1997 IEEE International Conference on Robotics and Automation

Image-based Prediction of Landmark Features for Mobile Robot Navigation

Gregory D. Hager, Computer Science
David Kriegman, Electrical Engineering
Erliang Yeh, Electrical Engineering
Christopher Rasmussen, Computer Science

Yale University, New Haven, CT 06520

Abstract

We have been developing an architecture for vision-based navigation which relies on continuous feedback from visual "landmarks" to control robot motion. In this approach, landmarks are consistently located and acquired as they come into view. To make this process efficient and robust, it is important that the image locations of these features can be predicted *from available image information. In this article, we discuss methods for direct image-based prediction of point and line features for a mobile system operating on a planar surface. Preliminary experimental results suggest that image-based prediction can be performed efficiently and with sufficient accuracy to ensure robust acquisition of navigational landmarks.*

1 Introduction

Most prior research on robot navigation has focused on developing methods for computing and/or controlling the position of a mobile system with respect to some geometric or topological map [1, 3, 9, 11, 12, 13, 14, 16]. However, there are applications where the ability to explicitly represent and reason about geometry is not essential. For example, consider a robot assigned to follow a routine delivery route on a flexible manufacturing floor, or a robot sentry assigned to patrol the perimeter of a recently established military encampment. In principle, there is no need to develop an extensive navigation system for these functions—the robot simply needs to constantly and reliably repeat the same path on a continuous basis. The system can be "programmed" by simply "showing" it the path that it should follow.

This problem offers a number of challenges to many published approaches to robot navigation. Given that the environment and the allowable paths through it may may change often and abruptly, it is unlikely that a prior geometric model will be available, eliminating approaches which rely on them [1, 11, 12]. For similar reasons, "situated" approaches which implicitly use strong assumptions about the environment are also inapplicable [3, 9]. While it is possible to arrange fiducial markers for such a task, it is onerous to erect and calibrate such markers. Finally, the fact that the robot may need to navigate accurately in large open areas suggests that topological approaches based on place recognition and approaches relying on range-limited sensing such as sonar [13, 14] may have difficulty supporting accurate and reliable motion.

Our aim is to develop a vision-based navigation system capable of performing these types of tasks. Our choice of vision is based on the fact that it can observe large areas to find useful "landmarks" for defining the path, and our desire to use passive sensing techniques. In a previous paper [6], we outlined the general architecture of such a vision-based navigation system. The central idea in this design is to constantly track image features used as landmarks for navigation. Such tracking is cheap and simple [7]; it quickly and continually reduces image information to the time history of a small set of feature locations. This time history is learned once (the teaching phase), and subsequent motion is defined by controlling the robot so as to replicate the learned feature time history. This provides constant and accurate control of position, yet avoids the overhead of computing an explicit geometric model of the environment.

One central component of this architecture is the ability to effectively predict and acquire landmarks as they come into view. Given a geometric model, prediction could be handled using a well-understood combination of odometry and estimation [11, 12]. In a purely sensor-based approach, prediction must be performed solely on image information. This problem is closely related to the *image transfer problem* discussed in the area of projective geometry applied to vision [2, 5, 8, 15]. In this paper, we describe specializations of image transfer methods for the two commonly used image features, point-features and line features, for a mobile system operating on a planar surface.

2 Problem Formulation

In an image- or appearance-based approach to navigation, the robot can be said to be at a place if what it sees corresponds, with some tolerance, to what can be seen at this place. Given that we know which location we are at from its appearance, we can use *visual tracking* to extend the definition of place to a range of locations, and require that all (or nearly all) of the world corresponds to some place. We then utilize the continuity of our representation to *predict* changes of view between places, thereby eliminating the need for a strong notion of recognition. Navigation is posed as the problem of moving from place to place—that is, from view to view—using techniques developed in the area of visual servoing [4, 10].

