Model

The geometric model of the aerial imaging system is illustrated in Figure 2. The notation used in this paper is as follows: An object A located at $\mathbf{X}_A$, is imaged by a camera C kept from a geoposition $\mathbf{X}_C$. Perspective imaging is assumed. The image point $\mathbf{x}_A$ is on the $z = -f$ plane of the camera coordinate system (x, y, z); however, the derivations will consider $z = f$ plane [11]. The angle θ_Z and the vector $\mathbf{X}_C$ denote the instantaneous orientation and geolocation of the camera. Also, θ_A describes the line of sight of A from the current location $\mathbf{X}_C$ of the camera C. The angle ϕ_a and the vector $\mathbf{x}_a$ are used to represent the direction and position of a, the image of A, measured in the 3-D frame of reference of the camera. The uppercase subscripts are used to indicate quantities that are prone to sensor inaccuracies; and, lowercase subscripts will

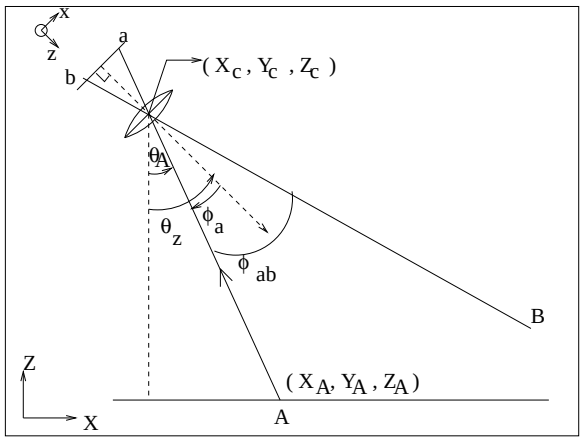

Figure 2: *Two objects A and B kept in an arbitrary locations are imaged by a video camera from $\mathbf{X}_C$. The z axis, called the optical axis, is shown in dotted lines. Also, the differential angle ϕ_{ab} is measured accurately at the image-resolution.*

be used to denote quantities which are measured using the observed image.

Lemma 1. Let A and B be two objects in the field of view of the camera C which is located at X_C. Then, the angle $\angle ACB$ subtended by the 3-D line segment AB at the pupil of the camera C can be measured accurately without any explicit knowledge of the exact geoposition and orientation of the camera.

Proof. We observe from Figure 2, that

$$\tan\theta_A = \frac{X_A - X_C}{Z_A - Z_C} \tag{6}$$

We assume that the video imaging system has already been calibrated, and its intrinsic parameters such as the location of the optical center within the given image, pixel dimensions, and focal length have been determined and made available. Also, we assume that the exact location of the point a, the image of A has been determined by standard image processing techniques at the best possible accuracy. Then, ϕ_a can be computed from the video-image based measurement of $\mathbf{x}_a$. Thus,

$$\phi_a = \arctan\left(\frac{x_a}{f}\right). \tag{7}$$

which satisfies the expression,

$$\phi_a = \theta_A - \theta_Z. \tag{8}$$

Also,

$$\begin{aligned}
\theta_A - \theta_B &= (\theta_A - \theta_Z) - (\theta_B - \theta_Z) && (9)\\
&= \phi_a - \phi_b && (10)\\
&= \arctan\tan(\phi_a - \phi_b) && (11)\\
&= \arctan\left[\frac{(x_a - x_b)f}{f^2 + x_a x_b}\right] && (12)\\
\theta_{AB} &= m_{ab}, && (13)
\end{aligned}$$

where m_{ab} is computed from x_a and x_b and treated as a single measurement derived through the image. Thus, even though X_C and θ_Z are not explicitly known, the difference $\theta_{AB} = (\theta_A - \theta_B)$ could be measured indirectly from the image. Moreover, the differential nature of the quantity θ_{AB} makes it insensitive to the orientation jitters of the camera.

3. Hough Transform based Hardware Solution

Given the image of three or more landmarks, the task is to identify the intersection of three toroids, each of which constrains the feasible camera location onto its surface.

