February 1988

Integrating Multiple Uncertain Views of a Static Scene Acquired by an Agile Camera System

Eric Krotkov
University of Pennsylvania

Ralf Kories
University of Pennsylvania

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-88-11.

Integrating Multiple Uncertain Views of a Static Scene Acquired by an Agile Camera System

Abstract

This paper addresses the problem of merging multiple views of a static scene into a common coordinate frame, explicitly considering uncertainty. It assumes that a static world is observed by an agile vision system, whose movements are known with a limited precision, and whose observations are inaccurate and incomplete. It concentrates on acquiring uncertain three-dimensional information from multiple views, rather than on modeling or representing the information at higher levels of abstraction.

Two particular problems receive attention: identifying the transformation between two viewing positions; and understanding how errors and uncertainties propagate as a result of applying the transformation. The first is solved by identifying the forward kinematics of the agile camera system. The second is solved by first treating a measurement of camera position and orientation as a uniformly distributed random vector whose component variances are related to the resolution of the encoding potentiometers, then treating an object position measurement as a normally distributed random vector whose component variances are experimentally derived, and finally determining the uncertainty of the merged points as functions of these variances.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-88-11.

INTEGRATING MULTIPLE UNCERTAIN VIEW OF A STATIC SCENE ACQUIRED BY AN AGILE CAMERA SYSTEM

Eric Krotkov
Ralf Kories

MS-CIS-88-11
GRASP LAB 135

Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104

February 1988

Acknowledgements: This work was performed at the University of Pennsylvania and was supported in part by NSF/DCR84-10771, U.S. Air Force F49620-85-K-0018, DARPA/ONR NOO14-85-K-0807, NSF grant MCS-8219196-CER, U.S. Army grants DAA29-84-K-0061, DAA29-84-9-0027, DEC Corp., IBM Corp., LORD Corp., and the German Federal Ministry of Defense.

Integrating Multiple Uncertain Views of a Static Scene Acquired by an Agile Camera System

Eric Krotkov

Ralf Kories

ABSTRACT

This paper addresses the problem of merging multiple views of a static scene into a common coordinate frame, explicitly considering uncertainty. It assumes that a static world is observed by an agile vision system, whose movements are known with a limited precision, and whose observations are inaccurate and incomplete. It concentrates on acquiring uncertain three-dimensional information from multiple views, rather than on modeling or representing the information at higher levels of abstraction.

Two particular problems receive attention: identifying the transformation between two viewing positions; and understanding how errors and uncertainties propagate as a result of applying the transformation. The first is solved by identifying the forward kinematics of the agile camera system. The second is solved by first treating a measurement of camera position and orientation as a uniformly distributed random vector whose component variances are related to the resolution of the encoding potentiometers, then treating an object position measurement as a normally distributed random vector whose component variances are experimentally derived, and finally determining the uncertainty of the merged points as functions of these variances.

Current Affiliation of Authors

Eric Krotkov
Robotics and Artificial Intelligence Group
L.A.A.S. du C.N.R.S.
7, avenue du Colonel Roche
31077 Toulouse Cedex
France
(ADDRESS FOR CORRESPONDENCE)

Ralf Kories
Fraunhofer-Institut fur Informations- und Datenverarbeitung
Sebastian-Kneipp-Str. 12-14
D-7500 Karlsruhe
Federal Republic of Germany

Acknowledgements

This work was performed at the University of Pennsylvania and was supported in part by NSF/DCR, US Air Force, DARPA/ONR, US ARMY, NSF-CER, DEC Corp., IBM Corp., LORD Corp., and the German Federal Ministry of Defense.

Keywords

Three-dimensional computer vision, multiple views, uncertainty, machine vision, robot manipulator.

Table of Contents

1. Introduction

This paper discusses an application of of an agile camera system capable of measuring three-dimensional object locations to the problem of acquiring and relating multiple views of an unknown world. The following generic scenario indicates the importance of multiple views, and serves as the basis for our experiments.

Imagine a device (for example, a gripper attached to a robot arm, a mobile robot, a moving conveyor, etc) operating in a workspace crowded with unknown or only partially known obstacles, which may occlude themselves, each other, and the device. A vision system on or near the device is to assist this operation, perhaps by inspecting or recognizing objects of interest, perhaps by detecting impending collisions, perhaps by tracking the moving device, or perhaps by acquiring and maintaining a spatial "map" to be used for planning collision-free trajectories. For the vision system to function well in these and similar scenarios, a single view of the workspace is, in general, insufficient. For example, it may be impossible to recognize complex objects without seeing them from several aspects, and the spatial map may be incomplete (thus allowing dangerous collisions) because some objects are occluded from one viewpoint.

Because of the complexity of the environment, it will be useful to acquire data from more than one viewing position. Multiple views allow exploration of a volume potentially much larger than the visual field of view, and enable otherwise occluded objects to become at least partially visible, thus increasing the completeness of the sampling of the environment. They also afford opportunities to authenticate the existence of objects seen again and again, thus increasing the density and reliability of the observations.

This paper concentrates on one function in this scenario — constructing a spatial map of the location of objects — to take advantage of the benefits offered by multiple views. In the next section it presents the background for this work, and formally defines the problems to be solved. In section 3, it presents a method of identifying the inter-frame transformation parameters together with their uncertainty. It reviews, in section 4, a method of visually identifying

three-dimensional point locations together with their uncertainty. Next, section 5 describes a method of estimating the uncertainty on the object locations referred to a common coordinate frame. In section 6 it demonstrates the efficacy of these methods by showing quantitative experimental results for a real example. The paper concludes by critically analyzing the performance of the methods, appraising the approach, and identifying applications and extensions of the research.

2. Background and Problem Definition

This section presents the background for and an overview of our approach, describing the tools we use to move the sensors to different viewing positions, to acquire three-dimensional data, to merge data from multiple views, and to describe measurement uncertainty. It then formally defines two problems of particular interest, situating them in the literature.

2.1. Background and Overview

Multiple views could be acquired in many ways: with fixed, spatially distributed sensors; with freely wandering sensors; or with sensors attached to a mobile but constrained device such as a robot arm, or another manipulator. For the experimental purposes of this paper, the movement between different viewing positions is accomplished by the agile stereo camera system [16, 20] illustrated in figure 1. Four servomotors position and orient the stereo cameras inside a supporting gantry, allowing horizontal and vertical translations, plus pan and tilt rotations.

In addition to moving the sensors to different viewing positions, this apparatus allows the two cameras to converge and diverge, and provides control over focal length, aperture diameter, and focusing distance. Taking advantage of these features, this device can acquire three-dimensional point locations by using the technique of cooperative focus and stereo ranging (described in section 4), which has been selected as a matter of availability rather than

necessity; it could be replaced by any of a wide variety of other techniques for three-dimensional data acquisition [12].

Perhaps the simplest method of analyzing data points from multiple views is to transform them into a common reference frame and merge them. With respect to the scenario, merging the data points amounts to adding them to the map. This approach has received considerable attention in the literature, which includes efforts to construct volumetric representations [7, 11, 22, 29, 30], and to recognize objects [13, 14]. However, much of this work assumes that the inter-view transformations are known, and does not explicitly consider uncertainty. Other work on combining multiple stereo views [1] concentrates on identifying sensed data points belonging to the same physical entity (using a priori knowledge that points lie on a plane) and on "intelligently forgetting" objects seen before (by matching three-dimensional features, which also helps to refine the estimates of the inter-frame transformations), rather than simple merging without any knowledge. One criticism of using a global reference frame argues that it is not appropriate for mobile robots equipped with odometers, since the uncertainty on the vehicle position grows as the robot wanders, eventually reaching unmanageable proportions [6]. However, it will be shown in section 3 that this criticism does not apply to positioning mechanisms like the camera system, which are agile but not truly mobile, so errors do not accumulate in the common reference frame.

The scenario requires two kinds of measurements from the vision system: of the sensor location, and of object locations. Both of these problems require measuring a location parameter, say $\underline{\Theta}$, which represents the true, ideal location of sensor or object. Because of imprecise and inaccurate sensing, each measurement $\underline{\mathbf{Z}}$ of $\underline{\Theta}$ is inexact and uncertain. The measurements will be described by the additive error model: $\underline{\mathbf{Z}} = \underline{\Theta} + \underline{\mathbf{V}}$, where $\underline{\mathbf{V}}$ represents the measurement error. Since calibration methods or techniques such as subtracting the mean can eliminate systematic errors, we assume that $\underline{\mathbf{V}}$ represents only a random error with mean zero.

The literature on the estimation and representation of spatial uncertainty contains a number of different characterizations of random errors. They have been interpreted algebraically as manifolds [5], and geometrically as spatial regions [27]. They have been treated probabilistically by using specific distributions (e.g., Gaussian [3, 6]), and by a general multivariate distribution (represented, e.g., by its covariance matrix [2, 15, 23, 24] or its information matrix [10]). The literature also documents a number of approaches to combining noisy measurements and propagating errors, including algebraic approaches [4, 25] and probabilistic techniques which propagate errors differentially [8], and by using the partial derivative matrices of the inter-frame transformations [26].

For the purposes of this paper, $\underline{\mathbf{V}}$ will be characterized by normal and uniform probability density functions (pdfs), whose second moments will be derived analytically for the joint position measurements, and empirically for the object position measurements. The propagated uncertainty will be derived directly from the first two moments of the random variables representing the measurements.

2.2. Problem Definition

Multiple views raise two particular problems that will receive attention in this paper, viz., identifying the transformation between two viewing positions, and understanding how uncertainties propagate as a result of applying the transformation. In the following formal problem statements, let A and B represent the sensor's position and orientation (it's *pose*), and let the point $\underline{\mathbf{P}}$ referred to coordinate frame A be written ${}^{A}\underline{\mathbf{P}}$. The homogeneous transformation defining the rotation and translation referring A to B will be written ${}^{A}\mathbf{T}_{B}$.

Problem 1. For an object point visible as ${}^{A}\underline{\mathbf{P}}$ from A which may or may not be visible as ${}^{B}\underline{\mathbf{P}}$ from B: (a) how can the transformation ${}^{A}\mathbf{T}_{B}$ such that ${}^{B}\underline{\mathbf{P}} = {}^{A}\mathbf{T}_{B}\ {}^{A}\underline{\mathbf{P}}$ be computed from A and B; and (b) what is its uncertainty $\mathbf{U}_{T}$?

If A and B are known, as they are for our camera system, then ${}^{A}\mathbf{T}_{B}$ is derivable from the forward kinematics of the positioning mechanism, as will be shown in section 3. Problem 1a is difficult if A and B are unknown. In this case, the rigid motion parameters can be determined from visual measurements by solving the *structure from motion* problem [28], or one of its variants. This may require identifying corresponding features from distinct views, which is a challenging task. The parameters can be more easily identified in some special cases, for example, by using a known target that is visible from every viewing position, and deducing the inter-frame transformation from the observed orientation of the target [21, 31]. That the target be visible from every viewing position is an important restriction, and in complex environments this condition may be difficult to satisfy, thus limiting the generality of this approach.

Problem 1b reduces to identifying the uncertainty on each of the coefficients of ${}^{A}\mathbf{T}_{B}$. Facing the same problem, Grimson [9] derives theoretical bounds on the errors in a transformation from object model coordinates to sensor coordinates. Our approach is to derive expected errors in the transformation from one sensor pose to another by analyzing the transformation parameters as random variables to determine $\mathbf{U}_{T}$.

Problem 2. Given that ${}^{B}\underline{\mathbf{P}} = {}^{A}\mathbf{T}_{B}\ {}^{A}\underline{\mathbf{P}}$ and the uncertainty on ${}^{A}\mathbf{T}_{B}$ (i.e., the solution to Problem 1), what is the uncertainty $\underline{\mathbf{U}}_{{}^{B}\underline{\mathbf{P}}}$ on ${}^{B}\underline{\mathbf{P}}$?

Note that this is a problem of identification, not of refinement or improvement, which decomposes into two sub-problems: (a) identifying the uncertainty $\underline{\mathbf{U}}_{{}^{A}\underline{\mathbf{P}}}$ on ${}^{A}\underline{\mathbf{P}}$; and (b) identifying the propagated uncertainty on ${}^{B}\underline{\mathbf{P}}$ as a result of applying an uncertain transformation ${}^{A}\mathbf{T}_{B}$ to an uncertain point ${}^{A}\underline{\mathbf{P}}$. Earlier work has provided a solution to Problem 2a that will be reviewed in section 4. Our approach to Problem 2b, addressed in section 5, is to directly derive the variance of the transformed point ${}^{B}\underline{\mathbf{P}}$ rather than to estimate its covariance matrix.

In summary, according to the scenario of constructing a spatial map of the location of objects in a complex, static environment, the sensors move to a number of different viewpoints, and at each the vision system identifies three-dimensional object locations by cooperative ranging. Then the data points gathered from the multiple views are transformed (using the solution of Problem 1) into a common reference frame, and merged (using the solution of Problem 2) into a single map, explicitly incorporating measurement errors. Modeling the data points at higher levels of abstraction is outside the scope of this investigation.

3. Identification of the Transformation Parameters and Their Uncertainty

This section presents a solution to Problem 1 for the agile camera system. First, for Problem 1a, it describes a procedure to use measured joint positions to identify the parameters of the transformation relating one viewing position to another. Then, for Problem 1b, it analyzes the uncertainty of each of the identified parameters.