In order to make this problem precise, we define the following terms (detailed more fully in [6]). We use the term *marker* to denote any visual entity that is in some way visually distinctive so that it can be tracked as the robot moves. A *scene* is a set of markers which are tracked concurrently in an image. A *sequence* is an ordered list of scenes containing the same set of markers. A *map* is a directed graph of sequences. In operation, the robot first "learns" a series of sequences (a map) by discovering, tracking, and recording the motion of visual features as it moves. Later, in order to move from its current location to a specified goal, the robot "follows" a series of stored sequences or partial sequences in the map, all directed head to tail, from its current scene to the chosen goal scene.

Central to this approach is the idea that the transition from one sequence to the next, a process which involves determining the image locations of a new set of markers, can be accomplished using only image information. This motivates the problem considered in this paper:

> Given the locations of n markers tracked through a sequence of images during learning, and the locations of $m < n$ of the same markers in an image acquired during navigation, predict the location of the remaining $n - m$ markers.

Specifically, the n markers are those needed to initiate tracking for the next sequence in the path. The m known markers are those common to both the current sequence and the next sequence — in other words, the markers whose location is already known. Our goal is to establish a lower bound on m as well as to describe algorithms for performing the prediction for different types of markers.

Within this paper, we simplify the problem by only considering the locations of the markers in two images within the sequence since this is sufficient for the prediction task. Thus, the information at our disposal is two scenes with n markers (the map images) taken at two different locations in the world, and a third scene (the current image) which contains a subset of m markers. We are to predict the location of each of the remaining $n - m$ markers.

The problem of predicting the location of markers or features *without* explicitly reconstructing their 3-D Euclidean location is known in photogrammetry as *image transfer* [2]. Using projective geometry and projective invariance, methods for performing image transfer have been developed for point and line features under a variety of assumptions about the configurations of the features, the camera model (orthographic projection, affine, perspective, projective), and availability of camera calibration [2, 5, 8].

For point features, Barrett *et. al.* show that linear methods can be used to transfer points from two map images to a third image if eight additional points are observed in all three images [2]. Hartley's methods for projective reconstruction of lines can be used for line transfer, and it requires observing thirteen lines in all three images [8]. In both of these general methods, the camera can be at an arbitrary 3-D position and orientation, and the camera may have different calibration parameters at each location. However, transfer can be accomplished with many fewer points or lines if the motion of the camera is constrained, if the camera parameters at each location are identical, and if some of the calibration parameters are available.

In particular, we will consider a mobile robot moving on a plane (the *ground plane*) with a camera whose optical axis is parallel to the ground plane. We will model the camera as pin-hole perspective, and without loss of generality assume unit focal length. We define a world coordinate system centered at the optical center of the first camera with the x and y axes spanning the image plane and the z-axis aligned with the optical axis. Now, consider three camera locations denoted by coordinate frames $0, 1$ and 2. For planar robot motion, the rotation matrix R and translation vector T between frames 0 and 1 take on the form

$$R = \begin{bmatrix} \cos\alpha & 0 & -\sin\alpha \\ 0 & 1 & 0 \\ \sin\alpha & 0 & \cos\alpha \end{bmatrix}, \quad T = \begin{bmatrix} T_x \\ 0 \\ T_z \end{bmatrix}. \quad (1)$$

Similarly, a rotation matrix S and translation vector U characterizes the relation of frames 0 and 2.

As the camera moves, the height of the camera center remains constant; the plane through the moving camera center is parallel to the ground plane and is

called the *horizon plane*. For any camera position, every point on the horizon plane projects to a single line in the image named the *horizon line* (the intersection of the horizon plane with the image plane).

3 Prediction

We now develop transfer methods for two types of markers: point-like markers and line-like markers.

