Proceedings of the 2002 IEEE
International Conference on Robotics & Automation
Washington, DC • May 2002

# Plane Based Relative Structure Recovery

P. Ravindran, and N. J. Ferrier*
Robotics Laboratory, University of Wisconsin, Madison, WI, USA 53706
prabu@robios6.me.wisc.edu ferrier@robios6.me.wisc.edu

## Abstract

*In this paper we provide a novel linear algorithm to estimate the direction of translation and the normal to the scene plane, given a stereo pair with known correspondences between the corners in the images. Planar parallax which arises naturally in planar scenes( and can also be defined with respect to virtual planes), is exploited to achieve this. Based on the parallax vectors and the homography induced by the scene plane we calculate the direction of translation and the normal to the scene plane. We also calculate the relative heights of points on the object with respect to the scene plane, given the height of the camera above the plane. The accuracy of the estimated normal with respect to noise in input data is also analyzed empirically.*

**Keywords:** Stereo Vision, 3D Reconstuction, Plane+Parallax, Homography/Collineation

## 1 Introduction

The presence of planes in a scene provides strong visual cues that can be exploited by mobile robots navigating indoor environments. A robot can use knowledge of the height of its "eyes" with respect to the ground plane, along with planar parallax induced by object points to determine the height of points( relative to the camera) above the ground. The recovered structure can be used for obstacle avoidance, navigation, object mapping and manipulation. Nature has exploited this relationship too. Frogs use their relationship with the ground plane( and the height of their eyes above the plane) to estimate the distance to prey on the plane [2]. By artificially elevating frogs above their prey, they snapped short of their prey as if their estimate of distance were determined primarily by their retinal elevation of the image of the prey. In this paper we present and demonstrate an algorithm that utilizes the robot's perceived planar parallax for structure recovery of points above the ground plane.

Planar parallax naturally arises when a scene contains a planar region. It has been studied and used for heirarchical structure estimation using image intensities, rather than sparse image features, in [9]. Criminisi et. al. [3] determine the height of points above a scene plane using areas induced by the parallax vectors on the scene plane. This method requires the heights of three points in general position to be known, as well as the homography between the base plane and the image plane. Tri-focal constraints for new view synthesis based on a reference-plane based formulation using the "plane+parallax" representation is derived in [7]. Without explicit reference to the parallax, the decomposition of the homography induced by the reference plane was used to detect points above the plane for obstacle avoidance [4]. Rother and Carlsson [8] also derive structure using a "plane+parallax" formulation. Their technique processes multiple frames as a batch process, requiring a number of frames before estimates can be obtained. Our method processes frames as they arrive, lending itself more naturally to mobile robot applications. (However, clearly some combination of these methods would produce a robust system).

In this paper we present a linear algorithm based purely on the parallax vectors and sparse image measurements to calculate the direction of translation between the two cameras and the normal to the scene plane. Our algorithm uses this information to estimate the height measure of object points relative to the height of the second camera, the heights being the height above the scene plane and measured in the same( but arbitrary) direction. We also present an empirical study of the sensitivity of the estimated normal with respect to noise in the input corner data.

The paper is organized as follows: in section 2 we lay down the notation to be used in the paper. Section 3 discusses the idea of plane+parallax representation. The theory behind our method is explained in section 4. Our method to estimate the direction of translation and the scene-plane normal are presented and then used to estimate the relative height measure of points on the object. We summarize our algorithm in section

*This research supported in part by NSF IRI-9703352.

5. Experimental results are presented in section 6. We conclude the paper with a discussion of the results and future work in section 7.