The sequential, CORDIC-structured and real-time computability of $\mathbf{X}(\alpha, \beta)$ suggests a possibility of computing the camera location through an approach based on the Hough Transforms [13]. A CORDIC structured hardware [7] would rapidly generate the circle and the toroid as follows. For example, the value of $\mathbf{X}(\alpha)$ can be computed as follows, using a recursive algorithm for Equation (3).

$$\begin{aligned}
\alpha_0 &:= 0; u_0 := 1; v_0 := 0; && (14)\\
\begin{bmatrix} u_k \\ v_k \end{bmatrix} &= \begin{bmatrix} 1 & -\delta \\ \delta & 1 \end{bmatrix} \begin{bmatrix} u_{k-1} \\ v_{k-1} \end{bmatrix} && (15)\\
\alpha_k &= \alpha_{k-1} + \delta && (16)\\
X(\alpha_k) &= X_O + h\mathbf{p} + u_k\mathbf{l} + v_k\mathbf{p} && (17)
\end{aligned}$$

where $\delta \approx \sin\delta$ is chosen as $\delta = 2^{-8}$, to perform the computation rapidly in hardware [7]. The computational accuracy of this approach could be further enhanced by considering the *three stage lifting procedure* proposed in [14] and [15].

In this context, the vote-accumulation and search take place in the XYZ space. The XYZ space is divided into a three-dimensional grid; and each cell is assigned a vote accumulator. The unknown camera position is traced by a toroid to be determined by locating two known landmarks in the image. The vote at each bin on the surface of the toroid would be suitably increased as the toroid is being traced. When there are more than two landmarks, it is likely that all toroids constructed from all landmark-pairs would intersect at

one point indicating the most likely location of the unknown camera position. Then, the computation of the camera location becomes one that of finding the cell that has the highest number of votes.

The design principles of Breshenham[16] algorithms used in computer graphics are suitably used to generate the toroidal surface based on the CORDIC structured computations of (4). Given two landmarks $\mathbf{X}_i, \mathbf{X}_j$. The first step is to choose an arbitrary point $\mathbf{X_a}$ in the 3-D grid of the accumulators, hence a voxel in the subspace of anticipated solution. Then, the normal to the plane passing through the landmarks and the said point is computed for applying (4). It is observed that, $\mathbf{X}(\alpha, \beta)$ for any β and $\alpha = \frac{\pi}{2}$, would represent a point on the circle which is at a maximum distance from the baseline IJ. Let, $\triangle\beta$ chosen such that,

$$0 < \triangle\beta < \frac{1}{h+R} \tag{18}$$

to extend the computation of (**??**) into tracing (4) iteratively. Then, two consecutive circles drawn to trace $\mathbf{X}(\alpha, \beta)$ and $\mathbf{X}(\alpha, \beta + \triangle\beta)$, would differ by a value less than one voxel, at $\alpha = \frac{\pi}{2}$. And, for all other values of α the distance between corresponding points will be even smaller. Thus, repeated application of (4) would produce a continuous, one voxel-thick and hollow toroidal surface. In general, the discretization could become a source of inaccuracies in the final outcome. The toroids are generated by incrementing the angles α and β, whereas the search space is divided along the X, Y, and Z axes. The error analysis of such generalized of Hough transform techniques could be found in [17].

It is of practical concern to address the possibility that the $XYZ-$volume to be covered by the accumulators may be too large. This could require more computational effort to solve the original problem than the previously known methods to compute the camera position. However, this potential disadvantage must be put in context with the traditional advantages of the Hough-transform based techniques. The underlying clustering ability of the proposed approach produces the net result which is also advocated by RANSAC [2]. Moreover, in many instances of terrain navigation we have access to a quick estimate of the camera position: for examples, 1) from data provided by GPS with known maximum errors, or 2) from a suitably processed estimate based on previous GPS data before the GPS becomes faulty. Hence, the search space can be limited into a relatively small volume. Also, we employ a multi-resolution approach to speed up the computation [8].