3.1. Transformation Parameter Identification

Let S be a coordinate frame representing the sensor pose, and C be the common coordinate frame to which all measurements are referred. The homogeneous transformation from S to C is

$$ {}^{S}\mathbf{T}_{C} = \begin{bmatrix} R_{11} & R_{12} & R_{13} & t_x \\ R_{21} & R_{22} & R_{23} & t_y \\ R_{31} & R_{32} & R_{33} & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{1} $$

where the R_{ij} comprise an orthonormal rotation matrix $\mathbf{R}$, and the t_i constitute a vector $\underline{\mathbf{t}}$ representing the position of the origin of S referred to C. This transformation can be used to refer measurements taken from any sensor pose to the common reference frame.

For convenience, C will be attached to and aligned with a particular frame at the "origin" of the camera system. This is no restriction, since any frame F could be defined as the common frame, as long as the transformation ${}^{C}\mathbf{T}_{F}$ such that ${}^{S}\mathbf{T}_{F} = {}^{S}\mathbf{T}_{C}\ {}^{C}\mathbf{T}_{F}$ is known. To specify the position of C, let x and y represent displacements along the horizontal and vertical axes (respectively) of the camera system gantry, with $C_x = C_y = 0$ at the centers of the respective axes, $x > 0$ and $y > 0$ for displacements to the left and the top (resp.), as seen from the front view in figure 1. We define the z-axis to make C left-handed, so that $z > 0$ for objects in the field of view, but note that no translation normal to the xy plane is mechanically possible for the camera system. To specify the orientation of C, let ϕ, θ, and ψ be the Eulerian angles under the $xy'x''$ convention corresponding respectively to tilt, pan, and roll, with $C_\phi = C_\theta = 0$ when the optic axes of the two unconverged cameras parallel the z-axis, i.e., look straight ahead. Since the cameras are mechanically unable to roll, $\psi = 0$.

Potentiometers sense the position of each servomotor, returning measurements in motor units (number of motor steps). The task now is to identify each of the transformation parameters in equation (1) from measurements of motor positions, which for the particular device are linearly related to the sensor poses.

Translation

We determine the translation vector $\underline{t} = [t_x, t_y, 0]^T$ from the camera system's horizontal and vertical servomotor positions. Let Ω_x and Ω_y (mm) denote the maximum horizontal and vertical travels of the sensors; M_{Cx} and M_{Cy} (motor units) represent the motor positions at pose C; M_x and M_y represent the motor positions sensed at pose S; M_{Ω_x} and M_{Ω_y} represent the number of possible motor positions. Defining unit steps per motor increment along each axis by

$$\delta_x = \frac{\Omega_x}{M_{\Omega_x}} \approx 1.44 \ \text{(mm/step)}, \quad \delta_y = \frac{\Omega_y}{M_{\Omega_y}} \approx 1.00 \ \text{(mm/step)} \ , \tag{2}$$

the horizontal and vertical translation components are

$$t_x = (M_x - M_{Cx})\,\delta_x \ \ (\text{mm}),\ \ t_y = (M_y - M_{Cy})\,\delta_y \ \ (\text{mm}). \tag{3}$$

Rotation

We identify the rotation matrix **R** from the camera system's pan and tilt servomotor positions. Using definitions analogous to those in the preceding section,

$$\delta_\phi = \frac{\Omega_\phi}{M_{\Omega_\phi}} \approx 0.20 \ \ (\text{deg/step}),\ \ \delta_\theta = \frac{\Omega_\theta}{M_{\Omega_\theta}} \approx 0.25 \ \ (\text{deg/step}) \ , \tag{4}$$

$$\phi = (M_\phi - M_{O\phi})\,\delta_\phi \ \ (\text{deg}),\ \ \theta = (M_\theta - M_{O\theta})\,\delta_\theta \ \ (\text{deg}) \ . \tag{5}$$

Under the $xy'x''$ Eulerian angle convention the rotation matrix is

$$\mathbf{R} = \text{Rot}(x'', 0)\ \text{Rot}(y', \theta)\ \text{Rot}(x, \phi) \ .$$

Abbreviating $\cos\theta$ by $c\theta$ and $\sin\theta$ by $s\theta$, and similarly for ϕ, equation (5) expands to

$$\mathbf{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\theta & 0 & s\theta \\ 0 & 1 & 0 \\ -s\theta & 0 & c\theta \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\phi & s\phi \\ 0 & -s\phi & c\phi \end{bmatrix} = \begin{bmatrix} c\theta & -s\theta s\phi & s\theta c\phi \\ 0 & c\phi & s\phi \\ -s\theta & -c\theta s\phi & c\theta c\phi \end{bmatrix} , \tag{6}$$

which is an orthonormal matrix with determinant +1. Thus, the rotation maintains the handedness of the coordinate frame S. If the handedness of S and C differ, we can adapt **R** by reversing the direction of any one of its eigenvectors.

3.2. Uncertainty of the Transformation Parameters

Following the approach to uncertainty described in section 2.1, we treat the measurements of the transformation parameters as random variables and derive their variances as functions of the limited resolution of the potentiometers sensing the servomotors' positions,

which experiments have shown to be much lower than the servomotors' mechanical resolution. These resolutions will be used to estimate the variances of the transformation parameters. Assembling all of these variances into one matrix, they can be expressed by

$$\mathbf{V}_T = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \sigma_{23}^2 & \sigma_{24}^2 \\ \sigma_{31}^2 & \sigma_{32}^2 & \sigma_{33}^2 & \sigma_{34}^2 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \left[\begin{array}{c|c} \mathbf{V}_R & \underline{\mathbf{V}}_t \\ \hline \underline{0} & 0 \end{array}\right], \tag{7}$$

where $\mathbf{V}_R$ is 3x3, and $\underline{\mathbf{V}}_t$ is 3x1. Note that $\mathbf{V}_T$ is not a transformation matrix. The uncertainty of the transformation parameters will be defined by a matrix $\mathbf{U}_T$, where $U_{Tij} = \sqrt{V_{Tij}}$.

Translation Uncertainty

The vector $\underline{\mathbf{V}}_t$ represents the variances of the components of translation vector $\underline{\mathbf{t}}$, which are entirely due to the resolution of the horizontal and vertical potentiometers. Using equation (3), a reading M_x of the horizontal motor position implies a translation $x = (M_x - M_{Cx})\,\delta_x$ (mm) along the x-axis. But because of the limited resolution of the potentiometer, the sensors could be at position x^* bounded by $x - \delta_x/2 \le x^* \le x + \delta_x/2$, and still produce the same motor position reading M_x.

Let us treat the actual position x^* as a random variable X. Although it is convenient to assume a normal distribution on X (reference [17, Chap. 6] considers this case), this is not particularly plausible, because there is no physical reason for the actual positions to cluster around a central value. It is more natural to assume that the actual sensor position lies, with uniform probability, anywhere within a certain interval defined by the resolution, suggesting that the pdf of X is uniform.

The uniform pdf is

$$U(X,a,b) = \begin{cases} (b-a)^{-1} & a \le x \le b \\ 0 & \text{Otherwise.} \end{cases}$$

A random variable X sampled from a uniform distribution has the following properties:

$$E[X] = \frac{a+b}{2} \quad , \quad E[X^2] = \frac{(b^3-a^3)}{3\,(b-a)} \quad , \quad Var(X) = \frac{(b-a)^2}{12} \quad .$$

From these properties, and the relation $\delta = b-a$, it follows that the variances of the translation components are

$$\underline{\mathbf{V}}_t = [V_{tx}\,, V_{ty}\,, V_{tz}\,, 0]^T = [\frac{\delta_x{}^2}{12}\,, \frac{\delta_y{}^2}{12}\,, 0\,, 0]^T \quad , \tag{8}$$

where δ_x and δ_y are given by equation (2). Using the numerical values from equation (2),

$$\underline{\mathbf{V}}_t = [0.17\,, 0.08\,, 0\,, 0]^T \quad (\mathrm{mm}^2), \quad \underline{\mathbf{U}}_t = [0.42\,, 0.28\,, 0\,, 0]^T \quad (\mathrm{mm}) \quad .$$

Rotation Uncertainty

The matrix $\mathbf{V}_R$ represents the variances of the rotation matrix coefficients, which are entirely due to the resolution of the pan and tilt potentiometers. As in the case of translation, this suggests a uniform probability distribution on the random variables describing the actual Eulerian angles. But a complication arises here, which does not occur in the translation case: while the motor positions are linearly related to the magnitude of the rotation angles, this is not the case for the sines and cosines of the angles which appear in the rotation matrix, since these trigonometric mappings are non-linear.

Using the uniform distribution, we derive the variance of $\cos(\Theta)$ and $\sin(\Theta)$ in appendix A. This suffices for the terms R_{11}, R_{22}, R_{23}, and R_{31} in equation (6), but not for R_{12}, R_{13}, R_{32}, and R_{33}, which are products of trigonometric terms. In general, if X_1 and X_2 are independent random variables, then

$$Var(X_1 X_2) = Var(X_1)\,Var(X_2) + Var(X_1)\,E^2[X_2] + Var(X_2)\,E^2[X_1] \quad . \tag{9}$$

Since θ and ϕ are independent, equation (9) can be used to compute the variance of the product terms. Using this relation, the components of $\mathbf{V}_R$ are

$$\mathbf{V}_R = \begin{bmatrix} Var(c\,\theta) & Var(-s\,\theta\ s\,\phi) & Var(s\,\theta\ c\,\phi) \\ 0 & Var(c\,\phi) & Var(s\,\phi) \\ Var(-s\,\theta) & Var(-c\,\theta\ s\,\phi) & Var(c\,\theta\ c\,\phi) \end{bmatrix}, \tag{10}$$

concluding the identification of the uncertainty on the transformation parameters.

In this formulation, the uncertainty of the transformation parameters $\mathbf{U}_T$ varies with the magnitudes of the rotations, but does not depend on the magnitudes of the translations. Note, however, that $\mathbf{U}_T$ does not vary with time. For a mobile robot equipped with odometric sensors, its position uncertainty accumulates with each instance of wheel slippage, so $\mathbf{U}_T$ would vary with time, and the formulation would have to incorporate this. For the agile camera system, the potentiometers encoding position can not "slip;" thus, identifying the transformation parameters is not a cumulative process, and the uncertainty does not depend on past parameter values.

4. Identification of Object Locations and Their Uncertainty

This section reviews a procedure solving Problem 2a presented elsewhere [17], but described here for completeness. First, it briefly presents a method for identifying three-dimensional object locations from one viewing position. Then, it presents a model of their uncertainty, again analyzing the components one by one. The final outcome is a set $\{(\underline{\mathbf{P}}, \underline{\mathbf{U}}_P)\}$ of estimated three-dimensional points, together with their estimated uncertainties.

4.1. Object Location Identification

The method for identifying point locations is a computer vision procedure in which focus ranging [18] and stereo ranging [19] operate together, cooperatively computing the three-

dimensional location of small patches of arbitrarily-shaped objects. It assumes that the scene is stationary and populated by multiple structured (not featureless) objects lying at unknown locations.

Preprocessing extracts features at positions (u, v) in the image. Focusing and stereo processes both make estimates Z_f and Z_s, respectively, of the range to the object point $\underline{\mathbf{P}} = [X, Y, Z, 1]^T$ projecting to (u, v). If one of the processes is unable to verify the range estimate of the other, both estimates are discarded; this increases the reliability of the position measurements by enforcing measurement consistency via mutual constraint. The Z component of $\underline{\mathbf{P}}$ is then computed as the maximum likelihood estimate given Z_f and Z_s; this provides an estimate of lower variance than either of the measurements alone. The X and Y components of $\underline{\mathbf{P}}$ are computed using the pin-hole lens model for a lens of focal length f as

$$X = \frac{u\,Z}{f} \quad , \quad Y = \frac{v\,Z}{f} \quad . \tag{11}$$

The procedure produces a set of points $\underline{\mathbf{P}} = [X, Y, Z, 1]^T$ referred to the sensor frame. Experiments on 75 different scenes have shown that the measurements are highly reliable and reasonably accurate.

4.2. Uncertainty of the Computed Locations

The variance σ_Z^2 of the normally distributed computed ranges Z has been experimentally determined [17, p. 116]; the square root of this figure — the expected range error σ_Z — is approximately one percent per meter: $\sigma_Z = 0.01\,Z^2$ (m). The measurement error is distance-dependent, and an uncertainty of 1 percent/m is to be interpreted as follows: for an object point 1 m away the uncertainty on its range is 1 percent, or 1 cm; for an object at 2 m distance the relative error is 2 percent, resulting in 4 cm uncertainty.