## 2 Notation

The pinhole model of the camera is used. We will let $\mathbf{C_i}$ represent the center of the $i$-th camera. The coordinates of a 3-D point expressed in frame $i$ will be written as $\mathbf{X_i} = [X_i, Y_i, Z_i]^T$. The $j$-th point in the $i$-th image will be represented by the vector $\mathbf{x_{ij}} = [x_{ij}, y_{ij}, 1]^T$. The parallax vector in the $i$-th image is represented as $\mathbf{l_i}$. $\mathbf{H_{ij}}$ will be homography that takes points in $j$-th image to the $i$-th image. The normal of the scene plane expressed in frame $i$ will be written as $\mathbf{n_i}$. The unit vector in the direction of translation of camera $j$ as expressed in camera frame $i$ is written as $\mathbf{t_{ij}}$. The rotation of frame $j$ with respect to frame $i$ is represented as $\mathbf{R_{ij}}$. The number of corners on the reference plane is denoted as $n_p$, while the number of corners on the object is denoted as $n_o$.

## 3 Plane + Parallax

In this section we provide a brief description of the "plane + parallax" representation. The images of points on a plane are related by a 3 × 3 collineation or the homography i.e

$$\mathbf{H_{12}x_{2i}} = \mathbf{x_{1i}}, \text{for } i = 1, \cdots, n_o. \qquad (1)$$

This collineation can be estimated upto a scale factor from four coplanar points in general position. If this homography is used to transfer the points which do not lie on the scene plane from image 1 to image 2 residual vectors $\mathbf{r_i} = \mathbf{H_{12}x_{2i}} - \mathbf{x_{1i}}$ will arise. These vectors are the parallax vectors( Figure 1). The parallax vectors depend on the height of the point above the scene plane and the translation between the two cameras. It is independent of the rotation between the two frames. The calibration parameters of the camera and the rotation between the two camera frames is captured in the reference-plane to image-plane homography. So we can decouple the rotation and the translation.

For the reminder of the paper we assume that the detected image corners have been segmented into points on the reference plane and points on the object. This segmentation can be achieved by using the invariance of the cross ratio of five coplanar points [10] or by randomly sampling the set of corner sorrespondences to fit a model to the data and select the model with the greatest support [5].

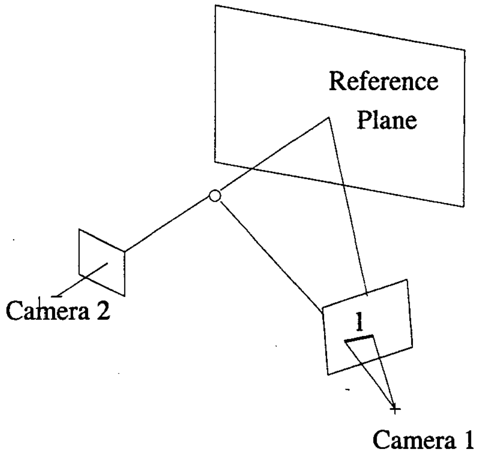


Figure 1: The geometry of "plane+parallax" representation. The open circle represents the world point that is being imaged. l represents the parallax vector

## 4 Theory

### 4.1 Estimation of Translation Direction

Image coordinates of points on the scene plane must satisfy Eq 1. For points not on the plane we can calculate the residual vectors $\mathbf{r_i}$

$$\mathbf{r_i} = \widehat{\mathbf{x_{1i}}} - \mathbf{x_{1i}}, \text{for } i = 1, \cdots, n_o \qquad (2)$$

where $\widehat{\mathbf{x_{1i}}} = \mathbf{H_{21}x_{2i}}$. We can clearly see that for each $i$, $(\widehat{\mathbf{x_{1i}}} - \mathbf{x_{1i}})$, $\mathbf{x_{1i}}$ and $\mathbf{t_{21}}$ are coplanar(ref Figure 1). It follows that

$$(\widehat{\mathbf{x_{1i}}} - \mathbf{x_{1i}}) \times \mathbf{x_{1i}} \cdot \mathbf{t_{12}} = 0, \qquad (3)$$

which on simplification yields

$$(\widehat{\mathbf{x_{1i}}} \times \mathbf{x_{1i}}) \cdot \mathbf{t_{12}} = 0. \qquad (4)$$