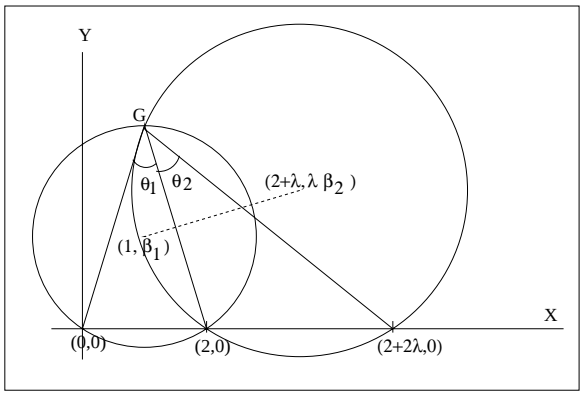

Figure 3: *Three colinear landmarks in the most canonical and parameterized configuration and the the uniquely found location G of the camera computed from θ_1 and θ_2 and λ.*

4. Special Configurations

There are special cases in which we can simplify the results of Section 2 so that faster computations can be achieved. The cases to be presented in this section are based on our observation that, if there are three colinear landmarks, then the camera position can be constrained to a unique circle in the world space.

Theorem 2. Given an image of three colinear landmarks, and the orientation of the principal plane of imaging; then, the geolocation the camera is uniquely determined.

Proof. Let L_1, L_2, L_3 be the three landmarks. By definition, we know the orientation of the principal plane and three points on the plane, hence it is uniquely fixed. Let C_{12} and C_{23} be the two uniquely defined circles constructed on the principal plane by applying Theorem 1, to the landmark-pairs L_1, L_2 and L_2, L_3 respectively.

Example 1. Consider an instance shown in Figure 2., where the landmarks are located at:$(0,0)$, $(2,0)$, and $(2+2\lambda, 0)$; $\lambda > 0$, on the X-axis; the pairwise differential angles subtended (at the camera) by the two consecutive pairs of landmarks are θ_1 and θ_2 respectively. Then, the camera position: (x_g, y_g) can be uniquely extracted as:

$$(x_g, y_g) = \left(\frac{2(1 - m\beta_1)}{m^2 + 1}, \frac{2m(1 + \beta_1)}{m^2 + 1} \right) \tag{19}$$

where, $\beta_1 = \cot\theta_1, \beta_2 = \cot\theta_2$ and, $m = \frac{\lambda\beta_2 - \beta_1}{1+\lambda}$. In a more general case we assume the landmarks to be at: $(0,0), (d_2, 0), (d_3, 0)$. Then, we let $\lambda = \frac{(d_3 - d_2)}{d_2}$ and scale the final solution by a factor of: $d_2/2$.

Theorem 3. Given an image of three colinear landmarks, then the geolocation of the camera is con-

strained to a circle on a plane perpendicular to the line passing through the landmarks; and, both the circle and the plane are unique.

Proof. Let L_1, L_2 and L_3 be the landmarks; and, let $\mathbf{l}$ be the direction cosine of the line L_{123} passing through these landmarks. Let $\mathbf{n}_P$ be the normal of a plane passing through the colinear landmarks. One such normal can be uniquely identified following the procedure used in the proof of Theorem 1; also $\mathbf{p} = \mathbf{n}_P \times \mathbf{l}$. Then, let G be the geolocation of the camera uniquely resolved by applying the second theorem in this context. Also, we apply the insight gained from the canonical Example 1 above. Then, in our case: $d_2 = L_{12}$ and $d_3 = L_{13}$ are the pairwise distance of the landmarks L_2 and L_3 respectively from L_1. Then, $\mathbf{X}_G = \mathbf{X}_{L_1} + (sx_g)\mathbf{l} + (sy_g)\mathbf{p}$, and, $s = \frac{d_2}{2}$ is the scale factor for applying the results of the canonical results from Example 1. Now, rigidly rotate the entire plane around the axis L_{123} by 2π radians. It will produce a circle of radius sy_g contained in a plane perpendicular to L_{123}; and, the plane is at a distance sx_g from the landmark L_1. Since G is unique and L_{123} is fixed, both the circle and the plane of the circle are unique. And, hence the claim.