3.1 Points

Equation and unknown counting can be used to show that a minimum of two point features are required to perform prediction. This leads to a set of quartic equations to be solved. Unfortunately, we have observed that the results are unacceptably noisy, perhaps due to inaccuracies in the camera calibration coupled with image noise propagating through nonlinear equations. Instead, we have developed a simpler method based on a specialization of the method by Barrett *et. al.* [2]. For a particular point $\mathbf{x}$ whose coordinates in the world frame (and the frame of camera 0) are $(x, y, z)^t$, the image coordinates (p, q) for camera 1 can be expressed as:

$$R(\mathbf{x} - T) = (x', y', z')^t,$$

and

$$(p - p0) = x'/z' \quad (q - q0) = y'/z'$$

where $(p0, q0)$ is the center of projection in internal camera coordinates. When the camera is fully calibrated, R has the form in (1). When it is uncalibrated, the cosines and sines become arbitrary numbers.

These expressions can be rearranged into a system of equations that are linear in the homogeneous coordinates of $\mathbf{x}$,

$$\begin{pmatrix} \mathbf{a}_1 \\ \mathbf{b}_1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \mathrm{diag}(p, q) \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}.$$

Denoting the image coordinates of the corresponding feature in image 0 by $(\bar{p}, \bar{q})$, the projection equations for camera 0 can be written similarly. Combining the two systems of equations, we arrive at a homogeneous equation of the form

$$[\mathbf{A} - \mathrm{diag}(p, q, \bar{p}, \bar{q})\mathbf{C}]\mathbf{X} = 0.$$

where $\mathbf{X}$ are the homogeneous coordinates of $\mathbf{x}$. A necessary condition for there to be a nontrivial value of $\mathbf{X}$ is that

$$|\mathbf{A} - \mathrm{diag}(p, q, \bar{p}, \bar{q})\mathbf{C}| = 0 \tag{2}$$

where $|\cdot|$ denotes the matrix determinant.

Expanding the determinant and grouping terms yields an equation involving sixteen unique combinations of observables multiplied by coefficients. These coefficients can be expressed as determinants of matrices composed of rows from $\mathbf{A}$ and $\mathbf{C}$. In the general case, it is shown in [2] that the first seven of these coefficients are zero. The remaining coefficients are multipliers of the terms $(p\bar{p}, p\bar{q}, q\bar{p}, q\bar{q}, p, q, \bar{p}, \bar{q}, 1)$. It follows that only eight points can have independent vectors of this form, and hence the determinant of a matrix composed of nine such point pairs vanishes. This provides a linear constraint on the values of the observations of these points [2] .

For the case of planar motion, we can specialize this result and show that the coefficients of $p\bar{p}, q\bar{q}$ and $\bar{p}$ are zero. This would mean that six points determine an invariant relationship. Suppose that we also know the image coordinates of the center of projection. Setting $p0$ and $q0$ to zero causes the coefficients of p and 1 to drop out. Hence, (2) becomes a linear homogeneous equation in $\mathbf{b} = (p\bar{q}, q\bar{p}, q, \bar{q})$.

As a result, given two views of four points, we can predict the location (s, t) of a fourth point in a third view from the other three points. Two linear equations in (s, t) can be established and solved. To set up these two equations, a 4×4 transfer matrix can be established with the four rows composed of the vector $\mathbf{b}$ given above computed for four distinct points. Let the first row correspond to the point being transferred, and let us decompose this matrix as

$$\begin{bmatrix} qs & pt & t & q \\ \mathbf{d}_1 & \mathbf{d}_2 & \mathbf{d}_3 & \mathbf{d}_4 \end{bmatrix}$$

where each $\mathbf{d}_i$ is a 3 element column vector, (p, q) is the location of the point in the first image, and (s, t) is the unknown coordinates in the third view. Expanding the determinant in terms of the minors of the matrix yields the following linear equation in (s, t)

$$\begin{aligned} &q|\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3| - q|\mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4|s + \\ &(p|\mathbf{d}_1, \mathbf{d}_3, \mathbf{d}_4| - |\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_4|)t \quad = 0. \end{aligned} \tag{3}$$

The second training view yields a second linear equation in (s, t), which can be readily solved. Since this result is independent of the scale of the image coordinate, it holds when the focal length or internal scaling parameters are unknown.