This range error propagates through to the computed X and Y components of $\underline{\mathbf{P}}$. From equation (11) it can be shown [17, p. 134] that a range error σ_Z introduces errors in the

computed X and Y positions:

$$\sigma_X = \frac{X\ \sigma_Z}{Z} \quad , \quad \sigma_Y = \frac{Y\ \sigma_Z}{Z} \quad . \tag{12}$$

Interpreting the squares of σ_X and σ_Y as the variances of the measurements of X and Y, the variance $\underline{\mathbf{V}}_P$ and uncertainty $\underline{\mathbf{U}}_P$ of each point $\underline{\mathbf{P}}$ located by cooperative ranging is

$$\underline{\mathbf{V}}_P = [\sigma_X^2\ ,\ \sigma_Y^2\ ,\ \sigma_Z^2\ ,\ 0]^T \quad , \quad \underline{\mathbf{U}}_P = [\sigma_X\ ,\ \sigma_Y\ ,\ \sigma_Z\ ,\ 0]^T \quad . \tag{13}$$

5. Estimation of the Uncertainty on the Transformed Positions

The position ${}^C\underline{\mathbf{Q}}$ referred to the common coordinate frame C of a point ${}^S\underline{\mathbf{P}}$ in sensor coordinate frame S is

$${}^C\underline{\mathbf{Q}} = {}^S\mathbf{T}_C\ {}^S\underline{\mathbf{P}} \quad . \tag{14}$$

The transformed position vector ${}^C\underline{\mathbf{Q}}$ is thus the product of an uncertain matrix and an uncertain vector. This section derives the propagated uncertainty on the transformed point. (Appendix E derives a first-order approximation of the propagated uncertainty from the covariance matrix of all the parameters.) Just as the components of ${}^C\underline{\mathbf{Q}}$ depend on ${}^S\mathbf{T}_C$ and ${}^S\underline{\mathbf{P}}$, so do the variances $\underline{\mathbf{V}}_Q$ depend on the variances of the transformation parameters $\mathbf{V}_T$ (equation 7), and the variances of the point location parameters $\underline{\mathbf{V}}_P$ (equation 13).

Let us begin with the X component of ${}^C\underline{\mathbf{Q}}$. From equation (14), $Q_X = T_{11}P_X + T_{12}P_Y + T_{13}P_Z + T_{14}$, which is abbreviated under the obvious substitutions to $Q_X = X_1 + X_2 + X_3 + X_4$. If the X_i are independent, then the variance of their sum is the sum of their variances. However, the terms P_X, P_Y, and P_Z all depend on a common parameter Z (cf. equation 11), and the terms T_{1j} , $1 \le j \le 3$, depend on the common parameters ϕ and θ (cf. equation 6). Thus, the X_i are not independent, and consequently, we must

incorporate the covariances between the X_i.

For this, let $\underline{\mathbf{X}} \equiv [X_1, X_2, X_3, X_4]^T$ and its 4x4 covariance matrix ${}^X\Lambda \equiv E[(\underline{\mathbf{X}} - E[\underline{\mathbf{X}}])(\underline{\mathbf{X}} - E[\underline{\mathbf{X}}])^T]$. The variance of Q_X is

$$Var(Q_X) = \sum_{i=1}^{4} \sum_{j=1}^{4} {}^X\Lambda_{ij} \quad . \tag{15}$$

The terms on the diagonal of ${}^X\Lambda$ are $Var(X_k)$, $1 \le k \le 4$. Since the point and transformation components are independent, each of the X_k, $1 \le k \le 3$, are products of independent variables, and equation (9) computes their variances. For k=4, $Var(X_4) = Var(T_{14})$. The off-diagonal terms involving T_{14} (i.e., ${}^X\Lambda_{4i}$ and ${}^X\Lambda_{i4}$, $1 \le i \le 3$) are zero, since the translation components are independent of the rotation components. The terms ${}^X\Lambda_{ij}$, $i \ne j$, $1 \le i, j \le 3$, are derived in appendix B. Now ${}^X\Lambda$ is completely defined, and equation (15) can be used to compute $Var(Q_X)$. The derivations of the variances of the Y and Z components proceed identically, and appendix B presents the details of their covariances.

The structure of equation (15) and the similar equations for $Var(Q_Y)$ and $Var(Q_Z)$ allow them to be expressed more descriptively by

$$\underline{\mathbf{V}}_Q = \mathbf{V}_T \underline{\mathbf{V}}_P + E^2[\mathbf{T}] \underline{\mathbf{V}}_P + \mathbf{V}_T E^2[\underline{\mathbf{P}}] + \underline{\mathbf{C}} \quad . \tag{16}$$

This formulation explicitly delineates the contributions of different error sources to the resultant uncertainty of the point position in the common reference frame. The term $\mathbf{V}_T \underline{\mathbf{V}}_P$ represents the combined contributions of the uncertainties in the transformation and the computed point. The term $E^2[\mathbf{T}] \underline{\mathbf{V}}_P$ represents the uncertainty due to applying the nominal transformation to an uncertain point. The term $\mathbf{V}_T E^2[\underline{\mathbf{P}}]$ represents the uncertainty due to applying an uncertain transformation to the nominal point position. The term $\underline{\mathbf{C}}$ represents the covariance of the parameters. This formulation is consistent with the interpretation in previous sections, that the vector $\underline{\mathbf{V}}_Q$ represents the variances of the components of ${}^C\underline{\mathbf{Q}}$, and the

vector $\underline{\mathbf{U}}_Q$ represents their uncertainties:

$$\underline{\mathbf{V}}_Q = \left[Var(Q_X), Var(Q_Y), Var(Q_Z), 0\right]^T \quad , \tag{17}$$
$$\underline{\mathbf{U}}_Q = \left[\sqrt{Var(Q_X)}, \sqrt{Var(Q_Y)}, \sqrt{Var(Q_Z)}, 0\right]^T \quad .$$

Because $\underline{\mathbf{Q}}$ is the product of $\mathbf{T}$, whose components are (in the limit of perfect potentiometer resolution) uniformly distributed, and $\underline{\mathbf{P}}$, whose components are normally distributed, the pdf of $\underline{\mathbf{Q}}$ is unknown. In a sense, this information has been lost by combining uniformly and normally distributed measurements. But for our purposes this loss is unimportant, since we use only the first two moments, which are explicitly available.

6. Experimental Results

The methods presented in the last three sections have been implemented as procedures and tested. This section presents experimental results on real data corresponding to the scenario of using multiple views to visually map a cluttered workspace in which a robot manipulator is operating. Specifically, it demonstrates the efficacy of the procedures performing multiple view analysis by observing the workspace from a number of different viewing positions, transforming the data points into the common reference frame C (cf. section 3.1), and analyzing their uncertainties.

Figure 2 illustrates the workspace used in the experiments: a gripper mounted on a robot arm, and a table cluttered with boxes. To name these objects with respect to the front view (figure 2a): the "LowBox" and "TopBox" are the two boxes lying to the left of the "Gripper;" the "Parcel" (really a parcel and the carton supporting it) lies to the right; and the "Base" (the base of the robot) lies below the Gripper. We manually measure the distances from the camera system gantry to the front surfaces of all of these objects in order to compare them as "ground truth" to the computed distances.

The experimental procedure begins by servoing the sensors to eight selected poses S_i, $1 \leq i \leq 8$, providing different views of the workspace. Figure 3 illustrates images digitized at each of the S_i, where S_1 coincides with C. Note that the Gripper is partially occluded by the TopBox from pose 5. Also note that some of the objects are either invisible (out of the common field of view) or partially visible in several of the images; for example, the Base is visible only from pose 8.

Equations (3), (5), and (6) derive matrices of the form of equation (1) for the transformations ${}^{S_i}\mathbf{T}_C$. Table 1 shows the computed transformation parameters for each of the eight poses. Only negative pan and tilt angles appear because of the geometry of the experimental setup; rotations in a positive sense would have moved the field of view away from the workspace.

At each S_i, the cooperative sensing procedure described in section 4.1 computes a set of three-dimensional points ${}^{S_i}\underline{\mathbf{P}}_j$. Over the eight poses, it locates a total of 156 points, some of which are ranged again and again (e.g., the Gripper), while others are visible with sufficient contrast from only one vantage point (e.g., the Base).

Equation (14) refers each of these points to C, creating a composite of the eight different views merged together. For the purposes of illustration, a clustering procedure [17, Chap. 7] fits bounding three-dimensional boxes around spatially coherent collections of points. (It might equally well have fit ellipsoids to the clusters, but these correspond poorly to the structure of the box-shaped objects.) Figures 4 and 5 illustrate side and top views of this composite, including both the points and their enclosing boxes. These graphs show that the structure of the scene has been preserved under all of the different transformations; they have not introduced systematic errors or bias.

However, the fidelity of the map to the workspace is not complete. Close inspection of the two figures reveals that some of the points do not line up exactly with the corresponding

labels on the ordinate axis. To quantify this "misalignment" for the Z component of the transformed points, let the range error be $\Delta_Z \equiv |Z - Q_Z|$, where Z represents the manually measured object distance expressed in frame C (we did not accurately manually measure the X and Y object positions). Table 2 presents the mean error of the N computed ranges for each object, over all eight sensor poses, as well as the standard deviation of the distribution of empirical errors. The table shows that the mean range errors increase with object distance, except for the Parcel. That this is an anomaly can be seen from the relatively large standard deviation, and can be explained by observing that some of the points on the sides of the Parcel are measured (appearing as the two points the top left of the Parcel in figure 4, which are visible from pose 2), and these points lie at different distances than the front surfaces, which are the distances labeled on the ordinate axes of the graphs, and used in the error calculation. The table also shows that the actual range errors are considerably smaller than the expected range error of one percent/m. This is a pleasant surprise, but to the extent that that figure is derived from experiments on approximately 3000 object points, we conclude that the present data comes from eight particularly accurate runs (perhaps due to the highly textured objects) rather than that the figure is too high.

For each point ${}^C\mathbf{Q}$, equation (17) computes the uncertainty $\underline{\mathbf{U}}_Q = [\sigma_{Q_X}, \sigma_{Q_Y}, \sigma_{Q_Z}, 0]^T$. We will illustrate these uncertainties graphically and analyze them numerically. For the purposes of illustration, we represent the uncertainty on a transformed point ${}^C\mathbf{Q}$ by adding six "satellite" points to the map, viz.,

$$[Q_X \pm U_{Q_X}, Q_Y, Q_Z]^T ; [Q_X, Q_Y \pm U_{Q_Y}, Q_Z]^T ; [Q_X, Q_Y, Q_Z \pm U_{Q_Z}]^T .$$

Figures 6 and 7 show top and side views of the boxes computed with and without the satellite points, i.e., with and without considering uncertainty. The figures show that the uncertainties increase with object distance, and that the Z uncertainties are considerably larger than for X

and Y.

To analyze the uncertainties numerically we compute the mean X, Y, and Z uncertainties over the N measurements of the same object at all poses, as well as the standard deviation of the uncertainty distribution. Table 3 records these statistics before applying the transformation, and table 4 records them after. Comparing these two tables in the mean uncertainty columns reveals that applying the transformation increases the X and Y uncertainties, but slightly decreases the Z uncertainty. That the uncertain transformation introduces position errors, as expected, accounts for the former. To account for the latter, surprising result, we observe that since the camera system can neither translate along nor rotate about the z axis, the Z uncertainty depends exclusively on the Z component of the object distance, which referring measurements to the common reference frame must in all cases decrease.

Both tables 3 and 4 show that the Z uncertainty is in all cases larger than the X and Y uncertainties; this is expected, since the angular resolution of the lenses is considerably better than the resolution of the cooperative ranging procedure. The mean Z uncertainties are consistent with the figure one percent/m quantified in table 2. The magnitude of the X and Y uncertainties, both before and after applying the transformation, without exception increase with Z, which reflects the dependence of the uncertainty on object distance. Before the transformation, the X and Y uncertainties depend only upon the sensed point locations in the local frame of reference; table 3 shows that this varies from object to object, as it should. After the transformation, the X and Y uncertainties depend also upon the transformation uncertainties. For example, since all points on the Base are measured from pose 8, which has a large tilt angle (cf. table 1), the mean Y uncertainty of the Base in table 4 is relatively large. As another example, the Gripper and TopBox have X uncertainties larger than Y in table 4, due to the fact that they are observed from a variety of poses where one transformation uncertainty is dominant; in the end the X uncertainty is larger, since both the horizontal and pan

axes have lower resolution than the vertical and tilt axes, respectively (cf. equations 2 and 4). We can explain the relative magnitudes of the X and Y uncertainties of the other objects similarly.

Finally, we observe that the contribution of the covariance terms $\underline{\mathbf{C}}$ in equation 16 is quantitatively negligible. Thus, the dependence on common variables of the point components and the transformation parameters has little effect on the uncertainty of the points referred to the common reference frame.

To summarize the experimental results, the empirical range errors are less than the expected range uncertainty, but consistent with the distance-dependent uncertainty model. The observed effect of applying the uncertain transformations to uncertain points is to increase their X and Y uncertainties, while changing the Z uncertainties relatively little, where the uncertainty magnitudes of all components are distance-dependent.

7. Discussion

For many visual tasks, a single view of a complex environment is not enough. This paper has formulated and presented solutions to two problems arising in the context of acquiring and merging uncertain information from multiple views of a static scene using the agile camera system as experimental support.

The solution to the first problem — how to identify the transformation between two viewing positions and its uncertainty — requires first identifying the forward kinematics of the agile camera system in order to convert motor positions into metric units, and then analyzing the position and orientation measurements as uniformly distributed random variables whose variance is related to the potentiometers' resolutions in order to quantify the transformation uncertainty. The experimental results confirm that the conversions are correctly implemented, since no systematic errors in the positions referred to the common frame appear. The

results also show a concrete example of gaining information about objects (the gripper) that are partially occluded from one viewpoint, but are visible from another, illustrating one practical benefit of multiple view analysis.

