Accurate and True Reconstruction of 3D Models with Closed-Form Solutions

Wee Kheng Leow and Zhiyong Huang

leowwk,huangzy@comp.nus.edu.sg

Technical Report No. TR12/98

11 December 1998

School of Computing
National University of Singapore
Lower Kent Ridge Road, Singapore 119260

Accurate and True Reconstruction of 3D Models with Closed-Form Solutions

Wee Kheng Leow and Zhiyong Huang
leowwk,huangzy@comp.nus.edu.sg
School of Computing, National University of Singapore
Lower Kent Ridge Road, Singapore 119260

Abstract

3D models of objects or scenes can be reconstructed from image sequences or multiple views. Reconstruction algorithms that adopt calibrated cameras generally produce more accurate results than those that adopt uncalibrated cameras, whereas the latter are more convenient to use. Depending on an application's need, one may trade off between reconstruction accuracy and convenience.

The objective of this research is to devise an accurate method for reconstructing the true and complete 3D models of small objects for use in applications such as virtual reality and model-based object recognition. This technical report focuses on the task of accurately recovering 3D coordinates of world points given a set of tracked 2D image points. To ensure reconstruction accuracy, the method uses a calibrated camera and a computer-controlled turntable. With this equipment setup, close-form solutions to the reconstruction problem are derived for various camera models including the perspective and the projective models. The existence of close-form solutions facilitate the construction of an efficient, accurate, and reliable optimization algorithm that performs true and complete reconstruction. Comprehensive experiments have been conducted to test the method's performance. Experimental results show that the method is more accurate and robust against noise than existing methods.

1 Introduction

3D models of objects or scenes can be reconstructed from image sequences or multiple views. These multiple views may be captured using several cameras, a single moving camera, or a single static camera looking at a moving object. The cameras may or may not be calibrated. Methods that adopt calibrated cameras generally produce more accurate reconstruction results than those that adopt uncalibrated cameras. On the other hand, the latter methods are more convenient and easier to use because they do not require prior calibration of the cameras. Depending on an application's need, one may trade off between reconstruction accuracy and convenience.

3D reconstruction is used in applications ranging from robot navigation, model-based object recognition, virtual reality to CAD/CAM. Different applications impose different needs on the reconstruction methods. For robot navigation, only the frontal views that the robots can see need to be reconstructed. The reconstruction should be performed as fast as possible. The accuracy of the reconstruction may not be critical as long as 3D features that are useful for navigation are recovered. On the other hand, accurate reconstruction is required for model-based object recognition, virtual reality, and especially CAD/CAM. In addition, these applications also require the reconstruction of the entire models of objects and scenes rather than just the frontal views.

The overall goal of our research is to devise an accurate method for reconstructing the complete 3D models of small objects. To ensure reconstruction accuracy, the method uses a calibrated camera and a computer-controlled turntable, similar to the setup in [13, 32, 33, 34]. Calibration of the camera's intrinsic and extrinsic parameters can be performed using standard calibration algorithms such as [10, 36] with the calibration object placed on the turntable. Specialized calibration algorithms adapted to rotating objects such as [33] can also be used. The turntable can be rotated precisely under computer control. This equipment setup permits accurate recovery of the entire 3D models of objects.[1]

This technical report focuses on the task of recovering the 3D coordinates of the world points on an object given a set of tracked 2D image points. The novelty and main contributions of this work include the following:

1. Derive closed-form solutions to the reconstruction problem under various camera models namely, perspective, para-perspective, weak perspective, affine, and projective cameras.

2. Devise an accurate, efficient, and reliable algorithm to perform true and complete reconstruction based on the closed-form solutions.

3. Perform comprehensive tests and comparisons of reconstruction results. This contribution is especially important because very few comparative studies of reconstruction methods have been published.

In the following sections, a comparative review of existing related methods will first be discussed in Section 2. Next, our reconstruction method will be described in Section 3. Extensive experiments have been performed to verify the accuracy and robustness of the method and they are described in Section 4. Fianlly, Section 5 summarizes the research work and concludes this technical report.

2 Related Works

There are several general approaches to reconstructing 3D models from images and they fall into the various shape-from-X categories. The shape-from-silhouette approach [18, 19] is particularly interesting because of its similarity to reconstruction by laser ranging, which has been successfully applied to recovering complete 3D models of objects [6]. However, shape-from-silhouette has the disadvantage that concave parts of an object may not appear on the silhouette and, therefore, cannot be recovered. In the following discussion, we will focus only on 3D reconstruction methods based on image sequence or multiple views.

View-based reconstruction methods can be categorized along the following four main dimensions: (1) the reconstruction approach, (2) the camera model, (3) calibrated vs. uncalibrated camera, and (4) the type of reconstruction achieved. The reconstruction approach can be roughly classified as either projective or Euclidean.

[1]The top and bottom views of an object can be recovered using the same equipment setup by placing the object on its side.

The projective approach typically assumes the perspective camera model and performs reconstruction in the projective space (e.g., [8, 12, 26, 27]). A transformation equation is set up to relate image points, camera parameters, and the world points to be recovered. An optimization algorithm is then applied to obtain the best fitting camera parameter values and the coordinates of the world points. Therefore, projective algorithms can recover camera parameters together with the world points. They do not require that the camera be calibrated before the reconstruction process. Another important advantage of this approach is that, in the projective space, the transformation equation can be written in a simple linear form thereby simplifying the solution methods.

Early projective algorithms typically reconstruct up to an arbitrary projective [3, 11] or affine transformation [8] of the target world points. By applying certain geometrical constraints or invariants, recent algorithms can recover up to a Euclidean (or similarity) transformation[2] [12, 20, 26, 27] of the true reconstruction. In particular, [27] allows variation of focal length while the others assume fixed focal length.

It has been argued in [23] that projective algorithms can be unreliable because they have too many free parameters to play with. They are free to overfit the data and, as a result, recover inaccurate camera parameters together with inaccurate world points. There seems to be some visual evidence to this claim. For example, in [26, 27], some straight edges in the actual scene appear wavy in the reconstructed models.

Another weakness of the projective approach is that the image points to be fitted to the transformation equation must be available in most of the image frames. Otherwise, there will be additional unknowns in the equation, which will complicate the optimization algorithm. Therefore, the projective approach is more appropriate for recovering the frontal views of objects or scenes from short image sequences. Complete reconstruction of entire objects would be difficult because at any one time, half the image points are hidden at the back of the objects.

The Euclidean approach performs the reconstruction directly in the Euclidean space. Most Euclidean methods recover at least a scaled version of the true reconstruction. Some methods can perform self-calibration and do not require calibrated cameras [7, 13, 25, 29, 31, 35]. Like the projective approach, these methods generally require that the image points be present in most of the image frames. Therefore, they are appropriate for reconstructing frontal views instead of the complete object models.

The main difficulty of the Euclidean approach is that perspective projection is a non-linear transformation in the Euclidean space. As a result, many Euclidean algorithms assume approximations of the perspective camera in order to simplify the transformation equations. For example, [7, 35, 37] assume orthographic projection, [5] uses weak perspective camera, [25] adopts para-perspective camera, and [29, 31] assume affine projection. By adopting these approximations, the transformation equations become linear equations which are easier to manipulate. The disadvantage of such an approach is that the approximations are appropriate only when the depth variations of the world points are small compared to the average depth. In recovering detailed 3D structures of small objects, these algorithms can become quite inaccurate (which will be shown in Section 4).

Of the methods that assume the perspective camera, not all of them adopt the full camera model. For example, [16, 21, 22, 24, 38] ignore focal length, aspect ratio, skew, and principal points; [33, 34] ignore aspect ratio, skew, and principal points; and [4, 15] ignore aspect ratio and skew. In general, the focal length can be assumed to be 1 without causing major reconstruction

[2] A Euclidean transformation is the composition of a rotation, a translation, and a uniform scaling.

error, except that the reconstruction is performed at best up to a scale factor. Other parameters such as the aspect ratio, however, may be too significant to be ignored (Section 4).

Of the methods that adopt the full perspective camera, [30] uses three calibrated cameras to recover the true 3D coordinates of objects with known or unknown motion, and [13] uses a single camera to perform quasi-Euclidean reconstruction under unknown but complete orbital camera motion. Although the method in [30] can potentially recover complete 3D models, only the reconstructed frontal views are shown in the article.

The methods by Matsumoto et al. [18], Mehren and Rodehorst [14, 19], and Szeliski and co-workers [13, 32, 33, 34] are the most similar to ours. All these methods attempt to recover Euclidean structures from rotating objects. In particular, the methods of Matsumoto et al. and Mehren and Rodehorst reconstruct object models from image silhouette. Mehren and Rodehorst also introduced another method that recover Euclidean structure from image point correspondence. The method assumes an affine transformation instead of a perspective one.

Szeliski's algorithm [32] recovers Euclidean structure from tracked image feature points. In the method, the object coordinate frame is attached to the object and rotates with it. As the object is rotated, feature points on the object are projected to the image points in a fixed image plane through a series of rotation and translation matrices. During reconstruction, image points are backprojected through a series of inverse transformations to "clouds" of object points in the 3D space. Neighboring object points are then merged using an uncertainty-weighted distance measure.

The methods of [33, 34] extend that of [32] by allowing the object to be displaced from the axis of rotation. The algorithm adopts an alternative and simplified version of the perspective transformation. In addition, it uses the Levenberg-Marquardt algorithm to recover both the 3D coordinates as well as the rotation and translation matrices. The method in [13] further extends those in [33, 34] by adopting an alternative formulation of the full perspective transformation and an uncalibrated camera that undergoes unknown but complete orbital motion around a static object (or equivalently, unknown but complete object rotation). This method recovers only a quasi-Euclidean structure whereas previous algorithms recover Euclidean structures.

In contrast to the methods of Szeliski and co-workers, in our method, the world coordinate frame is fixed onto the turntable and does not rotate with it.[3] A 3D world point on an object projects to a 2D image point in the image plane. As the world point is rotated, it casts new projection lines from its new 3D locations to new 2D points. By rotating the projection lines backward, the new world points would coincide with the original world point. Once this coincidence is achieved, the world point can be computed as the intersection of the backward rotated projection lines. Our algorithm is thus known as the *Ray Intersecting* (RI) method.

In comparison with existing works, the RI method has the following advantages:

1. It adopts a full perspective camera model and could, therefore, achieve very accurate reconstruction.

2. It recovers the true 3D coordinates of the world points instead of a Euclidean transformation of them.

3. It does not differentiate between pure rotation and orbital motion of objects. It will recover the true world coordinates regardless of whether the object is placed on the

[3]We distinguish between the object coordinate frame and the world coordinate frame. The object frame is attached to an object and moves with it. The world frame, on the other hand, is fixed and stationary, and an object may move within it.

rotation axis or off-centered.