3.1.1 Degeneracies

The degeneracies of this solution can be determined by analyzing when the linear system defined above looses

rank. This analysis was carried out using Mathematica, and it was found that the singularities of the linear system can be reduced to the following three cases:

1. One or more points are in the horizon plane and project to the horizon line. Since the horizon line is invariant under the allowed camera transformations, no information is obtained from the y coordinate of the feature.
2. If the three camera centers are collinear, the pencils of epipolar planes for all pairs of camera positions are identical. Therefore, the two equations in (s,t) given by (3) become linearly dependent. Note this includes the case when the camera does not move, but only rotates.
3. The final condition can be expressed as the vanishing of a determinant:

$$\begin{vmatrix} x_1 & z_1 & 1 \\ x_2 & z_2 & 1 \\ x_2 & z_3 & 1 \end{vmatrix} = 0 \qquad (4)$$

   This determinant vanishes when the projections of the three points onto the ground plane (the x-z plane) are collinear, i.e. when the three points lie on a vertical plane.

The first condition is easily avoided when choosing markers to include in the sequence, and hence never arises in practice. In the case of the last two conditions, it is still possible to compute the line in the image that the transferred point must lie on, hence the constraint still provides useful information for constraining marker search.

3.2 Lines

In this section, we consider the problem of using image transfer to predict the location of a line in an image under perspective projection, assuming that the camera is constrained to move in a plane. Recall that a line in a plane can be represented as a point in $\mathbb{P}^2$ whereas a 3-D line can be characterized as a point on a 4-D manifold. Thus, the problem of line transfer differs from that of point transfer; they are not duals. Note also that we are *not* using information about the location of the endpoints of measured line segments. Often the endpoints are difficult to localize because edge detectors break down near corners. Furthermore, when the endpoint is a t-junction, the images of the endpoint may not correspond to the same 3-D point because of occlusion.

Again, equation and unknown counting can be used to show that a minimum of four lines must be measured in three images to transfer a fifth line; however, the resulting system of equations is again highly nonlinear. Here we will consider a linear method to predict the location of a seventh line from six corresponding lines in three images and the seventh line observed in the two map images.

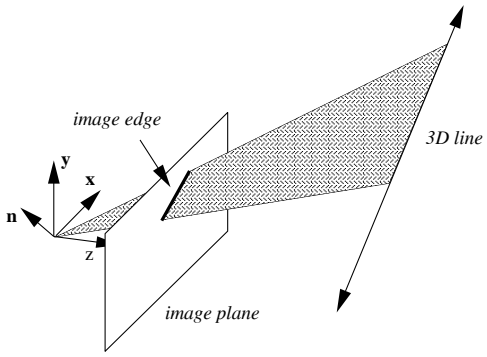

Figure 1. The image of a 3-D line can be represented by a normal $\mathbf{n}$ to the plane spanned by the line and camera center.

Expressing the rotation matrices in terms of the column vectors $R = [R_1 \mid R_2 \mid R_3]$ and $S = [S_1 \mid S_2 \mid S_3]$, we can define the following:

$$\begin{aligned} E &= R_1 U^T - T S_1^T, \\ F &= R_2 U^T - T S_2^T, \\ G &= R_3 U^T - T S_3^T. \end{aligned}$$

Taken together, these three 3×3 matrices form the so-called trifocal tensor. For point and/or line correspondences, methods have been developed for estimating the trifocal tensor when the cameras are uncalibrated and in arbitrary locations [8, 15].

Now, consider the image of a 3-D line as shown in Figure 1. The 3-D line and the center of projection define a plane. In turn, the intersection of this plane with the image plane defines a line which can be measured in the image. Conversely, from a line segment measured in an image, the normal to the plane can be determined. If image coordinates are (p,q) and the line's equation is $ap + bq + c = 0$, the coordinates of the normal in the camera frame are $\mathbf{n} = (a,b,c)$.