We define the parallax vector $\mathbf{l_i}$ for all the corners not on the scene plane as

$$\mathbf{l_i} = \widehat{\mathbf{x_{1i}}} \times \mathbf{x_{1i}}, \text{for } i = 1, \cdots, n_o \qquad (5)$$

The magnitude of the parallax vectors $\mathbf{l_i}$ form a natural weighting for the following least squares solution to find the direction of translation $\mathbf{t_{12}}$,

$$\sum\nolimits_i \mathbf{t_{12}}^T \mathbf{l_i} \mathbf{l_i}^T \mathbf{t_{12}} = 0. \qquad (6)$$

The solution to the above problem is the eigenvector of the matrix $\sum_i \mathbf{l_i}\mathbf{l_i}^T$ with the smallest eigenvalue.

In practice we include the parallax vector in the construction of the matrix $\sum_i \mathbf{l}_i\mathbf{l}_i^T$ only if its magnitude is above a certain threshold, i.e,

$$||\mathbf{l}_i||_E > \delta_1, \text{for } \delta_1 > 0, \tag{7}$$

where, $||\cdot||_E$ represents the euclidean norm of the vector. We must note that the epipoles can be estimated as the intersection of the residual vectors [1], but this method is ill-conditioned and hence our method is preferable.

### 4.2 Estimation of Normal

The homography matrix can be expressed in terms of the translation direction and the normal to the plane as,

$$\mathbf{H_{12}} = \lambda \mathbf{R_{21}}(\mathbf{I} - \frac{\mathbf{t_{12}n_2^T}}{d}) \tag{8}$$

where $\lambda$ is an arbitrary scale factor and $d$ is the scale factor of the translation relative to the normal. Using the above definition for $\mathbf{H_{21}}$ we get

$$\mathbf{t_{12}^T H_{12}^T H_{12} t_{12}} = k\left(1 - \frac{\mathbf{n_2^T t_{12}}}{d}\right)^2, \tag{9}$$

where $k$ is a constant. In the above manipulation we have made use of the fact that $\mathbf{R_{12}} \in SO(3)$.
Let $\mathbf{u}$ be a unit vector such that

$$\mathbf{u} \perp \mathbf{n_2}, \tag{10}$$

then

$$\mathbf{u^T H_{12}^T H_{12} u} = c, \tag{11}$$

where $c$ is a constant as $\mathbf{u} \cdot \mathbf{n_2} = 0$.
Let $\lambda_1$,$\lambda_2$ and $\lambda_3$ be the eigenvalues of $\mathbf{H_{12}^T H_{12}}$. Let $\mathbf{e}_1$,$\mathbf{e}_2$ and $\mathbf{e}_3$ be the corresponding eigenvectors. Additionally assume that the eigenvectors are rescaled such that $\lambda_2 = 1$ and that the eigenvalues are arranged in ascending order.
If $\alpha$ is a 3-vector, such that

$$\alpha = \alpha_1\mathbf{e_1} + \alpha_2\mathbf{e_2} + \alpha_3\mathbf{e_3}, \tag{12}$$

then

$$\alpha^T \mathbf{H_{12}^T H_{12}} \alpha = {\alpha_1}^2\lambda_1 + {\alpha_2}^2\lambda_2 + {\alpha_3}^2\lambda_3 \tag{13}$$

is the equation of an ellipsoid centered on the origin. The eigenvectors of $\mathbf{H_{12}^T H_{12}}$ are the major axes of the ellipsoid. The plane of directions of $\mathbf{u}$ will be the plane through the origin that intersects the ellipsoid in a circle. Based on the degeneracy of the eigenvalues there are three possible cases:

*Case(1) : All the three eigenvalues are the same.*
In such a case there is no solution. This corresponds to the case where there is no translation.

*Case(2) : Two of the eigenvalues are equal.*
In this case there is one solution. The normal vector $\mathbf{n}$ is the non degenerate eigenvector.