4. It does not require that the tracked image points be available in most of the image frames. It only needs to know the image points' rotation angles corresponding to each image frame. As a result, the RI method can be used to reconstruct complete object models.

5. Its optimization algorithm is simple, efficient, and reliable since it is based on the closed-form solution under full perspective camera.

Existing works that also provide closed-form solutions include [7, 16, 39, 40]. [7] provides a closed-form solution under othorgraphic projection and fixed rotation and translation. [16, 39] provide closed-form solutions under simplified perspective projections for scaled Euclidean reconstruction. [40] provides a closed-form solution for scaled Euclidean reconstruction based on four coplanar correspondences of two externally uncalibrated cameras. It is noted that some existing methods possess some of the above advantages but none possesses all of them.

3 Ray Intersecting Reconstruction Algorithm

For the sake of completeness, this section first provides an overview of the perspective coordinate systems. Next, closed-form solution under full perspective camera will be discussed, followed by a description of the efficient optimization algorithm for obtaining accurate reconstruction in practice. Finally, close-form solutions under projective, affine, para-perspective, and weak perspective cameras are also derived.

3.1 Perspective Coordinate Systems

Under perspective projection, a 3D point W with world coordinate $\mathbf{X} = (X, Y, Z)^\top$ is first mapped onto the camera coordinate system at $\mathbf{X}^c = (X^c, Y^c, Z^c)^\top$ by a rigid transformation (Fig. 1)

$$\mathbf{X}^c = \mathbf{R}\mathbf{X} + \mathbf{D} \tag{1}$$

where $\mathbf{R} = (\mathbf{R}_1^\top, \mathbf{R}_2^\top, \mathbf{R}_3^\top)^\top$ is the 3×3 rotation matrix and $\mathbf{D} = (D_X, D_Y, D_Z)^\top$ is the displacement vector that relate the camera and the world coordinate frames. The rotation matrix can be specified in terms of three Euler angles: pitch (vertical angle) ω, yaw (horizontal angle) ϕ, and roll κ (Fig. 2). Pitch ω is the rotation angle about the X-axis from the positive Z-axis to the positive Y-axis. Yaw ϕ is the rotation angle about the Y-axis from the positive X-axis to the positive Z-axis. Roll κ is the rotation angle about the Z-axis from the positive Y-axis to the positive X-axis. With Euler angles, the rotation matrix $\mathbf{R}$ can be expressed as

$$\begin{aligned}\mathbf{R} &= \begin{bmatrix} \cos\kappa & \sin\kappa & 0 \\ -\sin\kappa & \cos\kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & 0 & -\sin\phi \\ 0 & 1 & 0 \\ \sin\phi & 0 & \cos\phi \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\omega & \sin\omega \\ 0 & -\sin\omega & \cos\omega \end{bmatrix} \\ &= \begin{bmatrix} \cos\phi\cos\kappa & \sin\omega\sin\phi\cos\kappa + \cos\omega\sin\kappa & -\cos\omega\sin\phi\cos\kappa + \sin\omega\sin\kappa \\ -\cos\phi\sin\kappa & -\sin\omega\sin\phi\sin\kappa + \cos\omega\cos\kappa & \cos\omega\sin\phi\sin\kappa + \sin\omega\cos\kappa \\ \sin\phi & -\sin\omega\cos\phi & \cos\omega\cos\phi \end{bmatrix}\end{aligned} \tag{2}$$

which is an othornormal matrix, i.e.,

$$\mathbf{R}^\top\mathbf{R} = \mathbf{I} \tag{3}$$

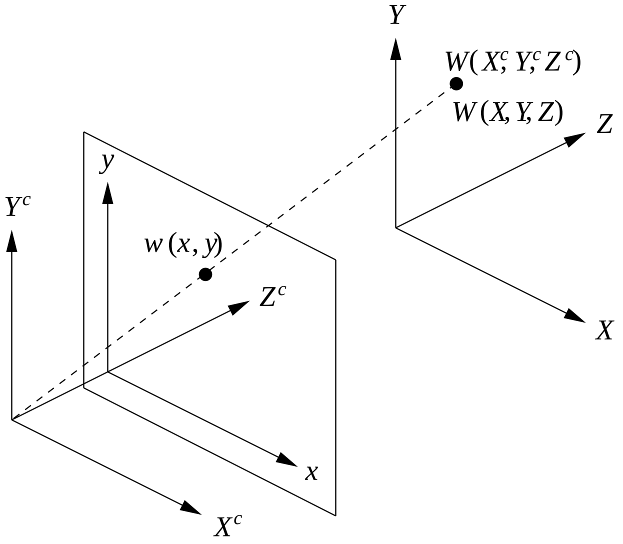


Figure 1: Left-hand coordinate systems for perspective projection. The X-Y-Z system denotes the world coordinate frame, the X^c-Y^c-Z^c system the camera coordinate frame, and the x-y system the image coordinate frame.

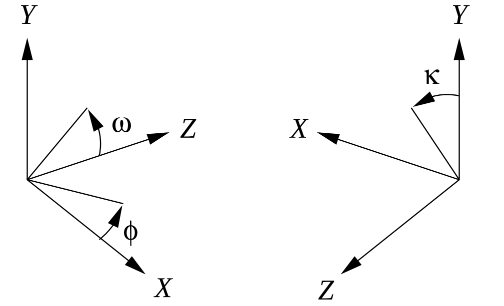


Figure 2: The three Eular angles: (a) pitch ω, (b) yaw ϕ, and (c) roll κ.

where $\mathbf{I}$ is the identity matrix.

The world point in the camera frame is projected to a 2D image point w at coordinate $\mathbf{x} = (x, y)^\top$ in the image plane. The projection is most conveniently expressed in the form of homogeneous coordinates. Define homogeneous coordinates $\mathbf{X}_h = (X_1, X_2, X_3, X_4)^\top$ and $\mathbf{x}_h = (x_1, x_2, x_3)^\top$ such that $(X, Y, Z) = (X_1/X_4, X_2/X_4, X_3/X_4)$ and $(x, y) = (x_1/x_3, x_2/x_3)$. The general mapping between $\mathbf{X}$ and $\mathbf{x}$ can be written in terms of a transformation matrix $\mathbf{T} = [T_{ij}]$:

$$\mathbf{x}_h = \mathbf{T}\mathbf{X}_h \tag{4}$$

and the transformation matrix $\mathbf{T}$ can be decomposed as follows [9, 17]:

$$\mathbf{T} = \mathbf{CPG} \; . \tag{5}$$

The 3×3 matrix $\mathbf{C}$ accounts for intrinsic camera parameters:

$$\mathbf{C} = \begin{bmatrix} f\xi & s & o_x \\ 0 & f & o_y \\ 0 & 0 & 1 \end{bmatrix} \tag{6}$$

where f is the camera focal length, ξ the aspect ratio, s the skew, and (o_x, o_y) the principal point (where the optic axis intersects the image plane). The 3×4 matrix $\mathbf{P}$ performs the projection operation, and the 4×4 matrix $\mathbf{G}$ accounts for extrinsic camera parameters, encoding the relative position and orientation between the world and the camera frames.

Under perspective projection, the projection matrix $\mathbf{P}$ and the extrinsic parameter matrix $\mathbf{G}$ are given by

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{7}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{R} & \mathbf{D} \\ \mathbf{0}^\top & 1 \end{bmatrix} . \tag{8}$$

Therefore, the transformation matrix $\mathbf{T}$ under perspective camera is

$$\mathbf{T} = \begin{bmatrix} f\xi\mathbf{R}_1^\top + s\mathbf{R}_2^\top + o_x\mathbf{R}_3^\top & f\xi D_X + sD_Y + o_x D_Z \\ f\mathbf{R}_2^\top + o_y\mathbf{R}_3^\top & fD_Y + o_y D_Z \\ \mathbf{R}_3^\top & D_Z \end{bmatrix} \tag{9}$$

and the image and world coordinates are related by

$$\begin{aligned} x &= \frac{f\xi(\mathbf{R}_1 \cdot \mathbf{X} + D_X) + s\,(\mathbf{R}_2 \cdot \mathbf{X} + D_Y)}{\mathbf{R}_3 \cdot \mathbf{X} + D_z} + o_x \\ y &= \frac{f(\mathbf{R}_2 \cdot \mathbf{X} + D_y)}{\mathbf{R}_3 \cdot \mathbf{X} + D_z} + o_y \end{aligned} \tag{10}$$

which specify a non-linear transformation between the world coordinate $\mathbf{X}$ and the image coordinate $\mathbf{x}$.

3.2 Solution Under Perspective Camera

By rearranging terms, Eq. 10 can be written as two equations of 3-D planes:

$$\begin{aligned} A_1X + B_1Y + C_1Z + D_1 = 0 \\ A_2X + B_2Y + C_2Z + D_2 = 0 \end{aligned} \tag{11}$$

where

$$\begin{aligned} A_1 &= f\xi R_{11} + sR_{21} + (o_x - x)R_{31} \\ B_1 &= f\xi R_{12} + sR_{22} + (o_x - x)R_{32} \\ C_1 &= f\xi R_{13} + sR_{23} + (o_x - x)R_{33} \\ D_1 &= f\xi D_X + sD_Y + (o_x - x)D_Z \\ A_2 &= fR_{21} + (o_y - y)R_{31} \\ B_2 &= fR_{22} + (o_y - y)R_{32} \\ C_2 &= fR_{23} + (o_y - y)R_{33} \\ D_2 &= fD_Y + (o_y - y)D_Z\ . \end{aligned} \tag{12}$$

These two planes intersect at a 3-D line given by the following parametric equation:

$$(X, Y, Z) = (X_0, Y_0, 0) + t(u, v, 1) \tag{13}$$

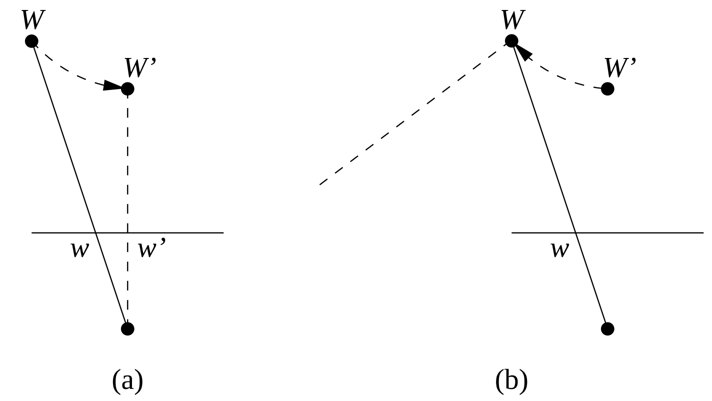


Figure 3: Ray Intersecting method. (a) World point W casts a projection line (solid line) onto the image point w on the image plane. As W rotates to its new position W', it casts a new projection line (dashed line) onto the image plane. (b) By rotating the new projection line backward, it now intersects the original line at the world point W.

where

$$\begin{aligned} X_0 &= (B_1D_2 - D_1B_2)/(A_1B_2 - B_1A_2) \\ Y_0 &= (D_1A_2 - A_1D_2)/(A_1B_2 - B_1A_2) \\ u &= (B_1C_2 - C_1B_2)/(A_1B_2 - B_1A_2) \\ v &= (C_1A_2 - A_1C_2)/(A_1B_2 - B_1A_2)\ . \end{aligned} \tag{14}$$

The formulation of this parametric equation requires that Z is not a constant. This requirement is valid since Eq. 10 describes a line of projection from a 3D world point to a 2D image point, which is not perpendicular to the Z-axis. Equation 13 can be written in the vector form

$$\mathbf{X} = \mathbf{X}_0 + t\mathbf{u} \tag{15}$$

where $\mathbf{X}_0 = (X_0, Y_0, 0)^\top$ and $\mathbf{u} = (u, v, 1)^\top$.