For three images of the same line, it is shown in a paper by Weng, Huang and Ahuja [18] that the normals to the corresponding planes are related by:

$$\mathbf{n}_2 \times \begin{bmatrix} \mathbf{n}_0^T E \mathbf{n}_1 \\ \mathbf{n}_0^T F \mathbf{n}_1 \\ \mathbf{n}_0^T G \mathbf{n}_1 \end{bmatrix} = \mathbf{0} \qquad (5)$$

where $\mathbf{n}_i$ denotes the normal to the plane defined by a line and camera center i. Hartley observes that for three uncalibrated cameras, this same set of equations

holds where E, F and G are arbitrary 3×3 matrices forming the trifocal tensor [8].

While the cross product in (5) defines three scalar equations, only two of these are linearly independent. Furthermore, the elements of the 3×3 matrices E, F and G enter linearly in this constraint. Five entries of E, F and G are constant because of the constrained camera motion, and so (5) can be expressed as a linear equation in only 12 unknowns rather than 27 unknowns. Therefore, given six lines measured in three views, we can define 12 homogeneous equations in $\mathbf{u} = (U_x, U_x, T_z, T_y, F_{11}, F_{31}, F_{13}, F_{33}, E_{11}, E_{31}, E_{13}, E_{33})$ which can be expressed in matrix form as $K\mathbf{u} = 0$ where K is a 12×12 matrix whose elements are functions of the image measurements ($\mathbf{n}_1, \mathbf{n}_2, \mathbf{n}_3$ for each line). $\mathbf{u}$ must lie in the kernel of K which is generically one dimensional except for degenerate motions or line configurations. This equation can be solved for the elements of E, F, U and T. To transfer a seventh line measured in two images and represented by $\mathbf{n}_0$ and $\mathbf{n}_1$, we note from (5) that

$$\mathbf{n}_2 = \begin{bmatrix} \mathbf{n}_0^T E \mathbf{n}_1 \\ \mathbf{n}_0^T F \mathbf{n}_1 \\ \mathbf{n}_0^T G \mathbf{n}_1 \end{bmatrix}. \tag{6}$$

3.2.1 Degeneracies

Like point transfer, there are degeneracies for line transfer. For the same reasons, transfer is impossible when two camera centers are coincident or when the three camera centers are collinear. Like the third degeneracy mentioned above, transfer is not possible if one of the lines lies in horizon plane. As noted in [18], there are other degeneracies for structure from motion from straight lines that are likely to apply to this case of image transfer under constrained motion.

3.3 Relaxing Assumptions

When formulating the transfer problem in Section 2, we assumed that the camera was neither tilted nor rotated about the optical axis; (i.e. that the optical axis is parallel to the ground plane, and the camera's y axis is pointing vertically). It is straightforward to use information derived from images to handle tilted or rotated cameras. For a moving camera, the focus of expansion (FOE) must lie on the horizon line. From a sequence of images from a moving camera, the motion of the FOE can be used to estimate the horizon line. Under our assumptions, the horizon line should be aligned with x axis in the image and pass through the principal point. Now, if the camera is tilted and rotated about the optical axis, the horizon line will be tilted and will not pass through the principal point. However, from the horizon line estimated using the motion of the FOE, it is possible to determine a transformation from image coordinates to a frame corresponding to a virtual camera that meets our assumption. The coordinates of all markers would then be transformed to the virtual camera frame, and the transfer methods of Sections 3.1 and 3.2 could be applied.

Image	Point Error Pixels	Line Error in Degrees		
		Line 1	Line 2	Line 3
c	(28.9, -5.6)	1.25	.72	.71
d	(9.2, -3.1)	1.85	1.12	1.46
e	(-12.2, -0.3)	.74	1.12	.93
f	(-6.7, -1.6)	1.29	.64	.11
g	(-13.9,-0.2)	1.07	.88	1.14
h	(-13.9, -1.6)	1.27	.48	.81

Table 1. A summary of the errors for point and line transfer for the examples in Figures 2 and 3. The point error is measured in pixels, and the line error is measured in degrees between the predicted and measured normal.

4 Experiments

The transfer methods described above have been tested on video images of a typical cluttered laboratory environment. In each case, two images were acquired and used to "train" the system by selecting a set of features by hand. In a series of subsequent images, a subset of the features were tracked, and the locations of one or more features were predicted from the information provided by tracking.