The solution to the second problem — how to estimate the uncertainty on a transformed point — requires analyzing the expected values of random variables to arrive at an estimate of the resultant uncertainty, which is the sum of terms representing (*i*) the combined transformation and point uncertainties; (*ii*) the uncertainty due to applying the nominal transformation to an uncertain point; (*iii*) the uncertainty due to applying an uncertain transformation to the nominal point; and (*iv*) the covariance between the transformation and point components. For the scenes studied, the mean uncertainties on the computed ranges are about two percent of the object distances, or approximately one percent/m. Perhaps more important than the actual values is that realistic quantitative estimates of the uncertainty in meaningful metric units are computed, in contrast to the often more theoretical results reported in the literature that still have to be proven in practice.

Although the implementation adequately demonstrates the principles of the approach to acquiring and merging uncertain information from multiple views, it is by no means a finished product. The remainder of this section discusses some improvements and extensions that might make it a more powerful system for applications such as inspecting or recognizing objects, detecting impending collisions, tracking moving objects, and visual mapping for planning collision-free trajectories.

The general approach to transformation parameter identification extends to positioning devices other than the agile camera system. For fixed sensors, one (manual) measurement of their relative positions suffices to determine the transformation parameters. For a mobile robot equipped with odometric sensors, they can be determined by trajectory integration. For sensors attached to positioning mechanisms, they are derivable from the forward kinematics of

the mechanism.

The use of the cooperative ranging technique for three-dimensional data acquisition imposes two important limitations: that the scene be static, and that the maps are quite sparse. Both of these limitations can be circumvented by using other data acquisition techniques that accommodate object motion and provide more dense data, perhaps with richer primitives than points. The basic machinery for processing multiple views would be unaffected by this.

The basic premise of referring all measurements to a global coordinate frame is inappropriate for devices equipped with odometric sensors, since the uncertainty on the transformation parameters may grow to unmanageable proportions over time. Further, to simply merge all measurements into a common reference frame is inadequate for most real applications of multiple view analysis, which may require object modeling at higher levels of abstraction. The clustering procedure illustrated in section 5 suggests one possible approach; it is not difficult to imagine using its box-like output for detecting and avoiding collisions.

Given the computed uncertainties, a natural extension is to use them to reason about the possibility or probability that a task can be successfully accomplished. We compute the first two moments of the pdf, and assume that they sufficiently describe the random errors for decision-making purposes. For some applications, more moments may be necessary, requiring probabilistic studies leading to more complete measurement models.

In summary, this work develops and demonstrates one approach to acquiring and merging uncertain three-dimensional information from multiple views. With refinement and development, it may allow the many applications of machine vision to profit from the practical advantages of processing information from multiple views.

Appendix A. Variances of Cosine and Sine, Uniformly Distributed

This appendix derives the variances of the functions cosine and sine of a uniformly distributed random variable Θ. In addition, it derives the first moment of their product.

Definition 1. The expectation of the random variable X with pdf $p(X)$ is

$$E[X] = \int_{-\infty}^{+\infty} x\ p(x)\ dx \quad .$$

Definition 2. The expectation of the function g applied to the random variable X is

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)\ p(x)\ dx \quad .$$

Definition 3. The variance of $g(X)$ is

$$Var(g(X)) = E[g(X)^2] - E^2[g(X)] \quad .$$

Definition 4. The pdf of a random variable Θ uniformly distributed on the interval $[a, b]$ is

$$U(\Theta, a, b) = \begin{cases} (b-a)^{-1} & a \le \theta \le b \\ 0 & \text{Otherwise.} \end{cases}$$

Definition 5. $b - a = \delta\ ;\ \Delta = \dfrac{\delta}{2}\ ;\ a = \theta - \Delta\ ;\ b = \theta + \Delta$.

Identity 1. $\sin(x \pm y) = \sin x\ \cos y \pm \cos x\ \sin y$.

Identity 2. $\cos(x + y) = \cos x\ \cos y - \sin x\ \sin y\ ;\ \cos(x - y) = \cos x\ \cos y + \sin x\ \sin y$.

Identity 3. $\sin 2x = 2 \sin x\ \cos x$.

Identity 4. $\cos 2x = \cos^2 x - \sin^2 x$.

Let us start by considering the first moment of the cosine function. By definition 2,

$$E[\cos\Theta] = \int_{-\infty}^{+\infty} \cos\theta \; U(\theta, a, b) \; d\theta \quad .$$

Substituting the uniform pdf from definition 4 and integrating yields

$$E[\cos\Theta] = \frac{1}{b-a} \int_{a}^{b} \cos\theta \; d\theta = \frac{\sin b \; - \sin a}{b-a} \quad .$$

Using definition 5, then identity 1, followed by algebraic manipulation leads to

$$\sin b \; - \sin a \; = \sin(\theta+\Delta) - \sin(\theta-\Delta) = 2\cos\theta \sin\Delta \quad ,$$

$$E[\cos\Theta] = \frac{2\cos\theta \sin\Delta}{\delta} \quad . \tag{A.1}$$

Now we consider the second moment of the cosine function. By definition 2,

$$E[\cos^2\Theta] = \int_{-\infty}^{+\infty} \cos^2\theta \; U(\theta, a, b) \; d\theta \quad .$$

Substituting the uniform pdf from definition 4 and integrating yields

$$E[\cos^2\Theta] = \frac{1}{b-a} \int_{a}^{b} \cos^2\theta \; d\theta = \frac{b-a \; + \sin b \cos b \; - \sin a \cos a}{2(b-a)} \quad .$$

Using definition 5, plus identities 1, 2, and 3, algebraic manipulation arrives at

$$\sin b \cos b \; - \sin a \cos a \; = \cos 2\theta \sin 2\Delta \quad ,$$

$$E[\cos^2\Theta] = \frac{\delta + \cos 2\theta \sin 2\Delta}{2\delta} \quad . \tag{A.2}$$

Finally, using equations (A.1) and (A.2) and definition 3, the variance of $\cos\Theta$ is

$$Var(\cos\Theta) = E[\cos^2\Theta] - E^2[\cos\Theta] \quad . \tag{A.3}$$

The derivation of $Var(\sin\Theta)$ is quite similar to the derivation of $Var(\cos\Theta)$. Omitting the details, the results are

$$E[\sin\Theta] = \frac{2\sin\theta\sin\Delta}{\delta} \quad , \tag{A.4}$$

$$E[\sin^2\Theta] = \frac{\delta - \cos 2\theta \sin 2\Delta}{2\delta} \quad , \tag{A.5}$$

$$Var(\sin\Theta) = E[\sin^2\Theta] - E^2[\sin\Theta] \quad . \tag{A.6}$$

This concludes the derivations of the variances. Now we consider the first moment of the product $\cos\Theta\sin\Theta$. From definition 2,

$$E[\cos\Theta\sin\Theta] = \int_{-\infty}^{+\infty} \cos\theta\ \sin\theta\ U(\theta, a, b)\ d\theta \quad .$$

Substituting the uniform pdf and integrating yields

$$E[\cos\Theta\sin\Theta] = \frac{1}{\delta}\int_{a}^{b} \cos\theta\ \sin\theta\ d\theta = \frac{\sin^2 b\ - \sin^2 a}{2\delta} \quad . \tag{A.7}$$

Some simplification of this equation results from considering the order of magnitude of certain variables in the experimental setup. As described in section 3.1, the value of Δ (half the potentiometer resolution) does not exceed $0.125°$. The difference $\sin 0.125° - 0.125° < 1.8x\,10^{-9}$; thus, the small angle approximation that $\sin\Delta \approx \Delta$ is extremely accurate. Similarly, the value of $\cos 0.125° = 1 - 2.4x\,10^{-6}$; for most purposes, this can be safely treated as unity. Making the small angle approximation that $\sin\Delta \approx \Delta$,

$$E[\cos\Theta\sin\Theta] \approx \sin\theta\cos\theta\cos\Delta \quad ,$$

and approximating $\cos\Delta \approx 1$, this reduces to

$$E[\cos\Theta\sin\Theta] \approx \cos\theta\sin\theta \quad , \tag{A.8}$$

with an accuracy on the order of one part in one million.

Appendix B. Covariances for the Transformed Point

This appendix derives the covariances required to compute the variances of the components of the transformed point, as described in section 5. It uses results on the covariance between the point components (appendix C) and between the transformation parameters (appendix D). Appendix E develops another approach, which approximately computes the covariance *between* the transformed point coordinates.

Definition 6. Let A_i represent any random variables. The covariance of A_i and A_j is

$$Cov(A_i, A_j) \equiv E[A_i A_j] - E[A_i]\, E[A_j] \quad .$$

Note that if A_i and A_j are independent, then $Cov(A_i, A_j) = 0$. If $i = j$, then $Cov(A_i, A_j) = Var(A_i)$. Also, it follows from the definition that the relation is reflexive, i.e., $Cov(A_i, A_j) = Cov(A_j, A_i)$.

Definition 7. Let $\underline{\mathbf{A}} \equiv [A_1, A_2, A_3]^T$. The covariance matrix for $\underline{\mathbf{A}}$ is defined by ${}^{A}\Lambda \equiv E[(\underline{\mathbf{A}} - E[\underline{\mathbf{A}}])(\underline{\mathbf{A}} - E[\underline{\mathbf{A}}])^T]$. In matrix form, the covariance matrix is

$${}^{A}\Lambda = \begin{bmatrix} Var(A_1) & Cov(A_1, A_2) & Cov(A_1, A_3) \\ Cov(A_2, A_1) & Var(A_2) & Cov(A_2, A_3) \\ Cov(A_3, A_1) & Cov(A_3, A_2) & Var(A_3) \end{bmatrix} .$$

Note that if the A_i are pairwise independent, then ${}^{A}\Lambda$ is a diagonal matrix. If our notation is to be consistent with section 5, then $\underline{\mathbf{A}}$ should have four elements. However, as described there, the covariances due to the fourth component are zero, so for simplicity we treat $\underline{\mathbf{A}}$ as a three-vector.

The diagonal terms of the covariance matrix are derived in section 5. The task here is to compute the off-diagonal terms, ${}^{X}\Lambda_{ij}$, ${}^{Y}\Lambda_{ij}$, ${}^{Z}\Lambda_{ij}$, $i \neq j$, $1 \leq i, j \leq 3$.

Let us begin with the covariances for the X component of the transformed point. From

section 5,

$$X_1 \equiv T_{11} P_X = \cos\theta \frac{u Z}{f} \;\; ; \;\; X_2 \equiv T_{12} P_Y = -\sin\theta \sin\phi \frac{v Z}{f} \;\; ; \;\; X_3 \equiv T_{13} P_Z = \sin\theta \cos\phi Z \;\; .$$

We proceed to compute $Cov(X_1,X_2)$, $Cov(X_2,X_3)$, and $Cov(X_1,X_3)$.

First we consider $Cov(X_1,X_2) \equiv Cov(T_{11}P_X, T_{12}P_Y)$. Since the transformation parameters and the location parameters are independent,

$$E[X_1]E[X_2] = E[T_{11}P_X]\,E[T_{12}P_Y] = E[T_{11}]E[P_X]E[T_{12}]E[P_Y] \;\; .$$

Since the covariance of T_{11} and T_{12} is very small (as in appendix D, on the order of 10^{-6}), the product of the expectations closely approximates the expectation of the products, and the previous expression can be rewritten approximately as

$$E[X_1]E[X_2] = E[T_{11}]E[P_X]E[T_{12}]E[P_Y] \approx E[T_{11}T_{12}]\,E[P_X]E[P_Y] \;\; . \qquad \text{(B.1)}$$

Since the transformation and point location parameters are independent, it follows that

$$E[T_{11}P_X\,T_{12}P_Y] = E[T_{11}T_{12}]\,E[P_X\,P_Y] \;\; . \qquad \text{(B.2)}$$

Subtracting (B.1) from (B.2) and applying definition 6 yields

$$Cov(X_1,X_2) \approx E[T_{11}T_{12}]\,Cov(P_X,P_Y) \;\; ,$$

where equations A.1, A4, and A.7 derive the expectation, and appendix C derives $Cov(P_X,P_Y)$.

Analyses parallel to the above for $Cov(X_1,X_3)$ and $Cov(X_2,X_3)$ arrive at

$$Cov(X_1,X_3) \approx E[T_{11},T_{13}]\,Cov(P_X,P_Z) \;\; ,$$
$$Cov(X_2,X_3) \approx E[T_{12},T_{13}]\,Cov(P_Y,P_Z) \;\; ,$$

completing the analysis for the X component. Using the reflexive property of definition 6,

and writing C_{ij} for $Cov(i, j)$, the covariance matrix is

$$ {}^{X}\Lambda \approx \begin{bmatrix} Var(X_1) & E[T_{11}, T_{12}]\, C_{XY} & E[T_{11}, T_{13}]\, C_{XZ} \\ E[T_{12}, T_{11}]\, C_{XY} & Var(X_2) & E[T_{12} T_{13}]\, C_{YZ} \\ E[T_{13}, T_{11}]\, C_{XZ} & E[T_{13} T_{12}]\, C_{YZ} & Var(X_3) \end{bmatrix} . $$

Derivations of the covariances for the Y and Z components of the transformed point follow the same pattern, yielding

$$ {}^{Y}\Lambda \approx \begin{bmatrix} Var(Y_1) & 0 & 0 \\ 0 & Var(Y_2) & E[T_{23} T_{22}]\, C_{YZ} \\ 0 & E[T_{22} T_{23}]\, C_{YZ} & Var(Y_3) \end{bmatrix} , $$

$$ {}^{Z}\Lambda \approx \begin{bmatrix} Var(Z_1) & E[T_{31} T_{32}\, C_{XY} & E[T_{31} T_{33}]\, C_{XZ} \\ E[T_{32} T_{31}]\, C_{XY} & Var(Z_2) & E[T_{32} T_{33}]\, C_{YZ} \\ E[T_{33} T_{31}]\, C_{XZ} & E[T_{33} T_{32}]\, C_{YZ} & Var(Z_3) \end{bmatrix} , $$

concluding the identification of the covariances for the transformed point.