The world coordinate system is fixed onto the turntable such that the turntable rotates about the Y-axis. As the world point W is rotated about the Y-axis, it projects new lines

$$\mathbf{X} = \mathbf{X}_{0i} + t_i\mathbf{u}_i \tag{16}$$

onto the image plane at various rotation angles θ_i, where $\mathbf{X}_{0i} = (X_{0i}, Y_{0i}, 0)^\top$ and $\mathbf{u}_i = (u_i, v_i, 1)^\top$. By rotating the new projection lines backward about the Y-axis by the corresponding angles, the new world points would now coincide with the original world point W (Fig. 3). The backward rotated projection lines are given by

$$\mathbf{X} = \mathbf{M}_i(\mathbf{X}_{0i} + t_i\mathbf{u}_i) \tag{17}$$

where

$$\mathbf{M}_i = \begin{bmatrix} \cos\theta_i & 0 & \sin\theta_i \\ 0 & 1 & 0 \\ -\sin\theta_i & 0 & \cos\theta_i \end{bmatrix} . \tag{18}$$

A closed-form solution of the 3D world point can now be obtained by solving for the intersection of any two backward rotated projection lines, for example, line 0 (Eq. 15) and line i (Eq. 17).

For notational convenience, denote $\mathbf{X}'_0 = \mathbf{M}_i\mathbf{X}_{0i}$ and $\mathbf{u}' = \mathbf{M}_i\mathbf{u}_i$. Then, the 3D world point is given by Eq. 15 at the following t:

$$t = \frac{(X'_0 - X_0)v' + (Y_0 - Y'_0)u'}{uv' - u'v} \; . \tag{19}$$

In practice, due to finite resolution and the presence of noise, the lines may not intersect at exactly one point. A good approximation of the intersection can be obtained by finding the point $\mathbf{Q}$ nearest to the lines. That is, find the $\mathbf{Q}$ that minimizes the total distance to the lines

$$D_T = \sum_i \left\| (\mathbf{Q} - \mathbf{X}_{0i}) - \frac{(\mathbf{Q} - \mathbf{X}_{0i}) \cdot \mathbf{u}_i}{\|\mathbf{u}_i\|^2} \mathbf{u}_i \right\| \tag{20}$$

The minimization can be performed efficiently and reliably given a good initial guess of the actual solution. This initial guess can be obtained from the intersection of any two projection lines using the closed-form solution (Eq. 19). In the current implementation, the minimization of Eq. 20 is performed using the simple and efficient Powell's agorithm [1, 28].

3.3 Solution Under Para-Perspective Camera

The para-perspective camera [2, 31] is an approximation of the perspective model. World points are first projected parallel onto an arbitrary fixed depth plane that is parallel to the image plane, and then perspectively onto to the image plane. The angles of projection onto the fixed depth plane are given by β_X in the X-Z plane and β_Y in the Y-Z plane, which are defined by the ray connecting the focal point and the object's centroid $\bar{\mathbf{X}}^c = (\bar{X}^c, \bar{Y}^c, \bar{Z}^c)^\top$ in the camera frame:

$$\begin{aligned}
\bar{\mathbf{X}}^c &= \frac{1}{N}\sum_{i=1}^{N}(\mathbf{R}\mathbf{X}_i + \mathbf{D}) &(21)\\
\beta_X &= \cot^{-1}(\bar{X}^c/\bar{Z}^c) &(22)\\
\beta_Y &= \cot^{-1}(\bar{Y}^c/\bar{Z}^c) &(23)
\end{aligned}$$

where N is the number of world points on the object.

The para-perspective camera differs from the perspective model in the projection matrix $\mathbf{P}$ and the extrinsic camera parameter matrix $\mathbf{G}$:

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & -C_X & C_X \\ 0 & 1 & -C_Y & C_Y \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{24}$$

$$\mathbf{G} = \begin{bmatrix} \mathbf{R} & \mathbf{D} \\ \mathbf{0}^\top & \bar{Z}^c \end{bmatrix} \tag{25}$$

where C_X and C_Y are defined in terms of the projection angles β_X and β_Y:

$$\begin{array}{l} C_X = \cot\beta_X = \bar{X}^c/\bar{Z}^c \\ C_Y = \cot\beta_Y = \bar{Y}^c/\bar{Z}^c \; . \end{array} \tag{26}$$

With this camera model, the transformation matrix $\mathbf{T}$ becomes:

$$\mathbf{T} = \begin{bmatrix} f\xi(\mathbf{R}_1^\top - C_X\mathbf{R}_3^\top) + s(\mathbf{R}_2^\top - C_Y\mathbf{R}_3^\top) & T_{14} \\ f(\mathbf{R}_2^\top - C_Y\mathbf{R}_3^\top) & T_{24} \\ \mathbf{0}^\top & \bar{Z}^c \end{bmatrix} \tag{27}$$

where

$$\begin{aligned} T_{14} &= f\xi(D_X - C_X D_Z) + s(D_Y - C_Y D_Z) + (f\xi C_X + sC_Y + o_x)\bar{Z}^c \\ T_{24} &= f(D_Y - C_Y D_Z) + (fC_Y + o_y)\bar{Z}^c \ . \end{aligned} \tag{28}$$

The image and world coordinates are thus related by

$$\begin{aligned} x &= \frac{1}{\bar{Z}^c}\{f\xi[(\mathbf{R}_1 - C_X\mathbf{R}_3)\cdot\mathbf{X} + D_X - C_X D_Z] + \\ &\qquad s[(\mathbf{R}_2 - C_Y\mathbf{R}_3)\cdot\mathbf{X} + D_Y - C_Y D_Z]\} + f\xi C_X + sC_Y + o_x \\ y &= \frac{1}{\bar{Z}^c}\{f[(\mathbf{R}_2 - C_Y\mathbf{R}_3)\cdot\mathbf{X} + D_Y - C_Y D_Z]\} + fC_Y + o_y \end{aligned} \tag{29}$$

which is a linear transformation between $\mathbf{X}$ and $\mathbf{x}$. Equation 29 can be written in the form of Eq. 11 with

$$\begin{aligned} A_1 &= f\xi(R_{11} - C_X R_{31}) + s(R_{21} - C_Y R_{31}) \\ B_1 &= f\xi(R_{12} - C_X R_{32}) + s(R_{22} - C_Y R_{32}) \\ C_1 &= f\xi(R_{13} - C_X R_{33}) + s(R_{23} - C_Y R_{33}) \\ D_1 &= f\xi(D_X - C_X D_Z) + s(D_Y - C_Y D_Z) + (f\xi C_X + sC_Y + o_x - x)\bar{Z}^c \\ A_2 &= f(R_{21} - C_Y R_{31}) \\ B_2 &= f(R_{22} - C_Y R_{32}) \\ C_2 &= f(R_{23} - C_Y R_{33}) \\ D_2 &= f(D_Y - C_Y D_Z) + (fC_Y + o_y - y)\bar{Z}^c \ . \end{aligned} \tag{30}$$

Equation 30 indicates that $\bar{\mathbf{X}}^c$ must be known in order to recover world points in the manner described in Section 3.2. In practice, it may be difficult to obtain $\bar{\mathbf{X}}^c$ accurately. One technique of overcoming this problem is to place the origin of the world frame at the object's centroid. In this case, $\bar{\mathbf{X}}^c = (D_X, D_Y, D_Z)^\top$, which can be obtained from the calibration of the camera parameters. In the simulation experiments (Section 4), for the purpose of performance comparison, $\bar{\mathbf{X}}^c$ will be computed directly from the actual world coordinates.

3.4 Solution Under Weak Perspective Camera

The weak perspective or scaled orthographic camera is an approximation of the para-perspective camera with projection angles β_X and β_Y equal to 90° [31]. In other words, world points are projected perpendicularly to the fixed depth plane. The transformation matrix $\mathbf{T}$ is thus given by

$$\mathbf{T} = \begin{bmatrix} f\xi\mathbf{R}_1^\top + s\mathbf{R}_2^\top & f\xi D_X + sD_Y + o_x\bar{Z}^c \\ f\mathbf{R}_2^\top & fD_Y + o_y\bar{Z}^c \\ \mathbf{0}^\top & \bar{Z}^c \end{bmatrix} \tag{31}$$

and the image and world coordinates are related by

$$\begin{aligned} x &= \frac{1}{\bar{Z}^c}[f\xi(\mathbf{R}_1\cdot\mathbf{X} + D_X) + s(\mathbf{R}_2\cdot\mathbf{X} + D_Y)] + o_x \\ y &= \frac{1}{\bar{Z}^c}[f(\mathbf{R}_2\cdot\mathbf{X} + D_Y)] + o_y \ . \end{aligned} \tag{32}$$

Comparing Eq. 32 with Eq. 10, we can see that the weak perspective camera approximates the perspective camera by replacing each $Z_i^c = \mathbf{R}_3 \cdot \mathbf{X}_i + D_Z$ by the average $\bar{Z}^c$. Equation 32 can be written in the form of Eq. 11 with

$$
\begin{aligned}
A_1 &= f\xi R_{11} + sR_{21} \\
B_1 &= f\xi R_{12} + sR_{22} \\
C_1 &= f\xi R_{13} + sR_{23} \\
D_1 &= f\xi D_X + sD_Y + (o_x - x)\bar{Z}^c \\
A_2 &= fR_{21} \\
B_2 &= fR_{22} \\
C_2 &= fR_{23} \\
D_2 &= fD_Y + (o_y - y)\bar{Z}^c \ .
\end{aligned}
\tag{33}
$$

This set of equations is the same as Eq. 30 with $C_X = C_Y = 0$. Similar to the case of para-perspective camera, an alignment technique can be used to avoid computing $\bar{Z}^c$: place the world frame's origin at the object's centroid and align the camera such that the object's centroid projects to the camera frame's origin. In this case, $\bar{\mathbf{X}}^c = (0, 0, D_Z)^\top$. For the purpose of performance comparison in Section 4, $\bar{Z}^c$ will be computed directly from the world coordinates.