*Case(3) : The three eigenvalues are distinct.*
There are two possible planes that cut the ellipsoid in circles. Each of the two possible planes passes through $\mathbf{e_2}$, the eigenvector corresponding to the middle eigenvalue. The angle between the plane and $\mathbf{e_3}$ in the plane of $\mathbf{e_1}$ and $\mathbf{e_3}$ is $\theta$ where,

$$\cos(\theta) = \pm\sqrt{\frac{1-\lambda_1}{\lambda_3 - \lambda_1}}. \tag{14}$$

The directions of the two possible plane normals are given by,

$$\mathbf{n_2} = \mathbf{e_2} \times (\cos(\theta)\mathbf{e_3} \pm \sin(\theta)\mathbf{e_1}). \tag{15}$$

Assuming that the latter case is true we now have four possible directions for the normal $\mathbf{n}$. This can be reduced to two choices by using the visibility constraint

$$\mathbf{n_2} \cdot \mathbf{e_z} < 0, \tag{16}$$

where $\mathbf{e_z}$ is the direction vector for the principal axis of the camera. Construct a $3 \times 3$ homography matrix, $\widehat{\mathbf{H_{12}}}$, using Eq( 8) for each of the estimated normals $\mathbf{n}$ of the scene plane. The following dot product is calculated for each of the constructed homographies,

$$\jmath(\mathbf{n_2}) = \frac{\mathbf{H_{12}^T H_{12}}}{|\mathbf{H_{12}^T H_{12}}|} \cdot \frac{\widehat{\mathbf{H_{12}^T}}\widehat{\mathbf{H_{12}}}}{|\widehat{\mathbf{H_{12}^T}}\widehat{\mathbf{H_{12}}}|}. \tag{17}$$

The $\mathbf{n_2}$ corresponding to the larger dot product $\jmath(\mathbf{n_2})$ is chosen as the scene plane normal. Thus we see that we have calculated the normal to the scene plane without explicitly computating the rotation of the camera between the two positions. Also the number of choices for the normals from which we have to choose is just two, as compared to four in [4]. We also give a measure to choose between the two choices.

### 4.3 Estimation of Heights

If we represent the height of the $i$-th point on the object as $h_i$ and the height of the second camera above the scene plane as $h_{c_2}$, then by applying the sine rule to the triangles in the epipolar plane we get

$$\frac{h_i}{h_{c_2}} = \frac{\sin(\phi)\sin(\theta)}{\sin(\theta+\beta)\sin(\beta+\phi)} \tag{18}$$

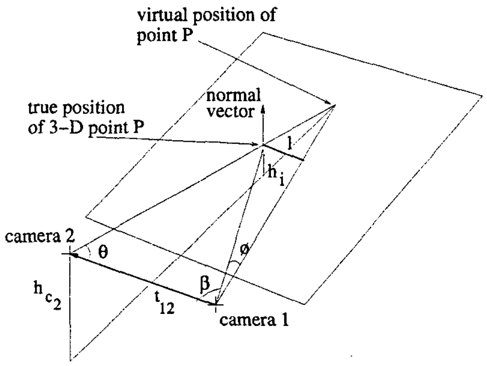


Figure 2: Scene geometry with the plane induced parallax and the angles we use illustrated.

where

$$\begin{aligned} \cos(\theta) &= \mathbf{x_{2i}} \cdot -\mathbf{t_{21}} && (19) \\ \cos(\beta) &= \mathbf{x_{1i}} \cdot \mathbf{t_{21}} && (20) \\ \cos(\phi) &= \widehat{\mathbf{x_{1i}}} \cdot \mathbf{x_{1i}} && (21) \\ \cos(\phi+\beta) &= \widehat{\mathbf{x_{1i}}} \cdot \mathbf{t_{21}} && (22) \end{aligned}$$

Note that $h_i$ and $h_{c_2}$ are measured above the plane in the same direction.