Figure 4 shows two training images (labelled a and b). Figure 4.b was taken about three feet to the left of Figure 4.a, looking in the same direction. The remaining images were taken at equal intervals as the robot moved forward approximately ten feet from its position in Figure 4.b. During training the image coordinates of all four features marked in Figure 4.a were known initially, then tracked to their positions in Figure 4.b. The crosses in the subsequent six images indicate the robot's predictions of the feature location. Table 1 shows errors in units of pixels for each image. The mean error magnitude of the x coordinate is 14.1 pixels, and for the y coordinate it is 2.1 pixels; the mean distance between the predicted and actual location of the point is thus about 14.3 pixels.

Figures 3.a and 3.b show two images of a laboratory scene with nine line segments that were detected using an implementation of the Canny edge detector. Six of these segments along with corresponding lines detected in a third image were used to estimate E, F and G. From these estimates, the location of three ad-

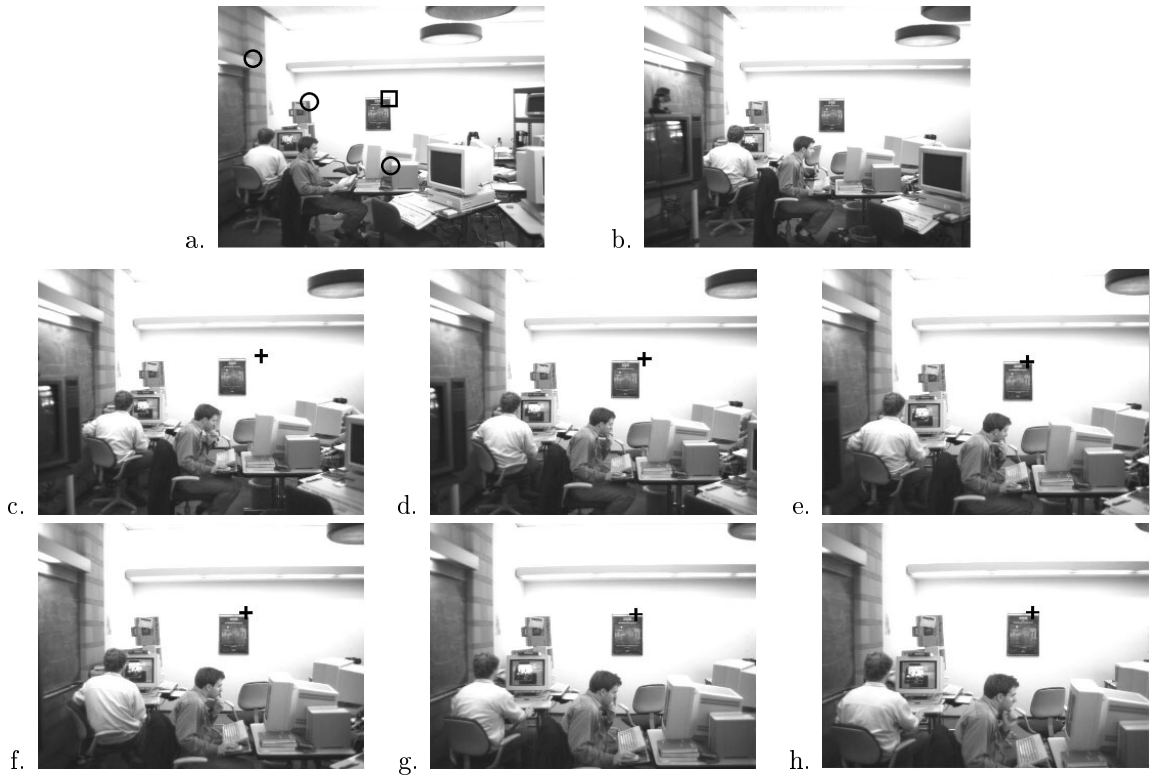

Figure 2. The location of four features (circles and box) are shown in the upper two "training" images. In the subsequent six images, the predicted location of one of the features is indicated by a cross.