3.5 Solution Under Affine Camera

The affine camera is a special case of the general projective camera. It has no constraint on the matrix elements of $\mathbf{T}$ except that $T_{31} = T_{32} = T_{33} = 0$:

$$
\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & T_{34} \end{bmatrix} .
\tag{34}
$$

The homogeneous coordinates of the image and world points are thus related by

$$
\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ 0 & 0 & 0 & T_{34} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}
\tag{35}
$$

and the image and world coordinates are related by

$$
\begin{aligned}
x &= \frac{1}{T_{34}}(\mathbf{T}_1 \cdot \mathbf{X} + T_{14}) \\
y &= \frac{1}{T_{34}}(\mathbf{T}_2 \cdot \mathbf{X} + T_{24})
\end{aligned}
\tag{36}
$$

where $\mathbf{T}_i = (T_{i1}, T_{i2}, T_{i3})^\top, i = 1, 2, 3$. Without loss of generality, T_{34} can be take as 1 and Eq. 36 can be written in the form of Eq. 11 with

$$
\begin{aligned}
A_1 &= T_{11} \\
B_1 &= T_{12} \\
C_1 &= T_{13} \\
D_1 &= T_{14} - x \\
A_2 &= T_{21} \\
B_2 &= T_{22} \\
C_2 &= T_{23} \\
D_2 &= T_{24} - y \ .
\end{aligned}
\tag{37}
$$

The world points can now be computed in the manner described in Section 3.2. In practice, the affine camera can be calibrated by applying least sqaures method to obtain the parameter values T_{ij} (see Section 4 for details).

3.6 Solution Under Projective Camera

The projective camera generalizes all the above-mentioned camera models. Its transformation matrix is given by

$$\mathbf{T} = \begin{bmatrix} T_{11} & T_{12} & T_{13} & T_{14} \\ T_{21} & T_{22} & T_{23} & T_{24} \\ T_{31} & T_{32} & T_{33} & T_{34} \end{bmatrix} . \tag{38}$$

The image and world coordinates are related by

$$\begin{aligned} x &= \frac{\mathbf{T}_1 \cdot \mathbf{X} + T_{14}}{\mathbf{T}_3 \cdot \mathbf{X} + T_{34}} \\ y &= \frac{\mathbf{T}_2 \cdot \mathbf{X} + T_{24}}{\mathbf{T}_3 \cdot \mathbf{X} + T_{34}} \end{aligned} \tag{39}$$

and the world points can be computed in the manner described in Section 3.2, with

$$\begin{aligned} A_1 &= T_{11} - xT_{31} \\ B_1 &= T_{12} - xT_{32} \\ C_1 &= T_{13} - xT_{33} \\ D_1 &= T_{14} - xT_{34} \\ A_2 &= T_{21} - yT_{31} \\ B_2 &= T_{22} - yT_{32} \\ C_2 &= T_{23} - yT_{33} \\ D_2 &= T_{24} - yT_{34} \ . \end{aligned} \tag{40}$$

Notice that Eq. 39 is a set of non-linear equations. Currently, there is no calibration algorithm for solving the T_{ij} values for true reconstruction.

4 Performance Experiments

Comprehensive experiments have been conducted to test the Ray Intersecting (RI) method's performance under various conditions and camera models. The projective camera is omitted in this study because of the lack of a calibration algorithm. The experimental results represent the (near) optimal solutions that can be achieved by any algorithm that assume the corresponding camera model. The results, therefore, provide a quantitative measure of the accuracies of various existing algorithms that adopt similar camera models. RI's performance is also directly compared with those of several existing reconstruction algorithms.

To have precise control over the experimental parameters, 2D image points were generated synthetically through a perspective camera. In practice, the perspective camera is a good approximation of the true projective camera because parameters that are not modelled by the perspective camera typically have negligible values.

To make the experiments more realistic, image coordinates were subjected to discretization errors due to the finite resolution of the camera's digitizer. These discretization errors were

Table 1: Parameter values of perspective camera used to generate synthetic image points.

focal length	f	0.8 cm
aspect ratio	ξ	1.2
skew	s	0.01 cm
principal point	(o_x, o_y)	$(0.01, 0.01)$ cm
resolution	r	1000 pixels/cm

imposed by converting real-valued image coordinates to integer-valued pixel positions and then back to real-valued image coordinates. In some experiments, additional pixel localization noise was added to test RI's performance in the presence of noise.

Eight sets of experiments were conducted to study RI's performance under various conditions:

1. different camera models,
2. inaccurate camera model,
3. presence of noise in pixel localization,
4. inaccurate object rotation,
5. occlusion and off-centered rotation,
6. object's depth,
7. camera focal length, and
8. comparisons with existing methods.

4.1 Experiment 1: Different Camera Models

This set of experiments compare the errors in reconstruction under perspective, affine, para-perspective, and weak perspective camera models. Synthetic image points were generated using the perspective camera with parameter values given in Table 1. The values of these intrinsic paramters were based on those of real calibrated cameras.

Twenty world points were randomly selected from a sphere of radius 20cm for testing. The centroid of the sphere was located at $(0, 20, 0)$. These world points were rotated about the Y-axis through an angle of 90° in steps of 10°. The projections of these world points on the image plane were computed for testing the RI algorithm.

Three experiments were performed with the camera located at different positions (Table 2, Fig. 4). In Experiment A, the camera frame was parallel to the world frame and its origin was aligned with the centroid of the sphere. In Experiment B, the camera frame was parallel to the world frame but its origin was *not* aligned with the sphere's centroid. In experiment C, the camera frame was *not* parallel to the world frame but its origin was aligned with the sphere's centroid. In all three experiments, the displacement matrix $\mathbf{D}$ was calculated such that the distance between the camera and the sphere's centroid remained fixed at 100cm. At

Table 2: Orientations and positions of the camera in Experiments 1A, 1B, and 1C. (O_X, O_Y, O_Z) is the origin of the camera frame in the world coordinate frame.

	rotation angles			camera's position		
Experiment	ω	ϕ	κ	O_X	O_Y	O_Z
A	$0°$	$0°$	$0°$	0	20	-100
B	$0°$	$0°$	$0°$	0	10	-100
C	$45°$	$0°$	$0°$	0	90.71	-70.71

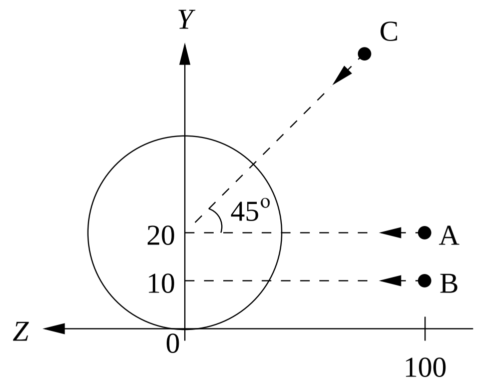


Figure 4: Orientations and positions of the camera in Experiments 1A, 1B, and 1C. This figure shows the view in the Y-Z plane.

this distance, the image formed in the image plane had a width of about 0.5cm or 500 pixels. The displacement matrix **D** can be calculated by substituting the coordinates of the sphere's centroid in the world and camera frames into Eq. 1.

The parameters of the affine camera were calibrated using a least squares method to fit seven calibration points to Eq. 36. Six of these points were located on the surface of the sphere and the remaining point was the sphere's centroid.

Test results are given in the Table 3. Column 2 gives the mean error averaged over the 20 recovered world points, where error is computed as the Euclidean distance between the actual and the recovered coordinates. Columns 3 gives the standard deviations of the errors. These test results show that reconstruction under perspective camera is at least 30 times more accurate than those based on the other camera models. Moreover, the standard deviations of errors under perspective camera are much smaller than those under the other cameras, demonstrating the robustness of RI under perspective camera.

The errors of reconstruction under affine camera are slightly larger than those under para-perspective camera in Experiments A and B because they are sensitive to the least squares calibration of affine parameters. If calibration were performed using the same world points as those presented for reconstruction, then the affine model would achieve the same accuracy as the para-perspective model.

In Experiments A and C, the camera frame's origin is aligned with the sphere's centroid. In these cases, the projection angles β_X and β_Y are equal to $90°$, and para-perspective camera reduces to weak perspective camera. Therefore, performance of weak perspective camera is as good as that of para-perspective camera. On the other hand, when the camera frame's origin

Table 3: Errors of reconstruction under various camera models and camera positions and orientations.

	Experiment A		Experiment B		Experiment C	
	error		error		error	
camera model	mean	std. dev.	mean	std. dev.	mean	std. dev.
perspective	0.038	0.021	0.032	0.013	0.032	0.018
affine	1.348	0.913	1.333	0.909	1.356	0.912
para-perspective	1.325	0.930	1.318	0.939	1.374	0.909
weak perspective	1.324	0.929	1.522	1.087	1.375	0.912

Table 4: Reconstruction errors due to inaccurate values of (a) aspect ratio, (b) skew, and (c) principal point. The fourth column gives the base case results (i.e., with correct parameter values).

(a)

aspect ratio	1.30	1.25	1.20	1.15	1.10	1.05	1.00
error	0.967	0.502	0.038	0.556	1.153	1.809	2.531

(b)

skew	0.020	0.015	0.010	0.005	0.000
error	0.088	0.078	0.038	0.049	0.078

(c)

principal point	0.020	0.015	0.010	0.005	0.000
error	1.681	0.848	0.038	0.835	1.665

is not aligned with the sphere's centroid, as in Experiment B, the projection angles β_X and β_Y are not equal to 90°, and weak perspective camera is not equivalent to para-perspective camera. Therefore, reconstruction under para-perspective camera is more accurate than that under weak perspective camera.

4.2 Experiment 2: Inaccurate Camera Model

Existing reconstruction methods often ignore certain intrinsic camera parameters in order to simplify the transformation equation (Section 2). This set of experiments examines the effects that inaccurate parameter values have on reconstruction errors. The aspect ratio, skew, and principal point are considered. Reconstruction tests were performed with various parameter values ranging from the correct ones to the default values, i.e., the values that correspond to the parameters being ignored. Other camera parameters and the image points were the same as those used in Experiment 1A.