Appendix C. Covariances Between the Point Components

This appendix derives the covariances between the point components

$$P_X = \frac{u\,Z}{f} \quad , \quad P_Y = \frac{v\,Z}{f} \quad , \quad P_Z = Z \quad ,$$

described in section 4. From this equation, and the fact that u and f are constants, the first moment of P_X can be written $E[P_X] \equiv E[\frac{u}{f}\,Z] = \frac{u}{f}E[Z]$, and the first moment of P_Z is $E[P_Z] \equiv E[Z]$. It follows that the product of the first moments is $E[P_X]\,E[P_Z] = \frac{u}{f}E^2[Z]$, and that the first moment of the product is $E[P_X\,P_Z] = \frac{u}{f}E[Z^2]$. Subtracting the two previous equations yields

$$E[P_X\,P_Z] - E[P_X]\,E[P_Z] = \frac{u}{f}\,(E[Z^2] - E^2[Z]) \quad ,$$

which, using the definitions of variance (definition 3) and covariance (definition 6), can be rewritten as

$$Cov(P_X\,,P_Z) = \frac{u}{f}\,Var(Z) \quad .$$

Analysis of the YZ and XY covariances proceeds as above, and yields

$$Cov(P_Y\,,P_Z) = \frac{v}{f}\,Var(Z) \quad , \quad Cov(P_X\,,P_Y) = \frac{u\,v}{f^2}\,Var(Z) \quad .$$

Appendix D. Covariances Between the Transformation Parameters

This appendix derives the covariances between the transformation parameters, which we assume are functions of uniformly distributed random variables. As in section 3, the transformation matrix is

$$\mathbf{T} = \begin{bmatrix} c\theta & -s\theta s\phi & s\theta c\phi & t_x \\ 0 & c\phi & s\phi & t_y \\ -s\theta & -c\theta s\phi & c\theta c\phi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} . \tag{D.1}$$

Since components only from the first row of $\mathbf{T}$ contribute to the variance of the X component of the transformed point (cf. section 5 and appendix B), and similarly for the other rows and point components, it suffices to consider the covariance between parameters in the same row of $\mathbf{T}$. In addition, since the translation component (the last element of the first three rows) is independent of the rotation terms, the covariance between these terms must be zero, so we can restrict the analysis to the first three columns of the first three rows of $\mathbf{T}$. Thus, the analysis proceeds in three steps, one for each row.

It will be useful to first calculate the covariance between $\sin\Theta$ and $\cos\Theta$, since many of the derived covariances will contain this term. From equations (A.1) and (A.4), and the small-angle approximation of $\sin\Delta$ (which appendix A has shown to be highly accurate), the product of the first moments is

$$E[\cos\Theta]\,E[\sin\Theta] = \frac{4\cos\theta\sin\theta\sin^2\Delta}{\delta^2} \approx \cos\theta\sin\theta \quad . \tag{D.2}$$

Equation (A.8) expresses the first moment of the product. Subtracting the two yields

$$Cov(\cos\Theta, \sin\Theta) \approx \cos\theta\sin\theta\,(1 - \cos\Delta) \quad . \tag{D.3}$$

As in appendix A, $\cos\Delta \approx 1$, so $Cov(\cos\Theta, \sin\Theta) \approx 0$.

With the preliminaries finished, let us turn to the covariances between the terms in the first row of equation (D.1). The covariance of the first and second components is by definition $Cov(T_{11}, T_{12}) \equiv Cov(\cos\Theta, -\sin\Theta\sin\Phi)$. Since Θ and Φ are independent, the product of the expectations is

$$-E[\cos\Theta]\,E[\sin\Theta\sin\Phi] = -E[\cos\Theta]\,E[\sin\Theta]\,E[\sin\Phi] \quad .$$

Again using the independence of Θ and Φ, the expectation of the product is

$$-E[\cos\Theta\sin\Theta\sin\Phi] = -E[\cos\Theta\sin\Theta]\,E[\sin\Phi] \quad .$$

Subtracting the product of the expectations from the expectation of the product and substituting the numerical value from equation (D.3) gives

$$Cov(T_{11}, T_{12}) = -E[\sin\Phi]\;Cov(\cos\Theta, \sin\Theta) \approx 0 \quad . \tag{D.4}$$

The covariance of the first and third components is by definition $Cov(T_{11}, T_{13}) \equiv Cov(\cos\Theta, \sin\Theta\cos\Phi)$, which using the same kind of analysis as above, yields

$$Cov(T_{11}, T_{13}) = E[\cos\Phi]\;Cov(\cos\Theta, \sin\Theta) \approx 0 \quad . \tag{D.5}$$

The covariance between the second and third components is by definition $Cov(T_{12}, T_{13}) \equiv Cov(-\sin\Theta\sin\Phi, \sin\Theta\cos\Phi)$. Since Θ and Φ are independent, the product of the expectations is

$$-E[\sin\Theta\sin\Phi]\,E[\sin\Theta\cos\Phi] = -E^2[\sin\Theta]\,E[\sin\Phi]\,E[\cos\Phi] \quad .$$

Again using this independence, the expectation of the product is

$$-E[\sin\Theta\sin\Phi\sin\Theta\cos\Phi] = -E[\sin^2\Theta]\,E[\sin\Phi]\,E[\cos\Phi] \quad .$$

Subtracting the product of the expectations from the expectation of the product gives

$$Cov(T_{12}, T_{13}) = -E[\sin\Phi]\,E[\cos\Phi]\;Var(\sin\Theta) \quad , \tag{D.6}$$

completing the identification of the covariances between terms in the first row.

In the second row of equation (D.1), the covariances of the first and second components and the first and third components are zero, since $T_{21} = 0$. From equation (D.3), the covariance between the second and third components is approximately zero.

In the third row of equation (D.1), the covariances of the first and second components and the first and third components are the same as for the first row (equations D.4 and D.5) but with a sign change. The covariance between the second and third components is the same as for the first row (equation D.6), but with $Var(\sin\Theta)$ replaced by $Var(\cos\Theta)$.

Appendix E. First-Order Uncertainty of the Transformed Positions

As an alternative approach to Problem 2 (defined in section 2), this appendix derives a first-order approximation of the propagated uncertainty on the transformed point. The approach involves what Ayache and Faugeras call "linearizing the problem," [2] and what Smith and Cheeseman call "approximating to first-order" [26], by working with the (approximate) covariance matrix of measured parameters. This differs from the approach taken in section 5, which works with the (exact) covariance matrices using derived parameters (e.g., $\cos\theta$ rather than θ). Since we investigated the approximate approach, we devote this appendix to its formulation, but have not experimentally studied it in any detail. One reason for this is some unresolved questions, which appear at the end of the appendix.

First, developing the formalism, we can rewrite equation (14) as

$$\underline{y} = f(\underline{x}) = f(t_x, t_y, \phi, \theta, P_X, P_Y, P_Z) \quad ,$$

where f is the nonlinear function described by the right-hand side of equation (14). The first-order Taylor series expansion of f is

$$\underline{y} = f(\hat{\underline{x}}) + J(\underline{x} - \hat{\underline{x}}) + \cdots \quad , \tag{E.1}$$

where J is the Jacobian matrix of partial derivatives $J \equiv \frac{\partial f(\underline{x})}{\partial \underline{x}}(\hat{\underline{x}})$.

The mean, to first order, is the function applied to the parameter means

$$\hat{\underline{y}} \approx f(\hat{\underline{x}}) \quad . \tag{E.2}$$

By definition, the covariance matrix is $C(\underline{y}) = E[(\underline{y} - \hat{\underline{y}})(\underline{y} - \hat{\underline{y}})^T]$. From equations (E.1) and (E.2), $C(\underline{y}) = E[(J\ (\underline{x} - \hat{\underline{x}})\ (J\ (\underline{x} - \hat{\underline{x}})^T]$. Using a transpose identity, this can be rewritten as $C(\underline{y}) = E[(J\ (\underline{x} - \hat{\underline{x}})\ (\underline{x} - \hat{\underline{x}})^T\ J^T]$. If the Jacobian is independent of the parameters, then the product of the expectations is identical to the expectation of the products:

$$C(\underline{y}) = E[J]\, E[(\underline{x} - \hat{\underline{x}})\,(\underline{x} - \hat{\underline{x}})^T]\, E[J^T] \quad . \tag{E.3}$$

From the definition of covariance, this is equivalent to $C(\underline{y}) = E[J]\, C(\underline{x})\, E[J^T]$. The expectation of the Jacobian is equal to the Jacobian, since J does not depend on the statistical properties of $\underline{x}$, so

$$C(\underline{y}) = J\; C(\underline{x})\; J^T \quad . \tag{E.4}$$

Now we apply this formalism to the problem of computing the propagated uncertainty on the transformed point. Differentiating, the Jacobian is

$$J = \begin{bmatrix} 1 & 0 & -P_Y s\theta c\phi - P_Z s\theta s\phi & Q_Z & c\theta & -s\theta s\phi & s\theta c\phi \\ 0 & 1 & -P_Y s\phi + P_Z c\phi & 0 & 0 & c\phi & s\phi \\ 0 & 0 & -P_Y c\theta c\phi - P_Z c\theta s\phi & -Q_X - t_x & -s\theta & -c\theta s\phi & c\theta c\phi \end{bmatrix} .$$

The covariance between the parameters is described by the matrix

$$C(\underline{x}) = \begin{bmatrix} \Lambda_t & 0 & 0 \\ 0 & \Lambda_R & 0 \\ 0 & 0 & \Lambda_P \end{bmatrix} ,$$

where

$$\Lambda_t = \begin{bmatrix} Var(t_x) & 0 \\ 0 & Var(t_y) \end{bmatrix} , \quad \Lambda_R = \begin{bmatrix} Var(\phi) & 0 \\ 0 & Var(\theta) \end{bmatrix} ,$$

$$\Lambda_P = \begin{bmatrix} Var(P_X) & Cov(X,Y) & Cov(X,Z) \\ Cov(Y,X) & Var(P_Y) & Cov(Y,Z) \\ Cov(Z,X) & Cov(Z,Y) & Var(P_Z) \end{bmatrix} .$$

The matrix $C(\underline{x})$ is diagonal because the translation, rotation, and point parameters are mutually independent. The 3x3 covariance matrix representing the uncertainty of the transformed

point is given by equation (E.4).

The vector $\underline{\mathbf{V}}_Q$ representing the variances of the components of ${}^C\underline{\mathbf{Q}}$ (cf. section 5) is related to the diagonal terms of the covariance matrix $C(\underline{y})$ by

$$\underline{\mathbf{V}}_Q = \left[Var(Q_X), Var(Q_Y), Var(Q_Z), 0\right]^T = \left[C_{11}(\underline{y}), C_{22}(\underline{y}), C_{33}(\underline{y}), 0\right]^T \quad .$$

The two important differences between this approach and that presented in section 5 are that the former computes the off-diagonal terms while the latter does not, and that the former is approximate while the latter is exact. This leads us to some questions which we have not yet answered. With respect to the first difference, how are the off-diagonal terms to be used in the assessment of the uncertainty on the components of the transformed point? And if they are not, then why bother to compute them? With respect to the second difference, how good is the first-order approximation in equation (E.1)? Is the Jacobian independent of the parameters, so that we can derive equation (E.3)? And finally, how do the results of the two approaches differ?