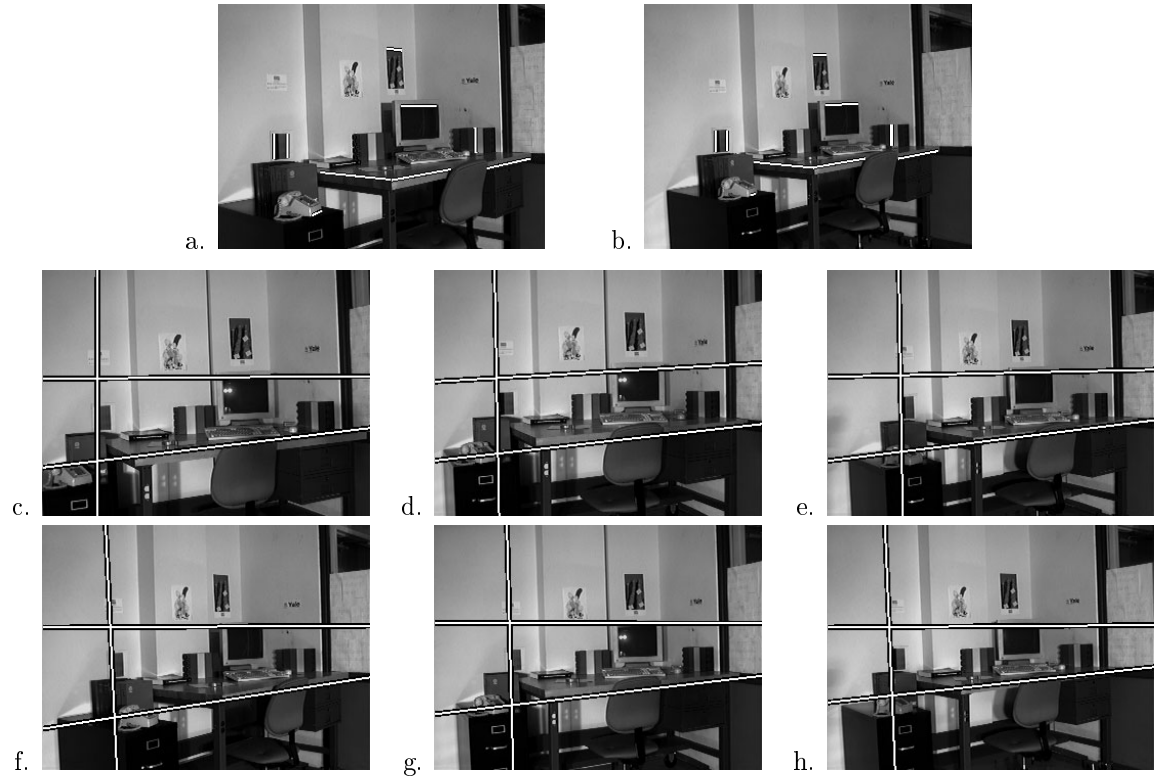

Figure 3. Nine line segments are shown in the upper two "training" images. Six of these are used to estimate E, F and G, and subsequently transfer the other three lines. In the subsequent six images, the predicted locations of the three lines are shown.

ditional lines were predicted. The remaining images in Figure 3 show the predicted location of the three lines in six images. The error between the predicted normal $\mathbf{n}_2$ from (6) and the measured normal can be expressed as the angle between these two vectors. Table 1 presents the error for all 18 predicted lines.

5 Discussion

As noted in the introduction, this work is part of a larger approach to navigation that is based on active vision and prediction. The results on transfer in this paper suggest that image-based prediction can be used reliably for transfer when combined with a search mechanism for matching previously observed markers to the new image.