Thus we see that our algorithm calculates the translation direction and the normal to the scene plane without explicitly calculating the rotation matrix between the two frames. Also we have shown how the relative heights of points above the reference plane can be calculated using the information derived above.

# 5 Algorithm

In this section we summarize the results of previous sections into an algorithm.

*Step (1):* The corners in the two images of the stereopair are extracted. We use the Plessey corner detector [6] and we fit a quadratic to the corner strengths to get the corners to sub-pixel accuracy.

*Step (2):* The epipolar constraint, which holds true for stereo pair images, is used to match the corresponding corners in the two images [11] .

*Step (3):* From the set of corner correspondences, we segment the corners into corners lying on the scene-plane and corners lying on the object( corners not on the scene-plane). A RANSAC based approach [5], using the homography to fit the model, is employed to realize this.

*Step (4):* Using all the points that are on the scene plane we get a good estimate of the homography induced by the plane.

*Step (5):* For $i = 1 \cdots n_o$, calculate the parallax vector $\mathbf{l_i}$ according to Eq(5). If $\mathbf{l_i}$ satisfies the threshold condition, Eq(7), it is used to construct the matrix $\sum_i \mathbf{l_i}\mathbf{l_i}^T$. The translation vector $\mathbf{t_{12}}$ is found as the eigenvector corresponding to the smallest eigenvalue of the matrix $\sum_i \mathbf{l_i}\mathbf{l_i}^T$.

*Step (6):* The normal to the scene-plane $\mathbf{n_2}$ is estimated using Eqs(15),(16) and (17).

*Step (7):* The relative heights of the points on the object are estimated using (18).

# 6 Experiments

The schematic of our experimental setup is shown in Figure 3(a). The image pair shown in Figure 4(a) and (b) is used to validate our algorithm. The images were taken in a controlled environment using the pan-tilt unit shown in Figure 3(b). The camera had a focal length of approximately $6.2mm$. The distance from the camera to the scene was approximately $30cm$. Corner correspondences and their segmentation was achieved as in *Step(2)* and *Step(3)* respectively of the algorithm in section 5.

For the three points on the object, shown in Figure 4(a) and Figure 4(b), the estimated relative heights are tabulated in Table 1. The monotonicity of the points' relative heights with respect to their real heights above the plane is obvious.

| *Point* | *Pair 1* | *Pair 2* |
|---|---|---|
| 1 | 0.984 | 0.967 |
| 2 | 0.969 | 0.923 |
| 3 | 0.953 | 0.944 |

Table 1: Estimated relative heights

The normal vector to the reference plane, as estimated by our algorithm, for the first image pair, is $(0.0745, -0.8512, 0.5195)^T$. The expected normal vector is $(0.1592, -0.8387, 0.5208)^T$ The results for the normal are accurate within 4 degrees of the expected value.

We used synthetic data to study the sensitivity of our method to noise. The zero mean Gaussian noise was added to the input synthetic data. The deviation of the estimated normal, $D$, from the expected value, for various values of the standard deviation parameter, was calculated as

$$D = \cos^{-1} \frac{\mathbf{n_2}.\widehat{\mathbf{n_2}}}{||\mathbf{n_2}||.||\widehat{\mathbf{n_2}}||} \qquad (23)$$