Test results in Table 4 show that certain parameters, such as the aspect ratio and the principal point, have large effects on reconstruction accuracy. Ignoring these parameters would result in large reconstruction error. In particular, reconstruction under a perspective model with

Table 5: Reconstruction errors in the presence of noise in image pixel locations. Noise level ranges from 0 to ±32 pixels.

noise (pixels)	0	1	2	4	8	16	32
noise (% image size)	0	0.2	0.4	0.8	1.6	3.2	6.4
perspective	0.038	0.073	0.133	0.265	0.510	1.007	2.005
affine	1.348	1.322	1.318	1.329	1.468	1.850	2.684
para-perspective	1.325	1.304	1.304	1.315	1.450	1.795	2.624
weak perspective	1.324	1.302	1.299	1.309	1.446	1.816	2.630

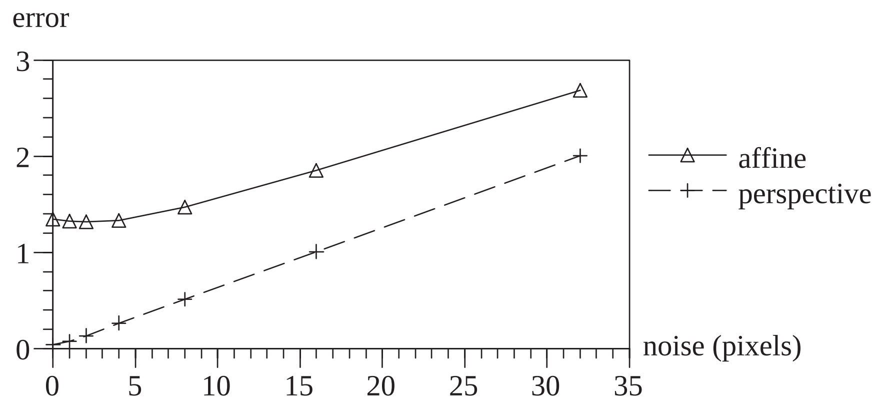


Figure 5: Reconstruction errors in the presence of noise in image pixel locations. The graphs for affine, para-perspective, and weak perspective models are very similar; only that for the affine camera is shown.

default aspect ratio or principal point results in an error that is larger than those under the other camera models with the correct parameter values (compare Tables 4 and 3). Therefore, by ignoring these parameters, the goal of using perspective camera model to obtain accurate reconstruction would have been defeated.

4.3 Experiment 3: Noise In Pixel Localization

When the camera digitizes an image, noise may affect the precise localization of image pixels. This experiment examines the effect of inaccurate pixel localization on reconstruction accuracy. Seven tests were conducted with random uniform noise ranging from 0 to ±32 pixels added to the horizontal and vertical pixel locations. With an image size of 500 pixels, the amount of noise added thus ranged from 0% to ±6.4% of the image size. The other camera parameters were set to the values given in Table 1.

The test results sumarized in Table 5 and Fig. 5 show that reconstruction under perspective camera is significantly more accurate and more robust against pixel noise than those under the other camera models. Even at a high level of noise of 16 pixels, the perspective model still performs better than the other models at no noise. Moreover, reconstruction error increases slowly with increasing noise, showing the robustness of the RI method against pixel localization error.

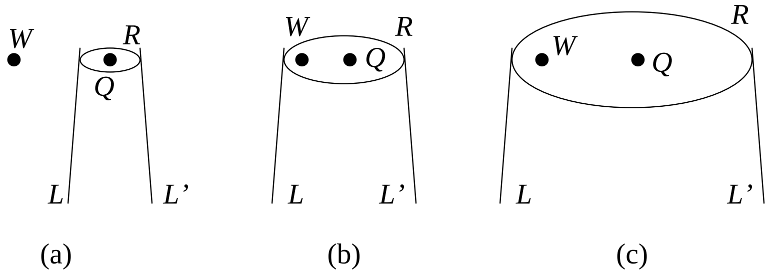


Figure 6: A conjecture of the effect of noise on reconstruction error. (a) In the absence of noise, the 3D projection lines L and L' under affine, para-perspective, and weak perspective cameras envelop a very small region R which does not include the correct world point W. The reconstruction process finds a point Q nearest to the lines, which is some distance (i.e., error) away from W. (b) With a small amount of noise, the region R grows large enough to include W, and Q is now nearer to W. (c) As the amount of noise increases further, the region R becomes so large that the nearest point Q starts moving away from W.

Reconstruction errors based on affine, para-perspective, and weak perspective models initially decrease slightly and then increase again with increasing noise. This surprising trend has been verified in two other tests that involved 50 and 100 world points randomly drawn from the same sphere. It is, therefore, a genuine phenomenon rather than a special case resulting from using only 20 random points in the test. One conjecture about this phenomenon is as follows: When there is no noise, the projection lines (Eq. 17) envelop a very small region in the 3D space that does not include the correct world point (Fig. 6). The reconstruction process amounts to finding a point in this region that is closest to all the lines, and this nearest point will be some distance away from the correct world point. With a small amount of noise, the region enveloped by the projection lines grows larger and closer to the correct world point, and reconstruction error thus decreases. However, as the noise level increases further, the region may become so large that the nearest point starts moving away from the correct world point. Therefore, reconstruction error initially decreases and then increases.

The above results provide quantitative measures of the accuracy of reconstruction under various camera models at different noise levels. On the other hand, Figs. 7 and 8 illustrate the visual quality of RI's reconstruction results. Figure 7(a) shows the wireframes of two synthetic objects. The width of the cube is 1cm. The camera is located at a distance of 5cm from the objects and oriented at 45°. Reconstruction results under perspective camera and affine camera are shown in Fig. 7(b) and 7(c), with the actual wireframes superimposed on the reconstructed ones. The results show that reconstruction under perspective camera has visually perfect results at noise levels below 8 pixels. In contrast, reconstruction under affine camera has visible error even at a low noise level of 0.5 pixel.

Figure 8 illustrates another visual quality of RI's reconstruction results. Figure 8(a) shows a synthetic checker-board pattern with a width of 2.4cm placed flat on the turntable. The camera was located at a distance of 5cm from the pattern and oriented at 45°. Reconstruction results under perspective camera and affine camera are shown in Fig. 8(b) and 8(c). Reconstruction under perspective camera has very good results at noise levels below 2 pixels. In particular, all the grid lines remain straight rather than wavy as are observed in [26, 27]. On the other hand, reconstruction under affine camera has significant error even at a low noise level of 0.5 pixel.

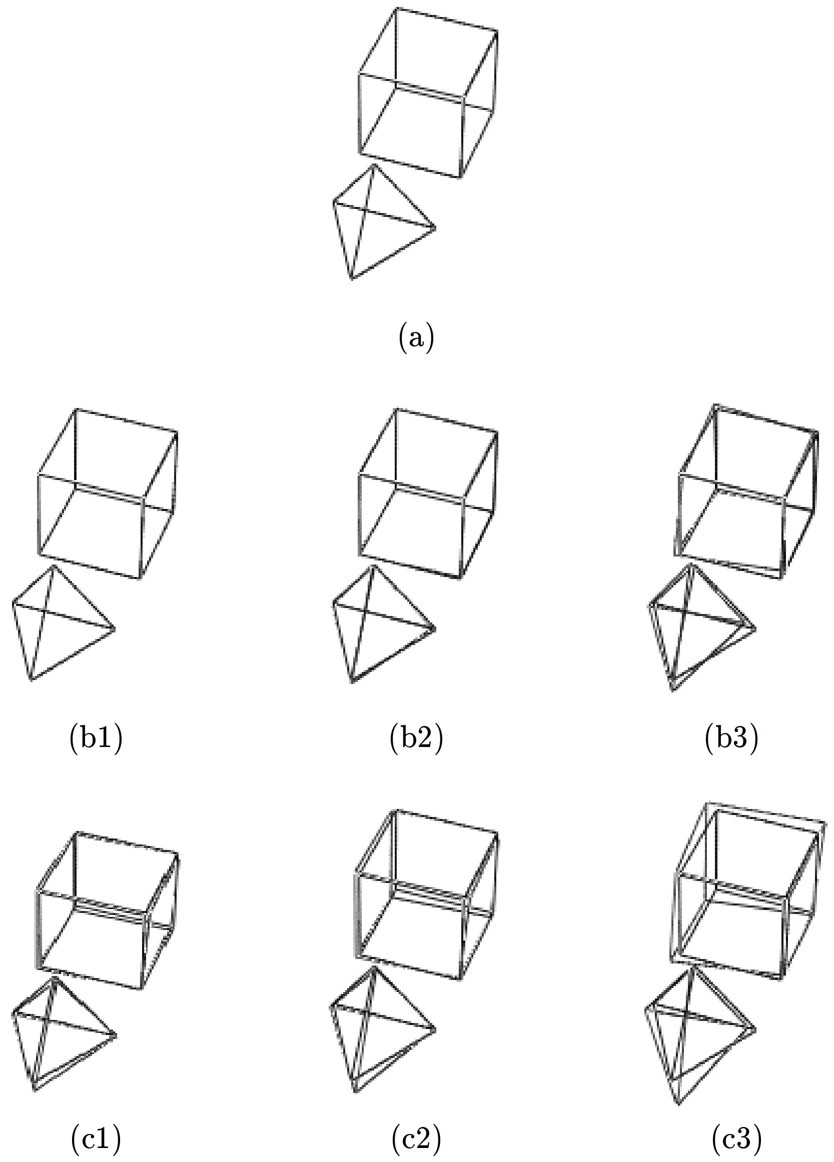


Figure 7: Qualitative comparison of the reconstruction of (a) the original objects under (b) perspective camera and (c) affine camera. Pixel localization noise of (1) 0.5, (2) 8, and (3) 32 pixels are added. Reconstruction under perspective camera has visually perfect results at noise levels below 8 pixels. On the other hand, reconstruction under affine camera has visible error even at a low noise level of 0.5 pixel.