References

1. Ayache, N. and O. D. Faugeras, "Building a Consistent 3D Representation of a Mobile Robot Environment by Combining Multiple Stereo Views," *Proc. Tenth International Joint Conference on Artificial Intelligence*, pp. 808-810, Milano, Italy, August, 1987.
2. Ayache, N. and O. D. Faugeras, "Building, Registrating, and Fusing Noisy Visual Maps," *Proc. First International Conference on Computer Vision*, pp. 73-82, IEEE Computer Society, London, June, 1987.
3. Bolle, R. M. and D. B. Cooper, "On Optimally Combining Pieces of Information, With Application to Estimating 3-D Complex-Object Position from Range Data," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 619-638, 1986.
4. Brooks, R. A., "Symbolic Reasoning Among 3-D Models and 2-D Images," *Artificial Intelligence*, vol. 17, pp. 285-348, 1981.
5. Brooks, R. A., "Visual Map Making for a Mobile Robot," *Proc. IEEE Conference on Robotics and Automation*, pp. 824-829, St. Louis, March, 1985.
6. Chatila, R. and J.-P. Laumond, "Position Referencing and Consistent World Modeling for Mobile Robots," *Proc. IEEE Conference on Robotics and Automation*, pp. 138-145, St. Louis, March, 1985.
7. Connolly, C. I., "Cumulative Generation of Octree Models from Range Data," *Proc. IEEE Conference on Robotics and Automation*, pp. 25-32, Atlanta, March, 1984.
8. Durrant-Whyte, H. F., "Integration, Coordination and Control of Multi-Sensor Robot Systems," *Ph.D. Dissertation MS-CIS-86-67*, Systems Engineering Department, University of Pennsylvania, Philadelphia, 1986.
9. Grimson, W. E. L., "Sensing Strategies for Disambiguating Among Multiple Objects in Known Poses," *IEEE Journal of Robotics and Automation*, vol. 2, no. 4, pp. 196-213, December, 1986.
10. Hager, G., "Active Reduction of Uncertainty in Multi-Sensor Systems," *Technical Report MS-CIS-86-76*, Computer Science Department, University of Pennsylvania, Philadelphia, 1986.
11. Hong, T. H. and M. O. Shneier, "Describing a Robot's Workspace Using a Sequence of Views from a Moving Camera," *Industrial Systems Division, National Bureau of Standards*, Washington, D.C., 1984.
12. Kanade, T., *Three-Dimensional Machine Vision,* Kluwer Academic Publishers, 1987.
13. Kim, H. S., R. C. Jain, and R. A. Volz, "Object Recognition Using Multiple Views," *Proc. IEEE Conference on Robotics and Automation*, pp. 28-33, St. Louis, March, 1985.
14. Korn, M. R. and C. R. Dyer, "3-D Multiview Object Representations for Model-Based Object Recognition," *Research Report RC11760*, IBM, March, 1986.
15. Kriegman, D. J. and E. Triendl, "A Mobile Robot: Sensing, Planning and Locomotion," *Proc. IEEE Conference on Robotics and Automation*, pp. 402-408, Raleigh, March, 1987.
16. Krotkov, E., J. F. Summers, and F. Fuma, "The Pennsylvania Active Camera System," *Technical Report MS-CIS-86-15*, Computer Science Department, University of Pennsylvania, Philadelphia, 1986.
17. Krotkov, E., "Exploratory Visual Sensing for Determining Spatial Layout with an Agile Stereo Camera System," *Ph.D. Dissertation MS-CIS-87-29*, Computer Science Department, University of Pennsylvania, Philadelphia, May, 1987.

18. Krotkov, E., "Focusing," *International Journal of Computer Vision*, vol. 1, no. 3, pp. 223-237, Kluwer Academic Publishers, Boston, October, 1987.
19. Krotkov, E. and R. Kories, "Ranging Errors with Verging Stereo Cameras," *Proc. DAGM-Symposium Mustererkennung 1987, Fachberichte Informatik 149*, pp. 217-222, Springer, Berlin-Heidelberg-New York, September, 1987.
20. Krotkov, E., J. F. Summers, and F. Fuma, "An Agile Stereo Camera System for Flexible Image Acquisition," *IEEE Journal of Robotics and Automation*, In press, 1988.
21. Lenz, R. K. and R. Y. Tsai, "Techniques of Calibration of the Scale Factor and Image Center for High Accuracy 3D Machine Vision Metrology," *IBM Research Report RC 54867*, October, 1986.
22. Martin, W. and J. K. Aggarwal, "Volumetric Descriptions of Objects from Multiple Views," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 150-158, 1983.
23. Matthies, L. and S. Shafer, "Error Modeling in Stereo Navigation," *Technical Report CS-86-140*, Carnegie-Mellon University, 1986.
24. Matthies, L. and T. Kanade, "Using Uncertainty Models In Visual Motion and Depth Estimation," *Preprints Fourth International Symposium of Robotics Research*, Santa Cruz, August, 1987.
25. O'Rourke, J. and N. I. Badler, "Model-Based Image Analysis of Human Motion Using Constraint Propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 2, no. 4, November, 1980.
26. Smith, R. and P. Cheeseman, "On the Representation and Estimation of Spatial Uncertainty," *International Journal of Robotics Research*, vol. 5, no. 4, pp. 56-68, Winter, 1987.
27. Snyder, M. A., "Uncertainty Analysis of Image Measurements," *Proc. ARPA Image Understanding Workshop*, pp. 681-693, Los Angeles, February, 1987.
28. Ullman, S., *The Interpretation of Visual Motion,* MIT Press, Cambridge, 1979.
29. Underwood, S. A. and C. L. Coates, "Visual Learning from Multiple Views," *IEEE Transactions on Computers*, vol. C, no. 24, pp. 651-661, 1975.
30. Veenstra, J. and N. Ahuja, "Octree Generation from Silhouette Views of an Object," *Proc. IEEE Conference on Robotics and Automation*, pp. 843-848, St. Louis, March, 1985.
31. Vemuri, B. C. and J. K. Aggarwal, "3-D Model Construction from Multiple Views Using Range and Intensity Data," *Proc. Computer Vision and Pattern Recognition Conference*, pp. 435-437, Miami Beach, June, 1986.

Table 1. Computed Transformation Parameters

Pose	x [mm]	y [mm]	ϕ [deg]	θ [deg]	Objects Visible	N
1	0.0	0.0	0.0	0.0	Gripper	23
2	153.0	0.0	0.0	0.0	Parcel, Gripper	11
3	153.0	-130.0	0.0	0.0	Parcel, Gripper	22
4	0.0	-130.0	0.0	0.0	TopBox, Gripper	25
5	-174.0	-130.0	0.0	0.0	TopBox	19
6	0.0	0.0	-3.4	0.0	TopBox, Gripper	17
7	0.0	0.0	-11.6	-2.1	LowBox	16
8	168.0	161.0	-11.6	-2.1	Base, Parcel	23

The computed transformation parameters relating the different sensor poses, where x corresponds to horizontal translations, y to vertical translations, ϕ to tilt rotations, and θ to pan rotations. The number N indicates how many points are ranged at each pose. Pose 1 coincides with the common reference frame C. For pose 2, the camera translates to the right, for pose 3 it translates downward, for pose 4 it translates back to the left, and for pose 5 continues to the left. For pose 6, it returns to pose 1 and tilts downward, and for pose 7 tilts farther down, and pans to the left. For pose 8 it maintains this orientation, but translates up and to the right.

Table 2. Mean Eror of Transformed Points

Object	Z [mm]	σ_Z^* [mm]	N	$\mu(\Delta_Z)$ [mm]	$\sigma(\Delta_Z)$ [mm]
LowBox	1701	28.9	16	12.2	11.2
TopBox	1701	28.9	36	10.6	10.6
Parcel	2133	45.5	44	27.0	29.6
Gripper	2510	63.0	56	20.4	13.7
Base	2820	79.5	4	32.7	5.6

The table first lists the Z component of the (manually measured) object distances expressed in the common reference frame C, and the expected range error σ_Z^* of 1 percent/m (cf. section 4.2) for that object distance. $\mu(\Delta_Z)$ represents the mean empirical error over the N points sensed on each object, taken over all eight sensor poses. $\sigma(\Delta_Z)$ represents the standard deviation of the distribution of errors over all sensed points.

Table 3. Mean Uncertainty of Points in Local Frame

Object	N	$\mu(\sigma_X)$ [mm]	$\mu(\sigma_Y)$ [mm]	$\mu(\sigma_Z)$ [mm]	$\sigma(\sigma_X)$ [mm]	$\sigma(\sigma_Y)$ [mm]	$\sigma(\sigma_Z)$ [mm]
LowBox	16	0.7	0.6	30.0	0.4	0.3	0.5
TopBox	36	0.9	0.5	29.2	0.4	0.2	0.5
Parcel	44	0.8	0.9	45.3	0.5	0.4	1.7
Gripper	56	1.2	1.3	62.7	0.8	0.6	1.2
Base	4	2.7	1.3	81.5	1.0	0.4	1.2

Table 4. Mean Uncertainty of Transformed Points

Object	N	$\mu(\sigma_X)$ [mm]	$\mu(\sigma_Y)$ [mm]	$\mu(\sigma_Z)$ [mm]	$\sigma(\sigma_X)$ [mm]	$\sigma(\sigma_Y)$ [mm]	$\sigma(\sigma_Z)$ [mm]
LowBox	16	4.5	7.0	29.4	0.1	0.1	0.5
TopBox	36	4.5	3.6	29.2	0.1	0.2	0.5
Parcel	44	5.6	6.9	44.8	0.1	2.8	1.7
Gripper	56	6.5	5.5	62.7	0.2	0.4	1.2
Base	4	8.2	17.4	79.8	0.3	0.2	1.2

The tables list the mean uncertainties $\mu(\sigma_X)$, $\mu(\sigma_Y)$, and $\mu(\sigma_Z)$, over the N points sensed on each object, computed over all eight sensor poses, and the standard deviations of the distribution of uncertainties $\sigma(\sigma_X)$, $\sigma(\sigma_Y)$, $\sigma(\sigma_Z)$, over all sensed points.

Figure Captions

Figure 1. The Agile Camera System.

Figure 2. Photographs of the Test Scene. This figure illustrates the test scene: a gripper attached to a robot arm, and a table cluttered with objects.

Figure 3. Digitized Images of the Test Scene. This figure illustrates images digitized at each of the eight sensor poses (described in table 1). The white overlaid box represents the field of view common to the focusing and stereo procedures; only object points inside the box are measured. The tape measure used for "ground truth" is visible at the bottom of the images from poses 7 and 8.

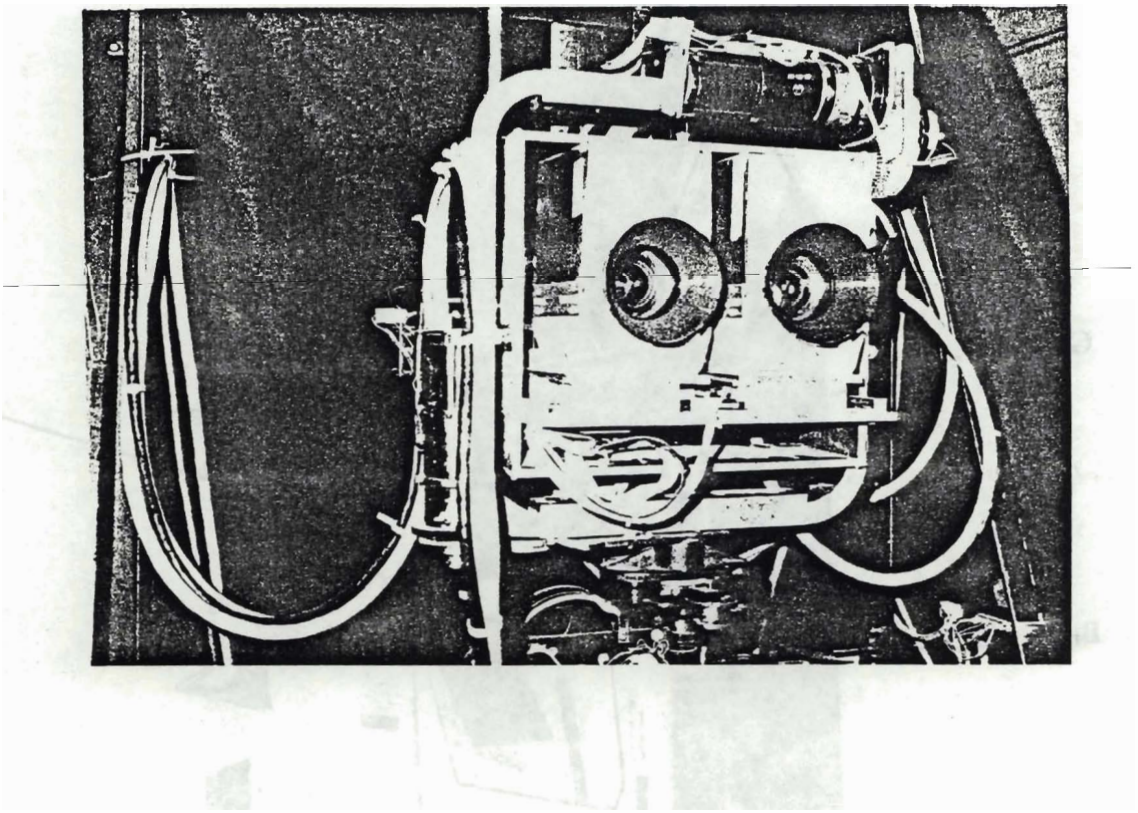
Figure 4. Top View of Points Referred to the Common Frame. This figure plots a top view of the computed range map. As described in section 3, the origin of the underlying left-handed coordinate frame is at the center of the right camera lens; the positive z-axis increases toward the objects, the x-axis points from the left to the right of the digitized images, and the y-axis points from bottom to top. The rectangles represent the output of the clustering procedure.

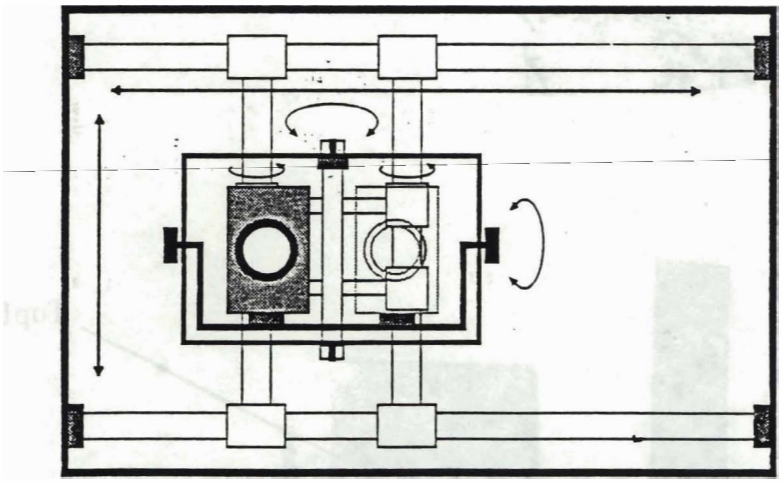
Figure 5. Side View of Points Referred to the Common Frame. This figure plots a side view of the computed range map, using the same coordinate frame as Figure 4.