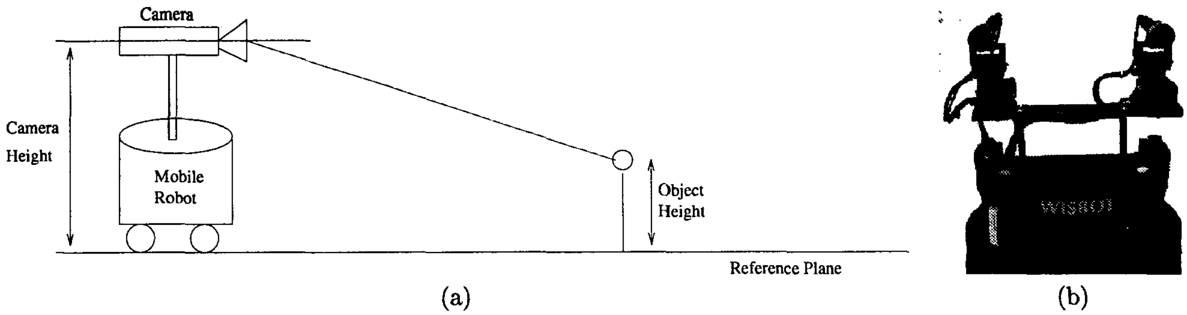


Figure 3: (a) The schematic of the experimental set-up. (b) The WISBOT pan-tilt unit used to make the images

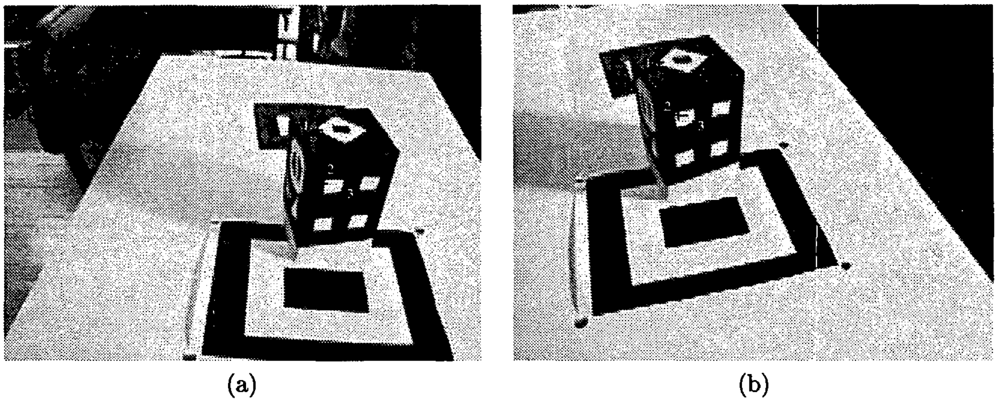


Figure 4: The first image pair used for validation of algorithm. (a) The left image. (b) The right image.

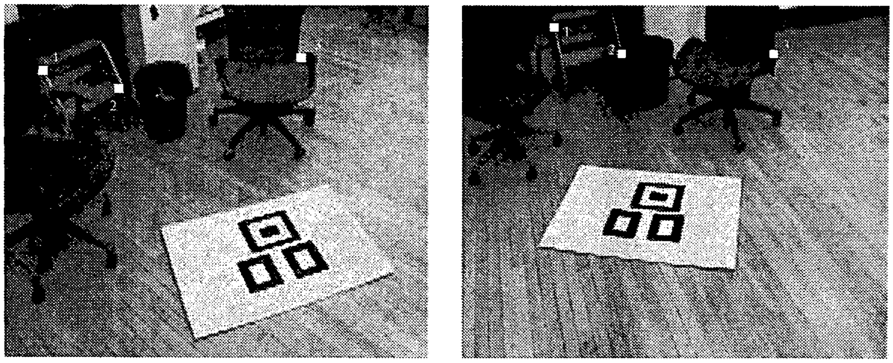
where, $\mathbf{n_2}$ is the estimated normal and $\widehat{\mathbf{n_2}}$ is the normal to the plane when the input data is noise free. For each value of the standard deviation we averaged the deviation of the normal, $D$, over ten noise datasets. This variation of accuracy of the estimate of the normal is shown in Figure 6(b).

## 7 Discussion and Future Work

We have demonstrated a simple algorithm to derive the relative heights of objects above a plane. Our algorithm processes frames as they arrive, in contrast to the batch processes which can get estimates of structure parameters only after acquiring a requisite number of frames [8]. Hence our algorithm is more suited for mobile robot applications.