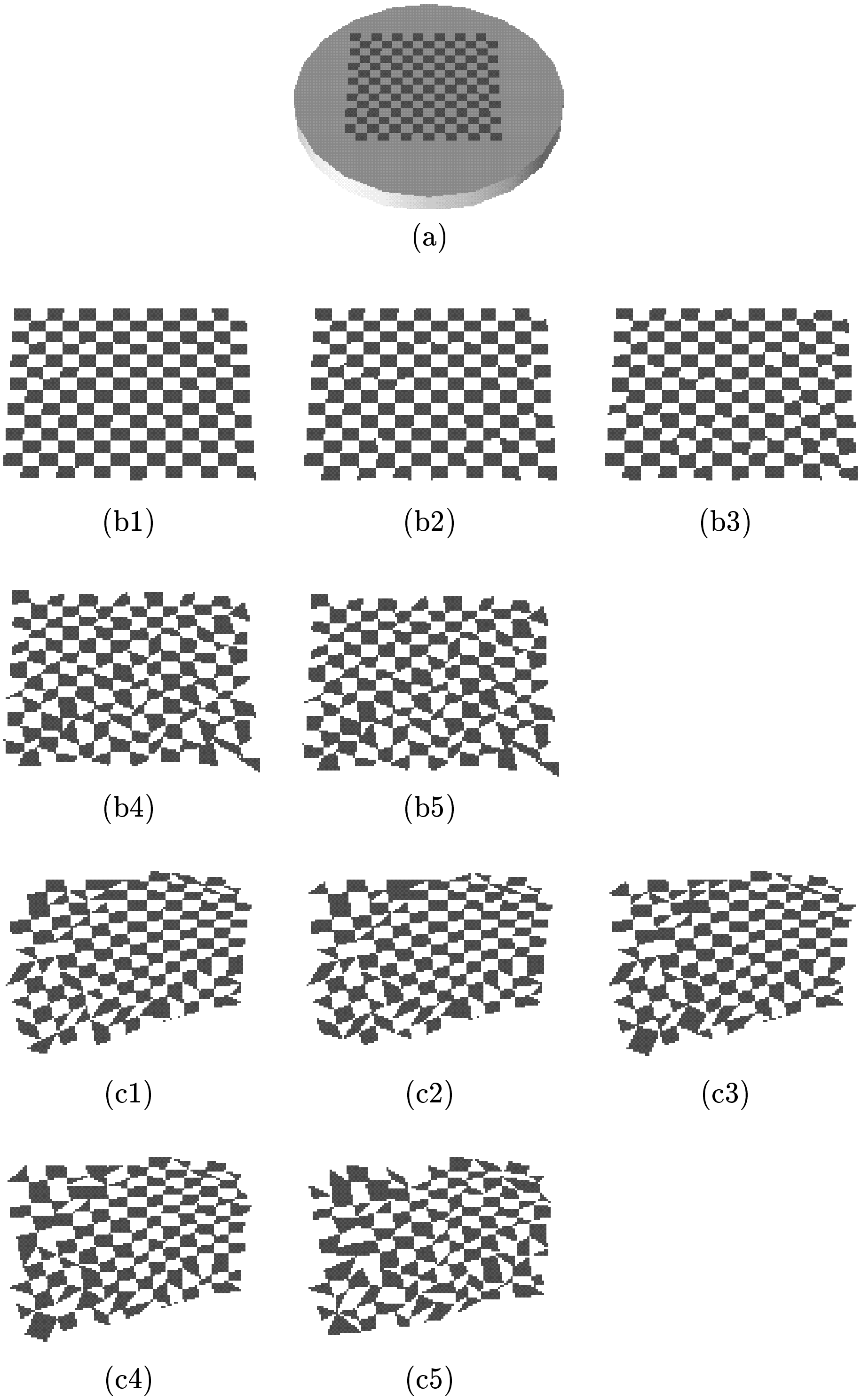


Figure 8: Visual quality of the reconstruction of (a) a checker-board pattern under (b) perspective camera and (c) affine camera at pixel localization noise of (1) 0.5, (2) 1, (3) 2, (4) 4, and (5) 8 pixels. Reconstruction under perspective camera has very good results at noise levels below 2 pixels. On the other hand, reconstruction under affine camera has significant error even at a low noise level of 0.5 pixel.

Table 6: Reconstruction error due to inaccuracy in object rotation.

noise (degree)	0.0	0.1	0.2	0.5	1.0	2.0	5.0	10.0	20.0
perspective	0.038	0.033	0.035	0.047	0.073	0.131	0.377	0.769	1.628
affine	1.348	1.346	1.348	1.354	1.322	1.277	1.247	1.288	1.570
para-perspective	1.325	1.318	1.319	1.321	1.299	1.261	1.229	1.282	1.604
weak perspective	1.324	1.322	1.324	1.330	1.300	1.256	1.231	1.278	1.582

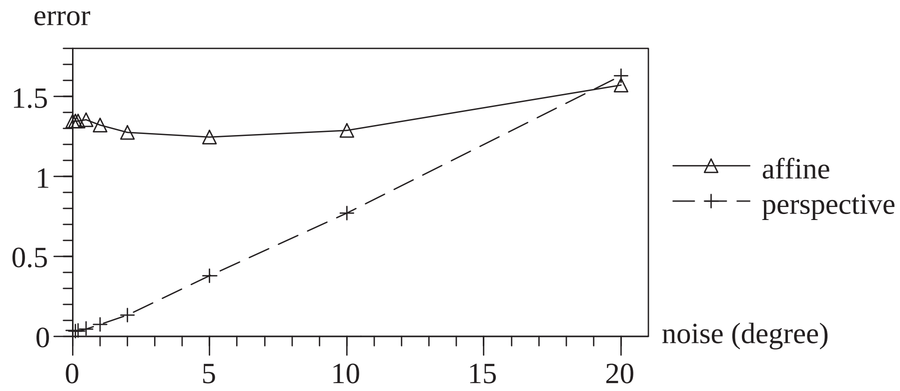


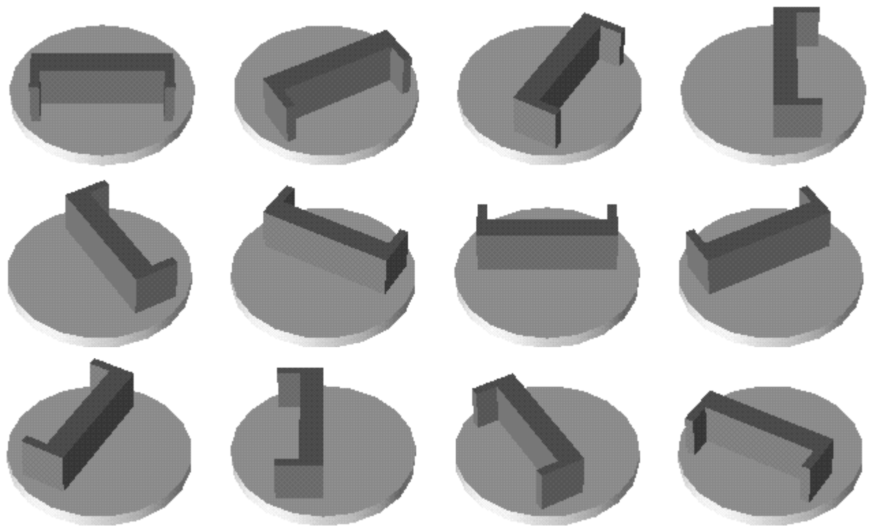
Figure 9: Reconstruction errors due to inaccuracy in object rotation. The graphs for affine, para-perspective, and weak perspective are very similar; only that for the affine camera is shown.

4.4 Experiment 4: Inaccurate Object Rotation

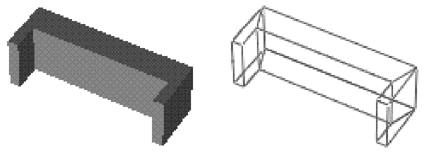
Practical turntables typically have a small amount of error in the angle of rotation. This experiment examines the effect of inaccurate object rotation on reconstruction error. The camera parameters and image points were the same as those given in Experiment 1A. Random uniform noise ranging from 0° to $\pm 20^\circ$ was added to the actual rotation angle. Test results are summarized in Table 6 and Fig. 9. Reconstruction under perspective camera again performs significantly better than those under the other camera models. Even with $\pm 10^\circ$ noise, error of reconstruction under perspective camera is still smaller than those under the other models at no noise. The test results show that RI is very robust against inaccuray in rotation angle. As in Experiment 3, errors of reconstruction under affine, para-perspective, and weak perspective cameras initially decrease slightly and then increase with increasing noise level.

4.5 Experiment 5: Occlusion and Off-Centered Rotation

As discussed in Section 2, reconstruction by shape-from-silhouette has the disadvantage that concave parts may not appear on the silhouette and, thus, may not be recovered. Figure 10 shows a synthetic U-shaped concave object placed at an off-centered position on a turntable. As the object is rotated, some parts of the object become occluded. Nevertheless, RI can correctly reconstruct such concave objects (Fig. 10b) even though the feature points on the object are visible only in some of the image frames. This test also shows that RI can perform the correct reconstruction regardless of whether the object's centroid is placed at the center of the turntable



(a)



(b)

Figure 10: Reconstruction of concave object placed at an off-centered location. (a) When a concave object is rotated, some parts of it become occluded. (b) Reconstructed result of the object shown in solid and wireframe. The apparent jaggedness of the edges is due to direct screen capture of the graphics display.

or is off-centered. In contrast, the methods in [33, 34] need to include a displacement vector in the transformation equation to handle off-centered rotation.

4.6 Experiment 6: Effects of Object's Depth

It is well known that affine and para-perspective camera models are good approximations of perspective camera when the depth variation $\max_i \Delta Z_i^c = \max_i (Z_i^c - \bar{Z}^c)$ of the world points on the object is small compared to the average depth $\bar{Z}^c$ of the object from the camera, i.e., when the depth ratio $\max_i \Delta Z_i^c / \bar{Z}^c$ is small. This experiment examines how small must the depth ratio be in order for the approximation to be valid. 2D feature points were generated in the same manner as described in Experiment 1A using the same camera parameters except that the camera's distance D_Z varied from the base case of 100cm to 5000cm. The radius of the sphere from which the world points were generated remained fixed at 20cm. Therefore, the depth ratio varied from the base case of 20% to 0.4%.

The test results, given in Table 7 and Fig. 11, show that reconstruction error based on perspective camera increases with decreasing depth ratio (i.e., increasing object depth). On the other hand, as the depth ratio decreases, the other camera models become better approximations of the perspective camera, and their reconstruction errors initially decrease. As the depth ratio decreases further, the field of view becomes so small that discretization error due

Table 7: Reconstruction errors at varying depth ratio. The affine, para-perspective, and weak perspective cameras achieve optimal performance at a depth ratio of 4%. They become good approximations of the perspective camera at depth ratio of 2%.

depth	100	200	500	1000	2000	5000
depth ratio	20%	10%	4%	2%	1%	0.4%
perspective	0.038	0.067	0.216	0.416	0.979	2.526
affine	1.348	0.675	0.385	0.448	0.965	2.517
para-perspective	1.325	0.660	0.390	0.448	0.980	2.529
weak perspective	1.324	0.667	0.385	0.448	0.965	2.517

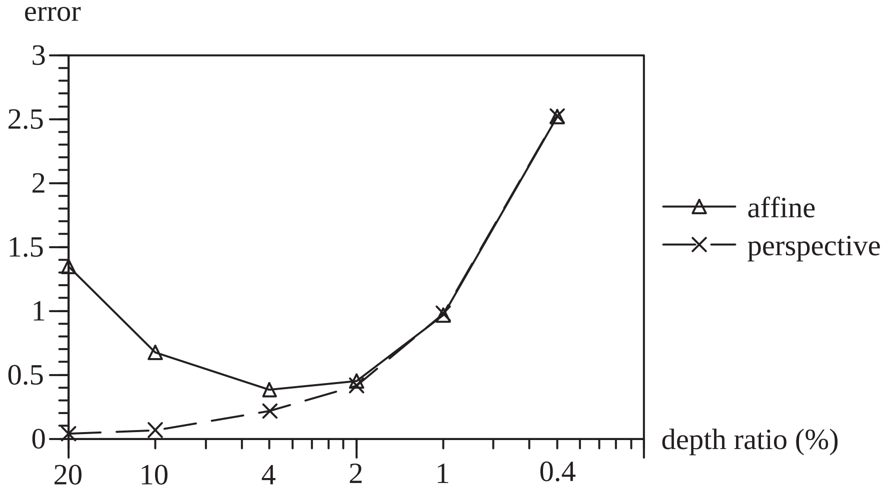


Figure 11: Reconstruction errors at varying depth ratio. The affine, para-perspective, and weak perspective cameras achieve optimal performance at a depth ratio of 4%. They become good approximations of the perspective camera at depth ratio of 2%.

to the limited resolution of the camera's digitizer begins to manifest and the reconstruction errors begin to increase. We call this critical depth ratio the *depth ratio at resolution limit*. It is dependent on the camera's focal length (or field of view) and resolution. These experimental results show that reconstruction based on affine, para-perspective, and weak perspective camera models achieve their optimal performance when the depth ratio is small enough but no smaller than the depth ratio at resolution limit. Below the depth ratio at resolution limit, reconstruction errors are approximately the same for all four camera models.