Figure 6. Measurement Uncertainties, Top View. Top view of the computed range map including measurement uncertainties. The solid rectangles represent the output of the clustering procedure, and the dashed rectangles represent the worst-case uncertainties (the best-case is not illustrated for aesthetic reasons).

Figure 7. Measurement Uncertainties, Side View. Side view of the computed range map including measurement uncertainties.

Figure 1. The Agile Camera System.





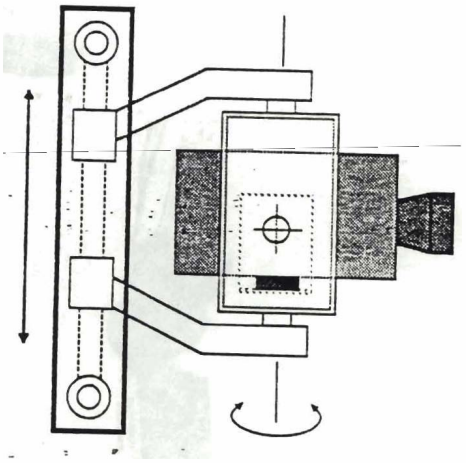


Figure 2. Photographs of the Test Scene.

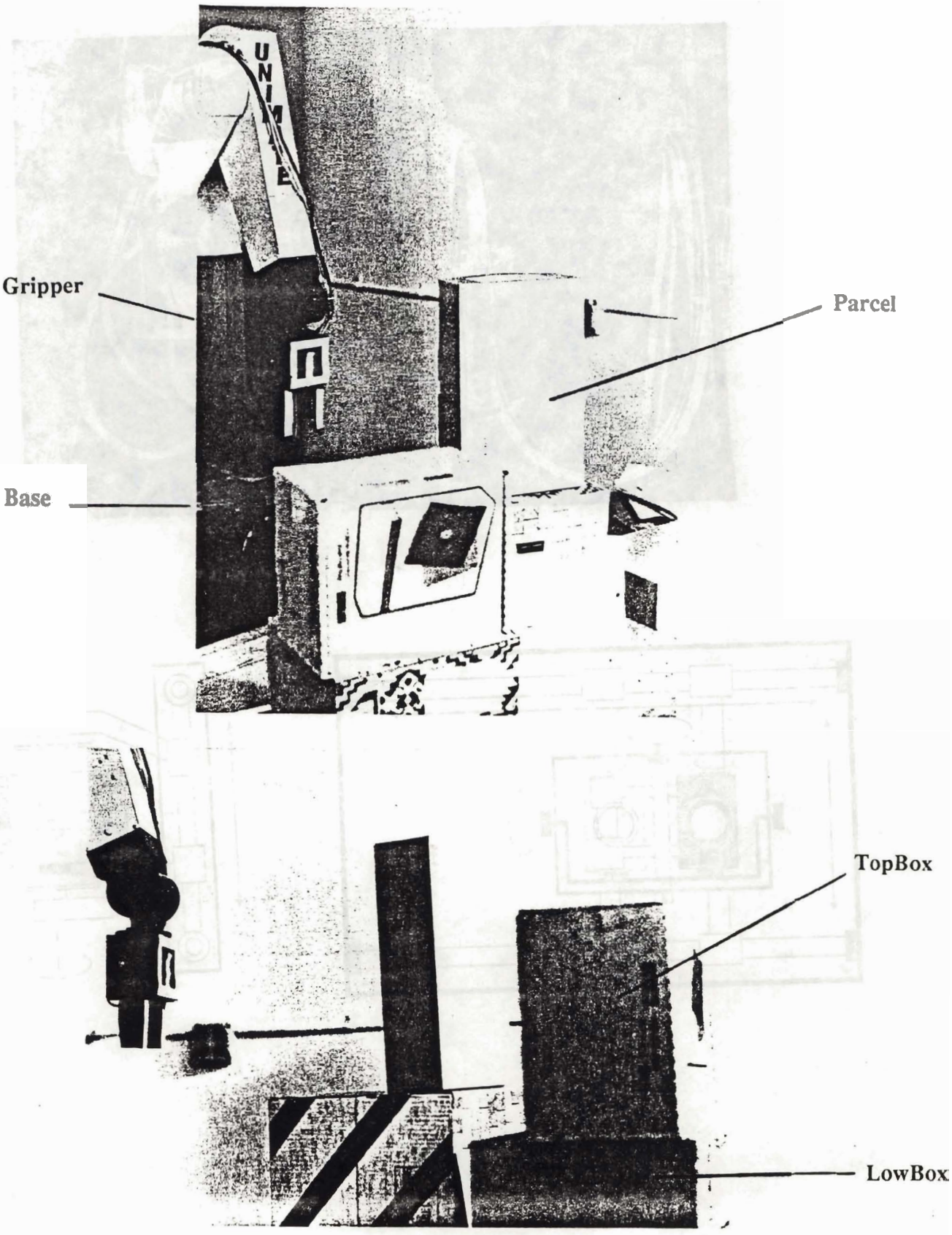


Figure 3. Digitized Images of the Test Scene.

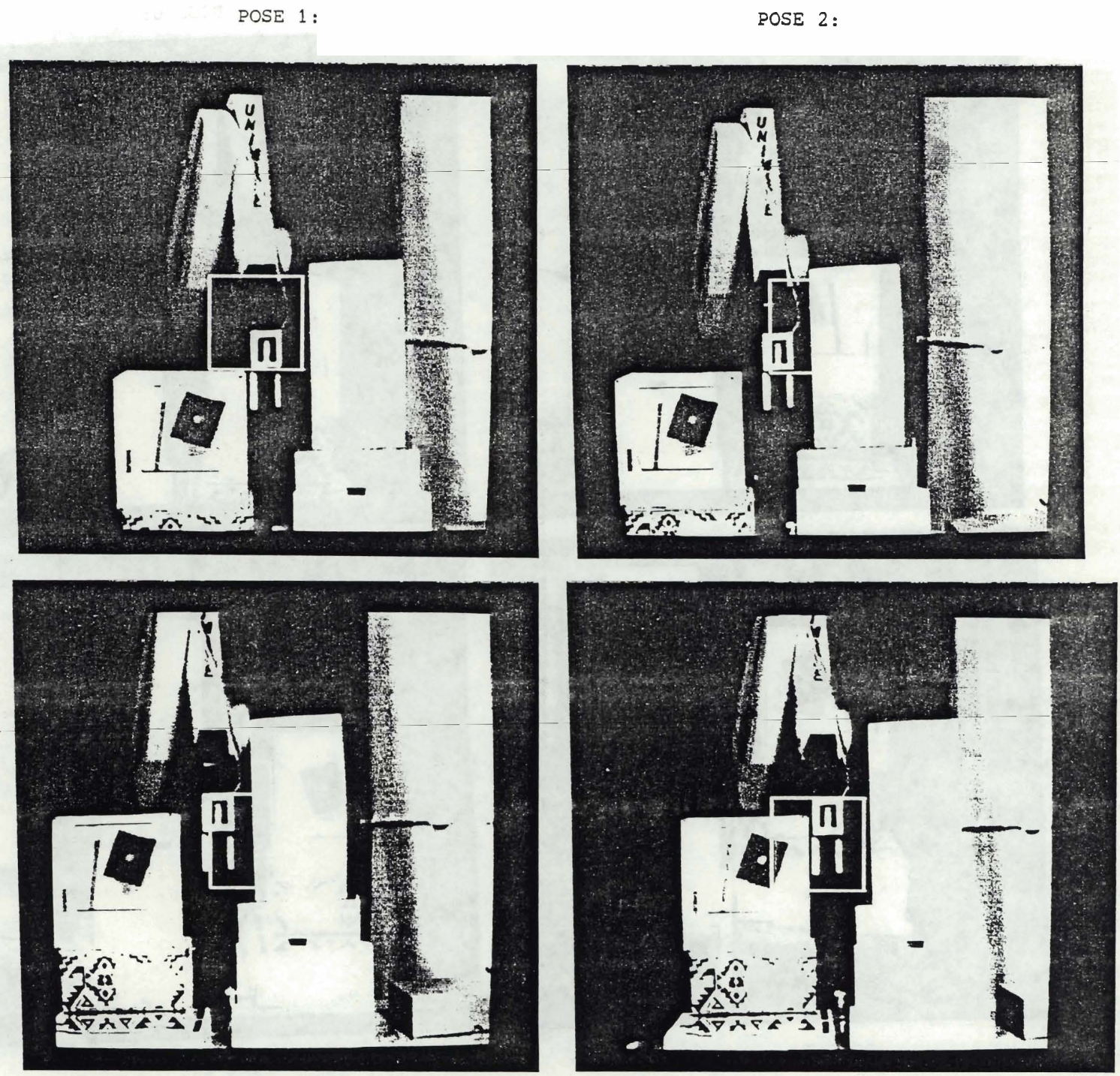
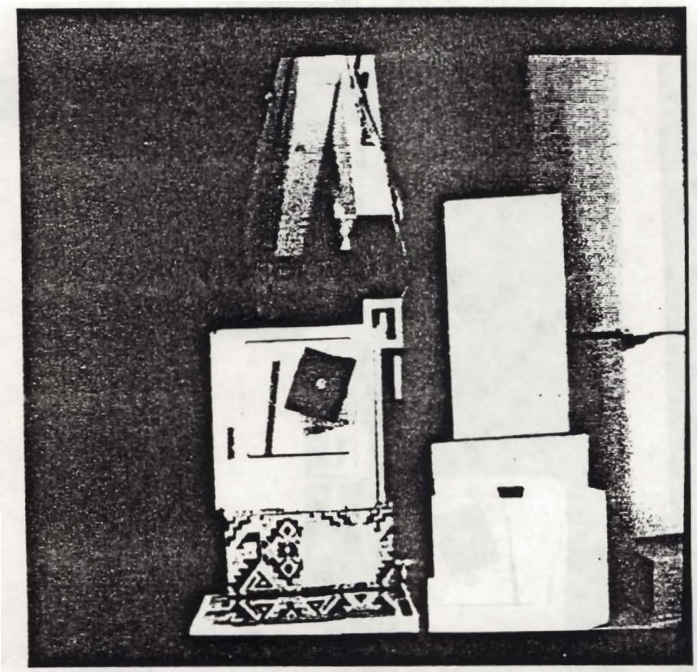
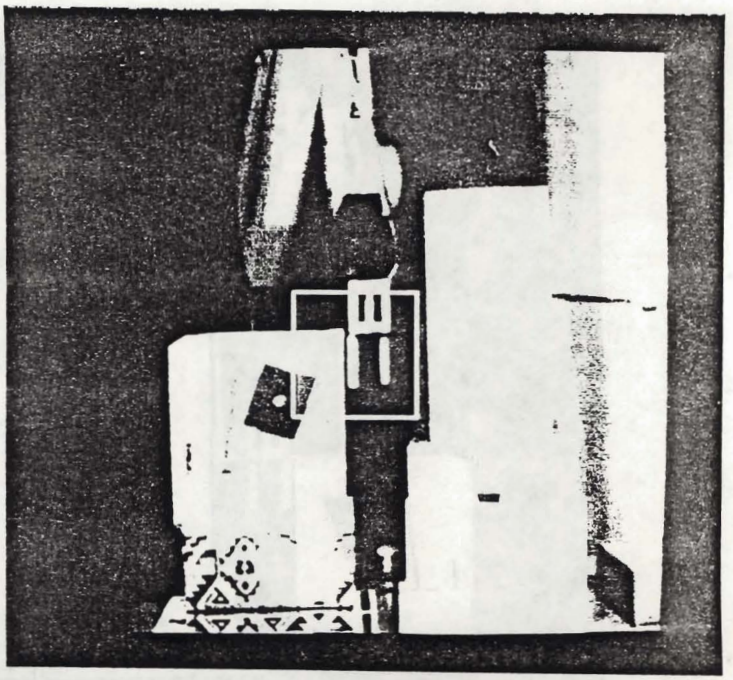


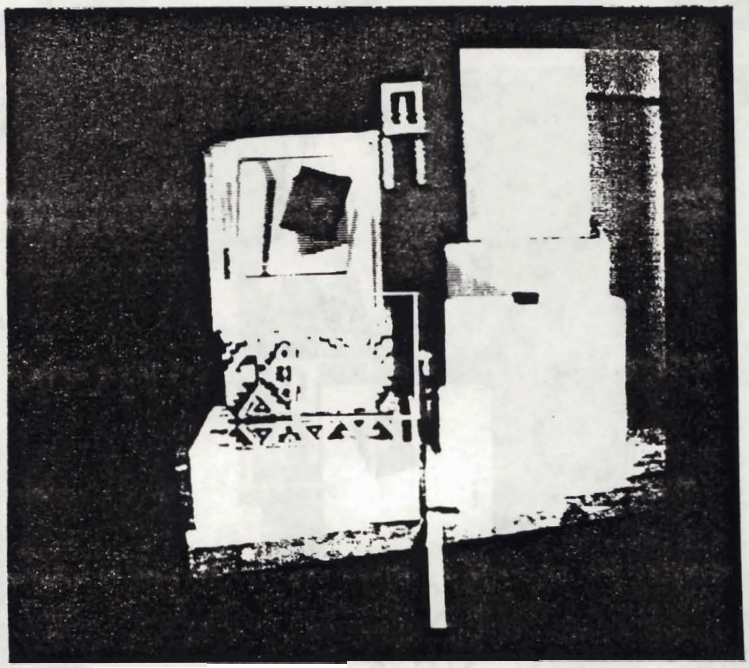
Figure 3 (continued). Digitized Images of the Test Scene.