Our current work is progressing toward integrating prediction with image search routines and image-based control, and in unifying the framework to support mixtures of point and line features. Recent work by Hartley has shown how a line matched in three images provides four constraints on the trifocal tensor while a point provide six constraints on the trifocal tensor. In our case, there are further constraints on the elements this tensor. We are still experimenting with various possibilities for using prediction for image-level search. Efficient search and image-level matching is clearly important, particularly when the prediction equations are singular or nearly so. We plan to compute the covariance of the predicted feature locations, and to use this to generate a search region. This formulation is particularly appealing as it allows for a unified formulation for singular and non-singular systems.

Prediction can also be used to control the motion of the robot. For example, it may be the case that the predicted image location is outside the image plane, suggesting the robot must pan the camera in order to acquire it. Likewise, the variance of the prediction can be used to control motion to improve prediction.

Acknowledgments This research was supported by ARPA grant N00014-93-1-1235, Army DURIP grant DAAH04-95-1-0058, by National Science Foundation grant IRI-9420982, and by funds provided by Yale University. D. Kriegman was funded in part under NSF Young Investigator award NYI IRI-9257990.

References

[1] S. Atiya and G. Hager. Real-time vision-based robot localization. *IEEE Trans. Robotics and Automation*, 9(6):785–800, 1993.

[2] E. Barett, M. Brill, N. Haag, and P. Payton. Invariant linear methods in photogrammetry and model matching. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 277–292. MIT Press, 1992.

[3] S. P. Engelson and D. McDermott. Error correction in mobile robot map learning. In *Proc. IEEE Int. Conf. Robotics and Automation*, Nice, France, May 1992.

[4] B. Espiau, F. Chaumette, and P. Rives. A New Approach to Visual Servoing in Robotics. *IEEE Trans. Robotics and Automation*, 8:313–326, 1992.

[5] S. Fairley, I. Reid, and D. Murray. Transfer of fixation for an active stereo platform via affine structure recovery. In *Int. Conf. Computer Vision*, pages 1100–1105, 1995.

[6] G. Hager and C. Rasmussen. Robot navigation using image sequences. In *Proc. AAAI*, pp. 938–943, 1996.

[7] G. Hager and K. Toyama. XVision: Combining image warping and geometric constraints for fast visual tracking. In *Proc. ECCV*, pp. 507–517. Springer Verlag, 1996.

[8] R. Hartley. A linear method for reconstruction from lines and points. In *Int. Conf. Computer Vision*, pp. 882–887, 1995.

[9] I. Horswill. Polly: A vision-based artificial agent. In *Proc. AAAI*, pp. 824–829, 1992.

[10] S. Hutchinson, G. Hager, and P. Corke. A tutorial on visual servo control. *IEEE Trans. Robotics and Automation*, 12(5):651–670, 1996.

[11] A. Kosaka and A. Kak. Fast vision-guided mobile robot navigation using model-based reasoning and prediction of uncertainties. *CVGIP: Image Understanding*, 56(3):271–329, 1993.

[12] D. J. Kriegman, E. Triendl, and T. O. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Trans. Robotics and Automation*, 5(6):792–803, December 1989.

[13] B. Kuipers and Y. Byun. A robot exploration and mapping strategy based on a semantic hierarchy of spatial representations. *Robotics and Autonomous Systems*, 8:47–63, 1991.

[14] J. L. Leonard and H. F. Durrant-Whyte. Mobile robot localization by tracking geometric beacons. *IEEE Trans. Robotics and Automation*, 7(3):376–380, June 1991.

[15] A. Shashua and M. Werman. Trilinearity of three perspective views and its associated tensor. In *ICCV95*, pp. 920–925, 1995.

[16] C. Taylor and D. Kriegman. Algorithms for vision-based exploration. In K. Goldberg, D. Halperin, J. Latombe, and R. Wilson, editors, *The Algorithmic Foundations of Robotics*. A. K. Peters, Boston, MA, 1995.

[17] S. Ullman and R. Basri. Recognition by a linear combination of models. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 13:992–1006, 1991.

[18] J. Weng, T. Huang, and N. Ahuja. Motion and structure from line correspondences: Closed-form solution, uniqueness, and optimization. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 14(3):318–336, March 1992.