4.7 Experiment 7: Effects of Camera Focal Length

There are two methods of improving the reconstruction accuracy at small depth ratio: (1) increase camera resolution and (2) increase camera focal length. The first method requires changing the camera itself, which is a more costly option. The second method only involves changing the camera lens and is more cost effective. This experiment studies how focal length, combined with depth ratio, affects reconstruction accuracy. In this experiment, depth ratio

Table 8: Reconstruction errors at varying focal length and depth ratio. The second column gives the focal length in multiples of the base case value of 0.8. The errors of reconstruction under para-perspective and weak perspective cameras are similar to that under the affine camera, and are omitted. Combinations of depth ratios and focal lengths that resulted in the world points going out of the field of view were omitted in the experiment.

(a) Perspective camera.

depth		100	200	500	1000	2000	5000
depth ratio		20%	10%	4%	2%	1%	0.4%
focal length							
0.8	1×	0.038	0.067	0.216	0.416	0.979	2.526
1.6	2×		0.029	0.109	0.245	0.410	1.265
4	5×			0.036	0.071	0.187	0.413
8	10×				0.035	0.085	0.270
16	20×					0.039	0.116
40	50×						0.043

(b) Affine camera.

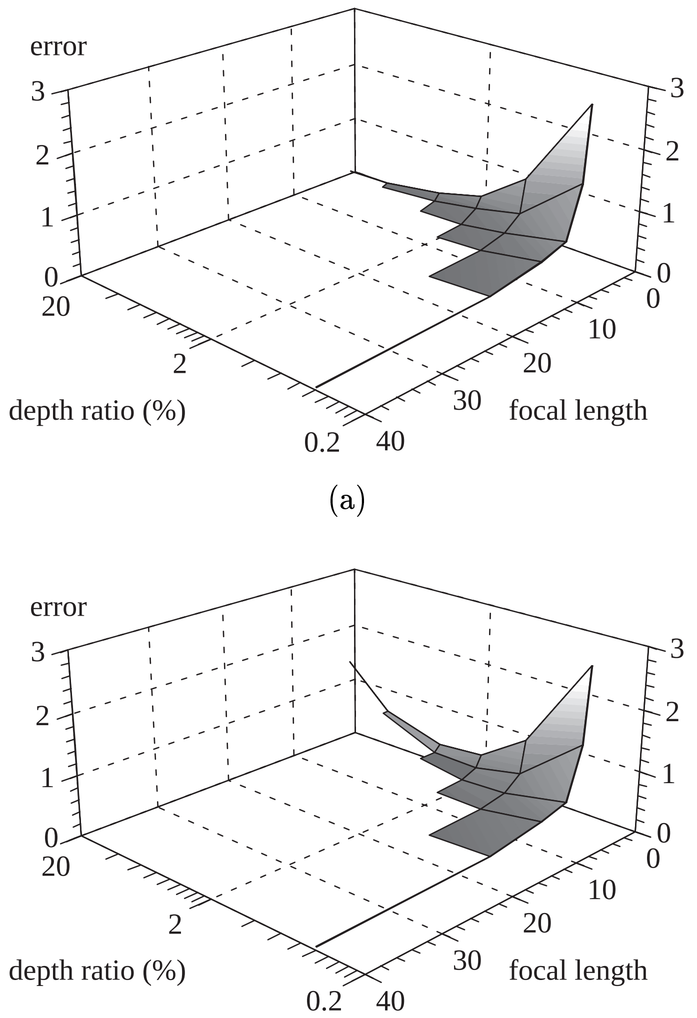
depth		100	200	500	1000	2000	5000
depth ratio		20%	10%	4%	2%	1%	0.4%
focal length							
0.8	1×	1.348	0.675	0.385	0.448	0.965	2.517
1.6	2×		0.665	0.280	0.269	0.423	1.259
4	5×			0.272	0.159	0.192	0.411
8	10×				0.128	0.109	0.279
16	20×					0.072	0.118
40	50×						0.054

varied from the base case of 20% to 0.4% (as in Experiment 6) and focal length varied from the base case of 0.8cm to 40cm. Some combinations of depth ratio and focal length were not permitted because large depth ratio (i.e., small depth) combined with large focal length would result in some of the world points going out of the field of view.

Test results are shown in Table 8 and Fig. 12. For reconstruction under affine (as well as para-perspective and weak perspective) camera, a decrease in depth ratio coupled with an increase in focal length can improve reconstruction accuracy over the base case result. Such improvement is not observed for reconstruction under perspective camera because it is already working under the correct camera model. Reconstruction accuracies under the affine, para-perspective and weak perspective cameras approach that of the perspective camera at small depth ratios and large focal length.

4.8 Experiment 8: Comparisons With Existing Methods

This section compares the performance of RI and existing methods. Although several articles have reported test results, most do not contain enough information for making fair and reason-



(a)

(b)

Figure 12: Errors of reconstruction at varying focal length and depth ratio under (a) perspective camera and (b) affine camera. The base case has a depth ratio of 20% and a focal length of 0.8. Combinations of depth ratios and focal lengths that resulted in the world points going out of the field of view were omitted in the experiment.

able comparison of performance. As demonstrated in the preceding experiments, factors such as noise in pixel localization, camera resolution, focal length, depth ratio, etc. can affect reconstruction accuracy. Therefore, performance comparison is possible, without reimplementing existing algorithms, only when sufficient information is reported in the articles.

The crucial information needed for performance comparison includes the following:

- Size of volume from which world points are drawn for testing.
- Focal length or field of view (FOV).
- Image size in pixels.

Among the existing articles, the papers by Mohr, Boufama, and Brand [20], Szeliski and Kang [33, 34], and Oliensis [24] contain enough information for making fair performance comparisons.

In the experiments, object size and camera parameters were carefully adjusted to match the test conditions reported in the articles.

4.8.1 Mohr, Boufama, and Brand

Mohr, Boufama, and Brand (MBB) [20] proposed an accurate reconstruction method based on the projective approach (Section 2). In the simulations, the world points were drawn from a volume of 162mm$\times$126mm$\times$60mm. The focal length used in the tests was 12.5mm. The reconstruction errors of their method are extracted from Table 3 of [20] and summarized in Table 8.

At first glance, a normalized error of 0.00002 seems remarkably accurate. However, we suspect that this level of accuracy is achieved in the simulation by assuming infinite camera resolution, and the error is a result of rounding off in floating-point computation. This claim explains why the normalized error jumps by 4 orders of magnitudes to 0.15 in the presence of only a small amount of noise of 0.2 pixel. This claim is verified in the following test that was performed.

In the test, the same 3D volume of sample points and camera focal length as those in [20] were used. This volume was placed at a distance of 50cm from the camera so that the whole volume fell inside the camera's field of view (about 22.6°). The camera frame was oriented parallel to the world frame and its origin was aligned with the centroid of the volume. Twenty world points were drawn from the volume for testing. The camera resolution varied from the base case of 1000 pixels/cm to infinity and noise level varied from 0 to ± 8 pixels. The other unspecified camera parameters were set to those given in Table 1.

Test results are summarized in Table 8 and Fig. 13. At a resolution of 1000 pixels/cm, reconstruction errors range from 0.10mm (with no noise) to 1.69mm (with 8 pixels of noise). Every order of magnitude increase in resolution lowers the errors by an order of magnitude. The limit is reached at a resolution of 1000000 when the rounding error of single-precision floating-point computation sets in. These test results explain the sudden drop of reconstruction accuracy of the MBB method described above. RI's reconstruction errors grow at lower rates with increasing noise than that of MBB (Fig. 13). In comparison, RI is more accurate and more robust against pixel localization noise than MBB.

4.8.2 Szeliski and Kang

The equipment setup used by Szeliski and Kang (SK) [33, 34] are similar to the RI method. In their simulation experiments, a camera with a focal length of about 1500 pixels and an FOV of 5° was used. The camera was oriented at 45° relative to the object. World points were drawn from a sphere of diameter 100 units, and they projected an image of 90$\times$90 pixels in size. Gaussian noise of various standard deviations were added to the pixel locations. The reconstruction errors, which are transcribed from Fig. 9b in [33], are summarized in Table 9.

Assuming a resolution of 1000 pixels/cm, a 90$\times$90 pixel image would have a width of 0.09cm. With this image size, a focal length of 1.03cm would provide an FOV of 5°. In our test, the sphere had a diameter of 10cm (i.e., 100 units at 1mm/unit) and its centroid was placed at $(0, 5, 0)$. The camera was oriented at 45° and placed at a distance of 150cm from the sphere's centroid so as to capture the entire sphere in its field of view. 100 world points were randomly drawn from the sphere for reconstruction test.