Consider another special case where there are four coplanar landmarks, and no three colinear. We extend the result from Theorem 3 to show that, with these four coplanar landmarks, we can construct two circles in the world space. Note that these two circles can intersect at no more than two points, and in that case one of the intersections will be on or below the ground plane, as shown in Figure 4. It means we can derive an analytic solution for the case of four coplanar landmarks from our constructions. However, in reality the inaccuracies introduced due to discrete sampling of the images with finite pixel size may result in a situation where these circles may not intersect. Obviously, our Hough transform-like approach provides a robust technique to accommodate that.

The most canonical form occurs when three non colinear land marks are observed. Then, the unknown position of the camera and the three landmarks form a tetrahedron, with the camera being located at its apex. The search for the 3-D position of the apex, is effectively reduced to solving the orthocenter of the triangle formed by three points on the ground plane which correspond to the final position of the apex when each of the side (two landmarks and the apex) is unfolded. We have been able to solve this problem using a 1-D search. Simulations have resulted only two real solutions. This is in accordance with [2].

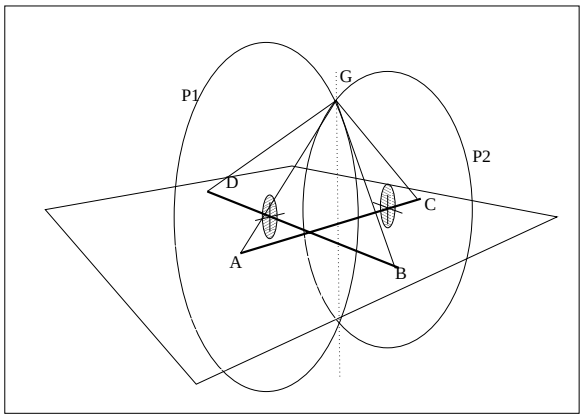

Figure 4: *Four coplanar landmarks in a canonical configuration are imaged by a camera at G; two circles are computed from the two sets of diagonals of the quadrilateral ABCD. The chord of intersection of the circles, shown in a dotted line, is perpendicular to the ground plane.*

5. Experimental Results

We conducted a set of experiments to validate the basic concepts. The tests were carried out on a Newport-Optics Optical Test Bench of size $9' \times 3' \times 6'$ fitted with 4 fully calibrated video cameras. The test scene was made of points in the three-dimensional space. We measured the location of these points and treated them as landmarks; and, acquired their image from a camera, whose precalibrated location was treated as the unknown geolocation system. These results will be presented at the conference.

Based on the experimental and simulation results, the following observations can be made.

1. The focal length of the camera should be short to obtain more reliable results.
2. The farther the distance between two landmarks, provided they can both appear in an image frame, the more reliable the measurement.
3. The closer the landmarks to the camera, the more accurate the results.

We found that our observations agree in general with the results of Weng at al. [18] and Lee [19]. Details on the algebraic methods for error estimation and analysis are discussed in [18].

6. Conclusions

A rigorous geometric analysis of visual landmark based terrain navigation has been presented. The new insights introduced in this paper may be summarized as:

1) an image of two landmarks helps to constrain the unknown geoposition of the aerial observer onto the surface of a unique toroid; 2) an image of three colinear landmarks reduces the same onto a uniquely defined circle, which is also perpendicular to the axis passing through the landmarks. We have also presented a Hough transform-like approach, facilitated by a class of CORDIC-structured computations based on the results of our analysis, that provides a fast and robust technique for compensating the GPS errors in locating the position of an aerial camera system.

Acknowledgement. This work was partially supported by the National Science Foundation, under Grants NSF-9210926, and NSF9612769. The authors wish to thank David Liese of NAVSEA, US Navy, and Scott Norton of Defense Technology Inc. VA.