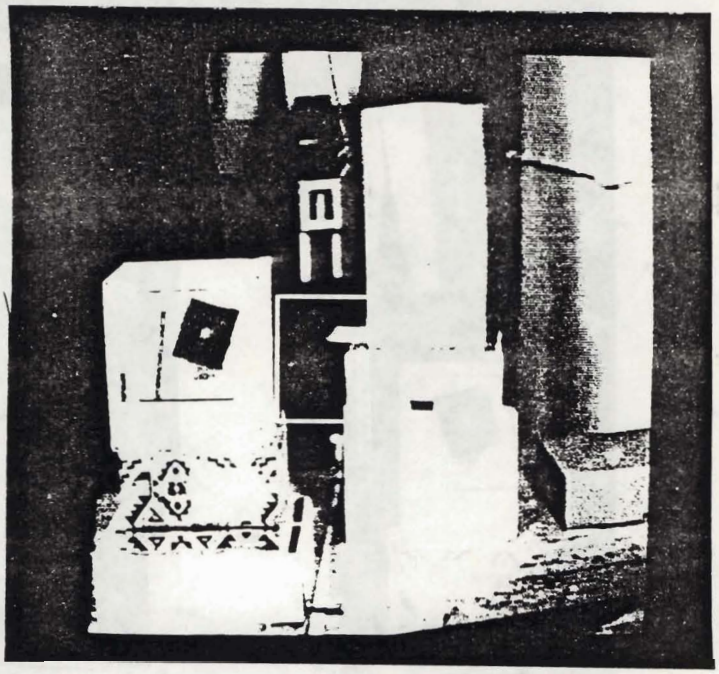
POSE 5:



POSE 6:







POSE 7:

POSE 8:

Figure 4. Top View of Points Referred to the Common Frame.

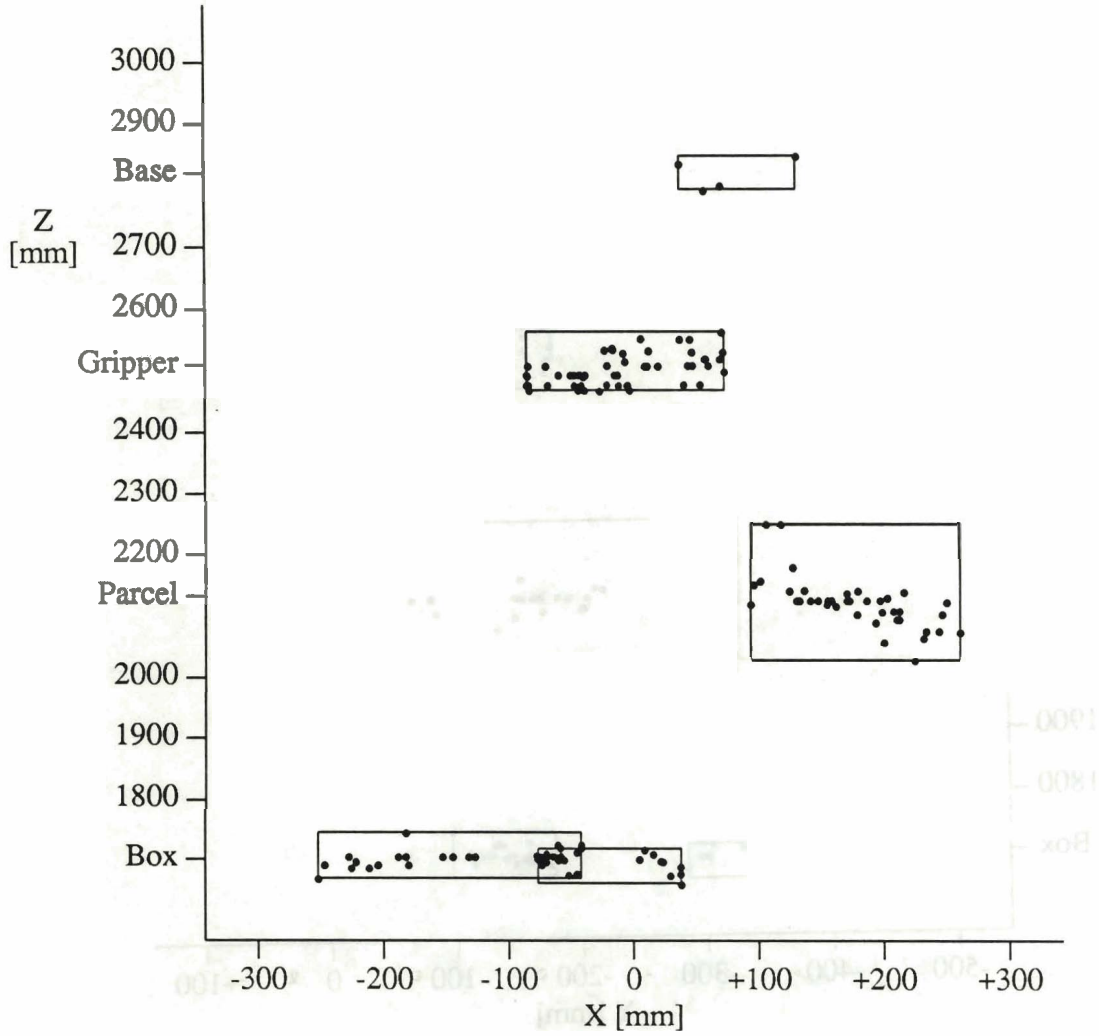


Figure 5. Side View of Points Referred to the Common Frame.

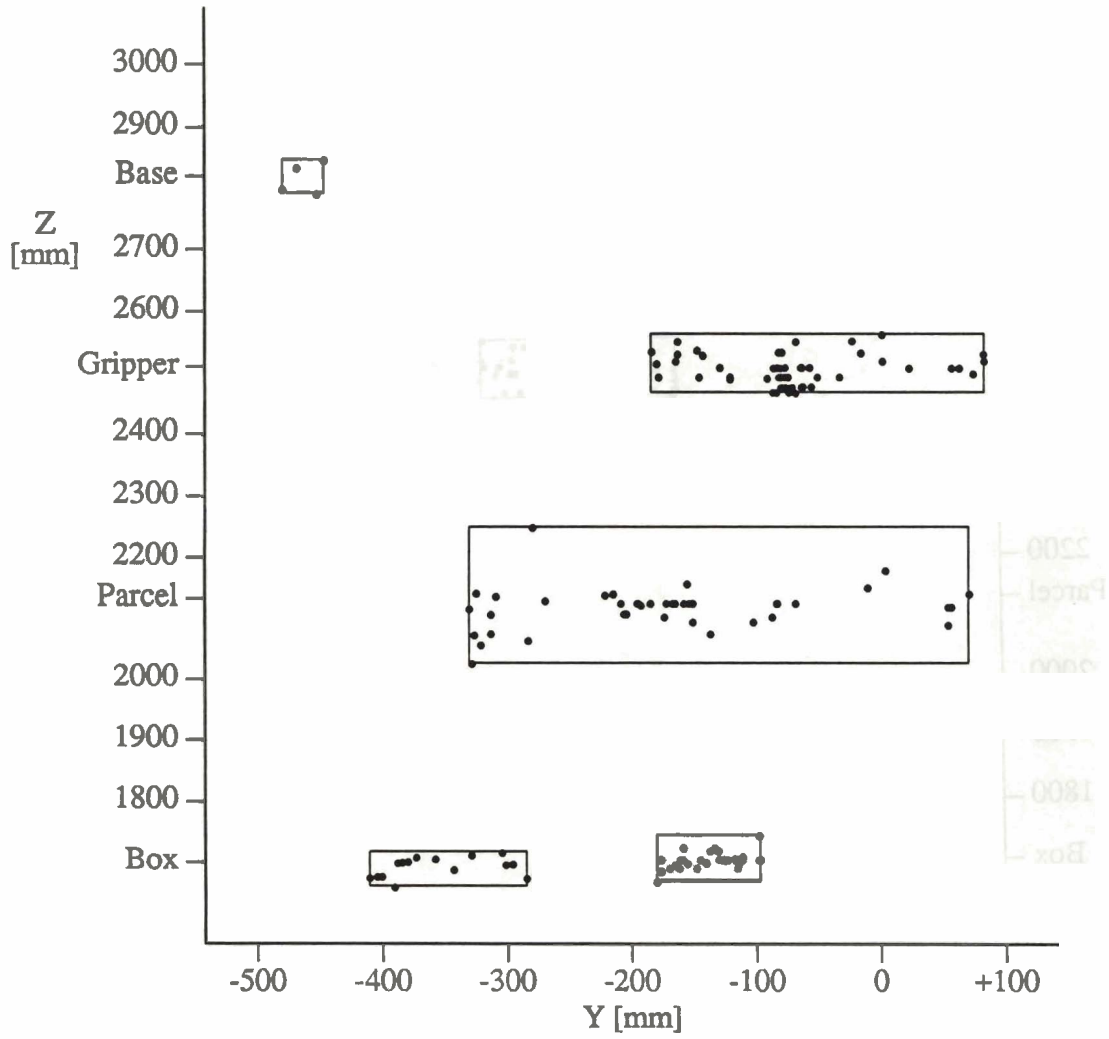


Figure 6. Measurement Uncertainties, Top View.

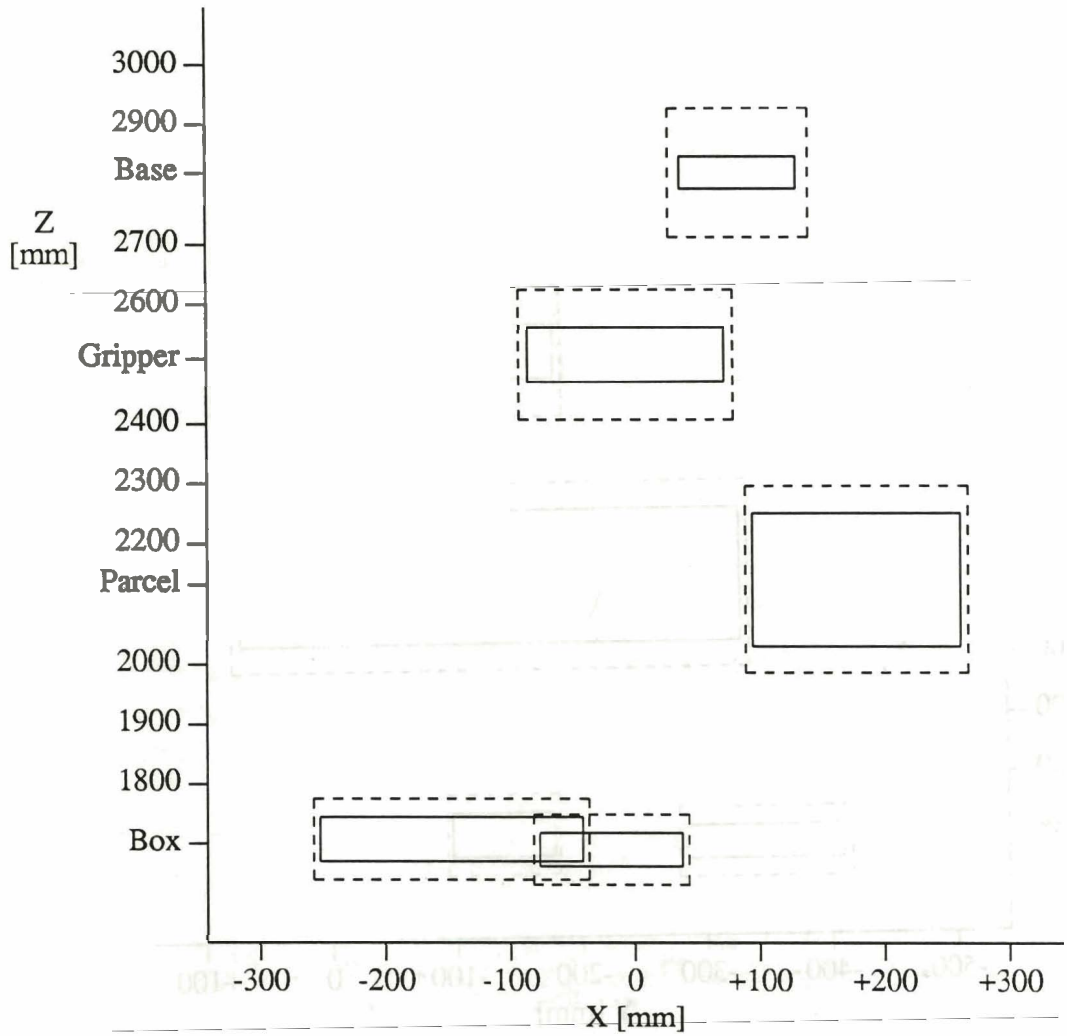


Figure 7. Measurement Uncertainties, Side View.