Test results are summarised in Table 9 and Fig. 14. At a noise level of 0.01 pixel, SK algorithm achieves a smaller error than RI algorithm. However, as the noise level increases,

Table 8: Comparison of the reconstrcution accuracy between MBB [20] and RI methods. For RI, resolution ranges from 1000 to ∞. The theoretical noise level at infinite resolution is 0. Therefore, only one test case is considered at infinite resolution. Test results for some cases are not available in [20].

resolution	noise (pixels)						
(pixels/cm)	0.0	0.2	0.5	1	2	4	8
MBB	0.00002	0.14829	0.37148	0.74624			
1000	0.10	0.13	0.13	0.21	0.42	0.83	1.69
10000	0.010	0.010	0.013	0.024	0.044	0.087	0.172
100000	0.0010	0.0012	0.0015	0.0020	0.0047	0.0088	0.0167
1000000	0.00010	0.00010	0.00010	0.00027	0.00044	0.00086	0.00174
∞	0.00010						

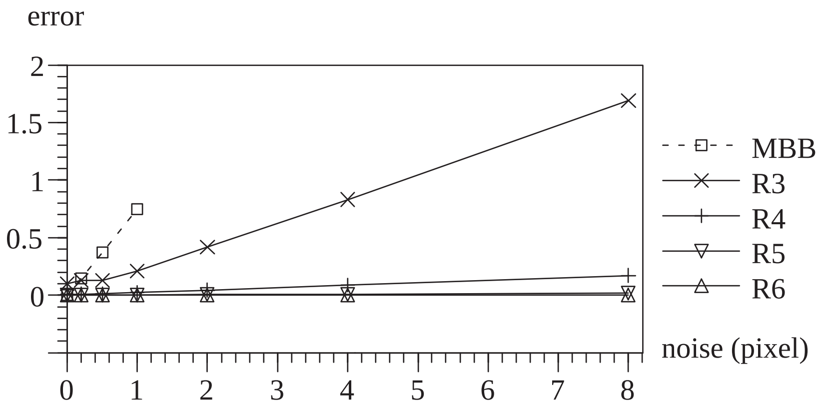


Figure 13: Comparison of the reconstrcution accuracy between MBB [20] and RI methods. R3–R6 refer to the error graphs of RI method, with Ri denoting a camera resolution of 10^i. Reconstruction errors of RI grow at lower rates than that of MBB.

RI's error grows at a much lower rate than that of SK. RI's reconstruction error is also smaller than that of SK at non-zero noise levels. RI is therefore much more accurate and robust in the presence of pixel localization noise than SK.

4.8.3 Oliensis

Oliensis's algorithm [24] performs Euclidean reconstruction under uncalibrated perspective camera using the Euclidean approach (Section 2). A total of 13 simulation experiments are reported in [24]. Experiments 1 to 5 are referred to as the easy cases which have smaller reconstruction errors than those in the other experiments. In these experiments, the image size was fixed at 512$\times$512 pixels but the FOV and depth ranges varied. For simplicity, RI is compared only with Experiments 1, 3, and 5 which have the same and the largest depth range of 20–100 among the 5 experiments. For RI, the FOV was also fixed at 60° which was the largest value among the three experiments.

At 1000 pixels/cm, the image size was calculated to be 0.512cm. This image size and an

Table 9: Comparison of reconstruction accuracy between SK [33] and RI methods. Gaussian noise of various standard deviations were used. Test results for some cases are not available in [33].

noise (pixels)	0.01	0.05	0.1	0.2	0.5	1	2	5
SK	0.14	0.77	1.43	2.77	7.02			
RI	0.48	0.47	0.45	0.51	0.74	1.38	2.44	3.82

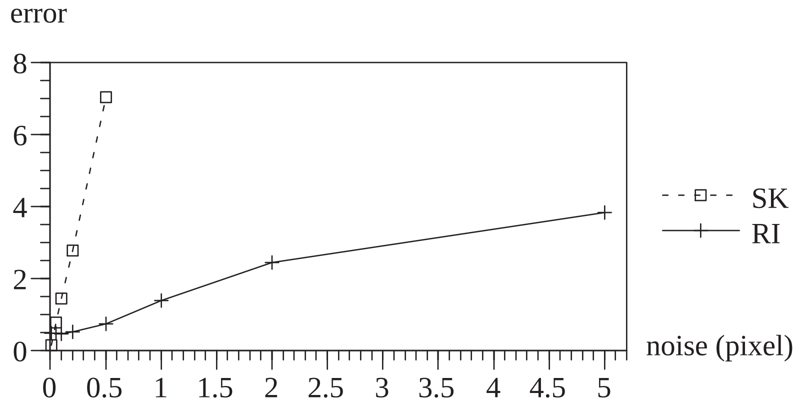


Figure 14: Comparison of reconstruction accuracy between SK [33] and RI methods. RI's reconstruction error grows at a much lower rate than that of SK.

FOV of 60° gave a focal length of 0.9cm. 30 sample points were drawn from a sphere of radius 4cm whose centroid was placed at $(0, 4, 0)$. The camera was oriented parallel to the world frame. It was positioned at a distance of 6cm from and was aligned with the sphere's centroid. At this distance, the depths of the points in the sphere would range from 20mm to 100mm matching Oliensis's experiment setup (at 1mm/unit). Gaussian and uniform noise of 0, 1, and 2 pixels were introduced. The test results summarized in Table 10 show that RI algorithm is more accurate in reconstruction compared to Oliensis's method.

5 Conclusions

This technical report focuses on the task of recovering 3D coordinates of world points from tracked 2D image points. A new method for 3D reconstruction called the Ray Intersecting (RI) method has been described. To ensure reconstruction accuracy, the RI method uses a calibrated camera and a computer-controlled turntable. It recovers the world points by computing the projection lines from the unknown world points to the known image points, backward rotating the projection lines, and then solving for the intersections between projection lines. Closed-form solutions under projective, affine, perspective, para-perspective, and weak perspective models have been derived. These close-form solutions lead to an efficient and reliable optimatization algorithm for 3D reconstruction under various camera models.

Comprehensive experiments have been performed to test RI's performance. Test results show that reconstruction under full perspective camera is significantly more accurate and more

Table 10: Comparison of reconstruction accuracy between Oliensis's algorithm [24] and the RI algorithm. U denotes uniform noise and G denotes Gaussian noise. Test results for some of the cases are not available in [24].

noise (pixels)	0	1U	2U	1G	2G
Oliensis		0.75		1.65	1.39
our	0.019	0.036	0.086	0.059	0.105

robust against noise than those under affine, para-perspective, and weak perspective camera models. Test results also show that RI is more accurate and more robust against pixel localization noise compared to existing reconstruction methods.

The RI method does not require that the tracked 2D image points be available in most of the frames in the image sequence. Instead, it uses the rotation angles of the image frames that contain the image points to perform reconstruction. Therefore, RI can be applied to recovering complete models of small objects including concave objects.

References

[1] F. S. Acton. *Numerical Methods That Work; 1990 corrected edition.* Mathematical Association of America, Washington, 1970.

[2] J. Y. Aloimonos. Perspective approximation. *Image and Vision Computing*, 8(3):179–192, 1990.

[3] P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In *Proceedings of European Conference on Computer Vision*, pages 683–695, 1996.

[4] T. J. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(6):497–513, 1991.

[5] R. Cipolla, Y. Okamoto, and Y. Kuno. Robust structure from motion using motion parallax. In *Proceedings of International Conference on Computer Vision*, pages 374–382, 1993.

[6] Brian Curless and Marc Levoy. A volumetric method for building complex models from range images. In *Proceedings of SIGGRAPH '96*, pages 303–312, 1996.

[7] C. H. Debrunner and N. Ahuja. A direct data approximation based motion estimation algorithm. In *Proceedings of International Conference on Pattern Recognition*, volume I, pages 384–389, 1990.

[8] O. D. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig. In *Proceedings of European Conference on Computer Vision*, pages 563–578, 1992.

[9] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera self-calibration: Theory and experiments. In *Proceedings of European Conference on Computer Vision*, pages 321–334, 1992.

[10] O. D. Faugeras and G. Toscani. The calibration problem for stereo. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 15–20, 1986.

[11] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 761–764, 1992.

[12] R. I. Hartley. Euclidean reconstruction from uncalibrated views. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision, LNCS 825.* Springer-Verlag, Berlin, 1994.

[13] S. B. Kang. Quasi-Euclidean recovery from unknown but complete orbital motion. Technical Report CRL 97/10, Cambridge Research Laboratory, 1997.

[14] R. Klette, K. Schlüns, and A. Koschan. *Computer Vision: Three-Dimensional Data From Images.* Springer-Verlag, Singapore, 1998.

[15] R. V. R. Kumar, A Tirumalai, and R. C. Jain. A non-linear optimatization algorithm for the estimation of structure and motion parameters. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–143, 1989.

[16] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, 1981.

[17] Q.-T. Luong, R. Deriche, O. Faugeras, and T. Papadopoulo. On determining the fundamental matrix: Analysis of different methods and experimental results. Technical Report 1894, INRIA, Sophia Antipolis, 1993.

[18] Y. Matsumoto, H. Terasaki, K. Sugimoto, and T. Arakawa. A portable three-dimensional digitizer. In *Proceedings of International Conference on Recent Advances in 3-D Digital Imaging and Modeling*, pages 197–204, 1997.

[19] D. Mehren and V. Rodehorst. Gestaltsanalyse komplexer Objekte bei kontrollierter Bewegung, Diplomarbeit, TU Berlin, Fachbereich Informatik, 1994.

[20] R. Mohr, B. Boufama, and P. Brand. Accurate projective reconstruction. In J. L. Mundy, A. Zisserman, and D. Forsyth, editors, *Applications of Invariance in Computer Vision, LNCS 825.* Springer-Verlag, Berlin, 1994.

[21] J. Oliensis. A linear solution for multiframe structure from motion. In *Proceedings of Image Understanding Workshop*, pages 1225–1231, 1994.

[22] J. Oliensis. Structure from linear or planar motions. In *CVPR*, 1996.

[23] J. Oliensis. A critique of structure from motion algorithms. Technical report, NEC Research Institute, 1998.

[24] J. Oliensis. A multi-frame structure-from-motion algorithm under perspective projection. Technical report, NEC Research Institute, 1998.

[25] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(3):206–218, 1997.

[26] M. Pollefeys, , and L. Van Gool. Self-calibration from the absolute conic on the plane at infinity. In *Proceedings of 7th International Conference on Computer Analysis of Images and Patterns*, pages 175–182, 1997.

[27] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of International Conference on Computer Vision*, 1998.

[28] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes in C, 2nd Edition.* Cambridge University Press, 1992.

[29] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(8):834–845, 1997.

[30] W. B. Seales and O. D. Faugeras. Building three-dimensional object models from image sequences. *Computer Vision and Image Understanding*, 61(3):308–324, 1995.

[31] L. S. Shapiro. *Affine Analysis of Image Sequences.* Cambridge University Press, 1995.

[32] R. Szeliski. Shape from rotation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 625–630, 1991.

[33] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. Technical report, Cambridge Research Laboratory, 1993.

[34] R. Szeliski and S. B. Kang. Recovering 3D shape and motion from image streams using non-linear least squares. *Journal of Visual Communication and Image Representation*, 5(1):10–28, 1994.

[35] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[36] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 3(4):323–344, 1987.

[37] S. Ullman. *The Interpretation of Visual Motion.* MIT Press, 1979.

[38] J. Weng, N. Ahuja, and T. S. Huang. Optimal motion and structure estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 144–152, 1989.

[39] J. Weng, T. S. Huang, and N. Ahuja. *Motion and Structure from Image Sequences.* Springer-Verlag, Berlin, 1993.

[40] Z. Zhang and A. R. Hanson. Scaled Euclidean 3D reconstruction based on externally uncalibrated cameras. In *IEEE International Symposium on Computer Vision*, pages 37–42, 1995.