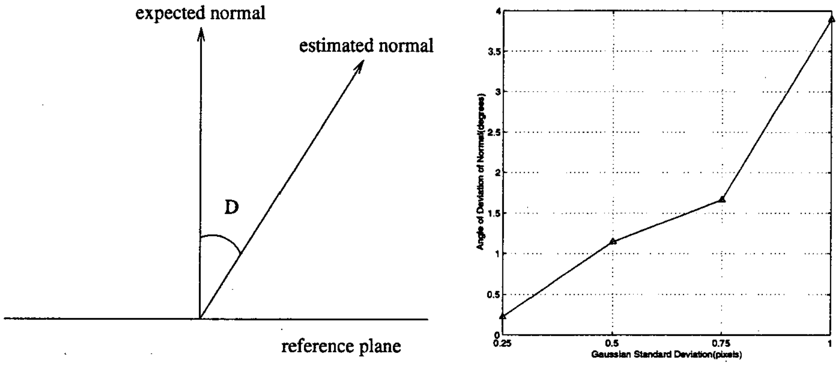
Our algorithm is convenient for a number of reasons. First, our algorithm calculates the translation direction weighted by the magnitude of the parallax vectors. This was found to be robust to noise introduced by small parallax vectors. Our algorithm also has the advantage that explicit computation of rotation matrix describing the relative orientation of the cameras is not required. Additionally, all our calculations are based on image measurements and none of the camera calibration parameters are required. Although calibration is not required, with partial scene calibration we can obtain an Euclidean reconstruction of the scene relative to a scene based coordinate system.

Our initial experiments indicate that plane+parallax will prove useful for our robot exploration applications. A number of issues must be addressed. Integrating data obtained across multiple frames in a efficient manner requires solving the viewpoint synthesis problem. The sensitivity of the algorithm and the numerical issues that arise as we get closer to critical motions of the camera are under investigation. We are currently analyzing how the parallax vectors of object points closer to the reference plane affect the algorithm. Once the investigations are complete we will have a clear idea of the algorithm's applicability and limitation in practice.



(a) (b)

Figure 5: The second image pair used for validation of algorithm. (a) The left image. (b) The right image.



(a) (b)

Figure 6: (a) Deviation of the normal, $D$, from its expected value. (b) Accuracy of the normal estimation with noisy data, using Eq(10). The value of, $D$, is averaged over ten noise datasets, for each value of the standard deviation.

# References

[1] P. Beardsley, D. Sinclair, and A. Zisserman. Ego-motion from six points. In *Insight meeting, Catholic University Leuven.* Feb. 1992.

[2] T. Collett and S.B. Udin. Frogs use retinal elevation as a cue to distance. *Journal of Comp. Physiol A*, 163:677–683, 1988.

[3] A. Criminisi, I. Reid, and A. Zisserman. Duality, rigidity and planar parallax, 1998.

[4] O. Faugeras. *Three-Dimensional Computer Vision: A Geometric Viewpoint.* MIT Press, Cambridge, Massachusetts, 1993.

[5] M. Fischler and R. Bolles. Random sampling consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. Assoc. Comput. Mach.*, 24(6):381–395, 1981.

[6] C. Harris and M. Stephens. A combined corner and edge detection. In *Proc. 4th Alvey Vision Conference*, pages 147–151, 1988.

[7] M. Irani, P. Anandan, and D. Weinshall. From reference frames to reference planes: Multi-view parallax geometry and applications. In *ECCV (2)*, pages 829–845, 1998.

[8] C. Rother and S. Carlsson. Linear multiview reconstruction and camera recovery. In *ICCV, July 2001.*

[9] H. Sawhney. 3d geometry from planar parallax. In *CVPR*, pages 929–934, 1994.

[10] D. Sinclair and A. Blake. Quantitative planar region detection. In *IJCV, 1996.*

[11] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry,. *Artificial Intelligence, December 1995*, 78:87–119, 1995.