References

[1] P. Rives, P. Bouthemy, B. Prasada, and E. Dubois. Recovering the Orientation and the Position of a Rigid Body in Space from a Single View . Technical report, INRS-Telecommunications, 3, place du Commerce, Ile-des-Soeurs, Verdun, H3E 1H6, Quebec, Canada, 1981.

[2] M.A. Fischler and R.C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, Vol. 24 (No. 6):381–395, Jun. 1981.

[3] R. Horaud, B. Conio, O. Leboulleux, and B. Lacolle. An Analytic Solution for the Perspective 4-Point Problem. In *Computer Vision and Pattern Recognition, 1989. Proceedings CVPR'89., IEEE Computer Society on*, 1989.

[4] D. Lowe. Three-dimensional Object Recognition from Single Two-dimensional Images . *Artificial Intelligence*, Vol. 31:355–395, 1987.

[5] G.B. Chatterji, P.K. Menon, and B. Sridhar. GPS/Machine Vision Navigation System for Aircraft. *IEEE Trans. on Aerospace and Electronic Systems*, Vol. 33 (No. 3):1012–1025, Jul. 1989.

[6] B. Sridhar and A.V. Phatak. Analysis of Image-Based Navigation System for Rotocraft Low-Altitude Flight. *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. 22 (No. 2):290–299, Mar./Apr. 1992.

[7] H.M. Ahmed, J.M. Delosme, and M. Morf. Highly Concurrent Computing Structures for Matrix Arithmetic and Signal Processing. *Computer*, Vol. 15 (No. 1):65–82, Jan. 1982.

[8] J. Illingworth and J. Kittler. The Adaptive Hough Transform. *IEEE Trans. on Pattern Ananlysis and Machine Intelligence*, Vol. 9 (No. 5):690–698, Sep. 1987.

[9] Sidney Cornbleet. Geometrical Optics Reviewed: A new light on an old subject. *Proceedings of The IEEE*, Vol. 71(No. 4):471–502, April 1983.

[10] W. Gellert, S. Gottwald, M. Hellwich, H. Kastner, and H. Kustner. *The VNR Concise Encyclopedia of Mathematics*. Van Nostrand Reinhold, NY, 2nd edition, 1989.

[11] Guna Seetharaman, Harold Szu, Biju Nair, and Scott Norton. Passive Surveillance from Video image-sequences and imprecise GPS and INS sensors. In *Proceedings of The International Workshop on Information and Decision Fusion, Washington DC*, 1996.

[12] H. Szu and R. Hartley. Nonconvex Optimization by Fast Annealing. *Proc. of The IEEE*, Vol. 75:1538–1540, Nov. 1987.

[13] P.V.C. Hough. Method and Means for Recognizing Complex Patterns, U.S. Patent 3069654. 1962.

[14] W. Sweldens. The Lifting Scheme: A custom-design construction of biorthogonal wavelets. *Appl. Comput. Harmonic Analysis*, 3:186–200, 1996.

[15] J. Liang and T. D. Tran. Fast multiplierless approximation of the DCT with the lifting scheme. *IEEE Trans. on Signal Processing*, 49:3032–3044, Dec 2001.

[16] Roger Williams. *Procedural Elements in Computer Graphics*. McGraw Hill Publishers, 1980.

[17] D.H. Ballard. Generalizing the Hough Transform to Detect Arbitrary Shapes. *Pattern Recognition*, Vol. 13 (No. 2):111–122, 1981.

[18] J. Weng, T.S. Huang, and N. Ahuja. Motion and Structure from Two Perspective Views: Algorithm, Error Analysis and Error Estimation. *IEEE Trans. on Pattern Ananlysis and Machine Intelligence*, Vol. 11 (No. 5):451–467, 1989.

[19] C.H. Lee. Time Varying Images: The Effect of Finite Resolution on Uniqueness. *Computer Vision, Graphics, and Image Processing*, Vol. 54 (No. 3):325–332, 1992.