

On the Use of Error Propagation for Statistical Validation of Computer Vision Software

Xufei Liu, *Member, IEEE*, Tapas Kanungo, *Senior Member, IEEE*, and
Robert M. Haralick, *Fellow, IEEE*

Abstract—Computer vision software is complex, involving many tens of thousands of lines of code. Coding mistakes are not uncommon. When the vision algorithms are run on controlled data which meet all the algorithm assumptions, the results are often statistically predictable. This renders it possible to statistically validate the computer vision software and its associated theoretical derivations. In this paper, we review the general theory for some relevant kinds of statistical testing and then illustrate this experimental methodology to validate our building parameter estimation software. This software estimates the 3D positions of buildings vertices based on the input data obtained from multi-image photogrammetric resection calculations and 3D geometric information relating some of the points, lines and planes of the buildings to each other.

Index Terms—Statistical analysis, multivariate hypothesis testing, 3D parameter estimation, error propagation, software validation, software engineering.

1 INTRODUCTION

NUMEROUS computer vision problems can be posed as neither parameter estimation problems (for example, estimate the pose of the object) or hypothesis testing problems (for example, which of the N objects in a database occurs on a given image). Since the input data (such as images or feature points) to the vision algorithms are noisy, the estimates produced by the algorithms are noisy. In other words, there is an inherent uncertainty associated with the results produced by any computer vision algorithm. These uncertainties are best expressed in terms of statistical distributions and the distributions' means and covariances.

Implementations of computer vision algorithms typically run into thousands of lines of code. Furthermore, the algorithms are based on many approximations and numerous mathematical calculations. Software validation in this context means an objective verification that the output produced by the software is what was intended by the algorithm designer. While software validation is an integral part of software engineering [15], we find (to the best of our knowledge) that there is almost no attention paid to this topic in computer vision. The numerical analysis community has a much better culture when it comes to scientific software validation, but numerical software is not statistical in nature. The validation methodologies dictated by software engineering principles (e.g., coverage analysis, test

cases, bounds testing, etc.) do not address the software validation issues of computer vision software.

One way to verify that the software implementation and the theoretical calculations are correct is by providing the algorithm input data with known (controlled) statistical characteristics, which is possible since the input data can be artificially generated, and then verifying that the estimated output is actually distributed as what was predicted by theoretical calculations.

In practice, since many of the estimation problems are multidimensional, testing whether the means and covariances of the empirical distribution and predicted distribution are identical is easier than testing whether or not the shapes of the two distributions are identical. In this paper, we summarize various statistical tests regarding the means and covariances for the case when the random estimates can be assumed to be multivariate Gaussian.

While error propagation has been used in computer vision by various authors, the details of covariance propagation methodology presented in this paper can be found in [17] and the reconstruction problem has been described in Liu's thesis [23]. Faugeras described the use of error propagation in his book [13] and Kanatani [19] used error propagation to compute Cramer-Rao bounds for the variances of the estimated parameters. More recently, Chowdhury and Chellappa [9] used error propagation for estimating the effect of feature estimation noise in the structure-from-motion problem, on the first and second-order statistics of the structure parameter estimates. In the literature mentioned above, researchers have used the notion of error propagation to compute the variances of certain parameters of interest. However, how the error propagation methodology can be used as a software engineering tool to validate the implementation of algorithms is not addressed.

In many cases, it is not possible to use the standard error propagation paradigm for computing the output parameter

- X. Liu is with Cisco Systems, 600 Lanidex Plaza, Parsippany, NJ 07054. E-mail: xuliu@cisco.com.
- T. Kanungo is with the IBM Almaden Research Center, 650 Harry Road, San Jose, CA 95120. E-mail: kanungo@us.ibm.com.
- R.M. Haralick is with the Computer Science Department, Graduate Center, City University of New York, 365 Fifth Avenue, New York, NY 10016. E-mail: haralick@netscape.net.

Manuscript received 19 Dec. 2003; revised 11 Jan. 2005; accepted 31 Jan. 2005; published online 11 Aug. 2005.

Recommended for acceptance by T. Tan.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-0438-1203.

distribution. In such cases, Cho et al. [8] propose the use of bootstrap estimate (Efron and Tibshirani [12]) for the variance. Courtney and Thacker [10] have advocated software testing in vision systems and have used CAD-based 3D scene models [11] and bootstrapping for benchmarking the algorithms. More recently, Bromiley et al. [5] have used covariance propagation for coregistration in the medical imaging domain.

The issue of testing software that deals with statistical issues has been of major concern in the field of statistics itself for many years (see Wilkinson [30]). McCulloch [27], [28] proposed various tests for estimation, random number generation, and statistical distributions and then applied them on various software packages like SAS, SPSS, and S-Plus. Altman and McDonald [1] used the methods proposed by McCulloch to show that the statistical computations used by political scientists can vary dramatically depending on the software package used for the computations. Chen et al. [7] used similar methods to show that the statistical packages available on the Web can give dramatically different results. Underlying statistical computations are numerical computations. For a discussion of numerical accuracy of computations, see Gill et al. [16].

The paper is organized as follows: In Section 2, we describe our statistical methodology for validating computer vision software. In Section 3, we provide a background material on statistical hypothesis testing. In Section 4, we outline a 3D parameter estimation problem for the RADIUS computer vision project [26]. Theoretical error propagation for estimating the distributions of the 3D parameters is described in Section 5. The experimental details and the test statistics used for computing the empirical distributions are described in Section 6. The validation results comparing theoretical distributions with empirical distributions are described in Section 7.

2 A STATISTICAL METHODOLOGY FOR VALIDATING COMPUTER VISION SOFTWARE

To validate computer vision software, two checks have to be performed. The first check is that the theory is correct: The theoretically derived null distributions of the test statistics are actually correct. The second check is that the software is correct: The implementation is exactly what the theory dictates. Both the checks can be performed by computing the empirical distributions of certain observable variables and comparing them with the corresponding theoretically derived distributions.

The sequence steps required to perform the software validation therefore are:

1. Model the specific computer vision problem as a continuous¹ parameter estimation problem.
2. Derive the optimal solution in a closed form.
3. Assume a distribution for the input measurement noise.
4. Theoretically (using error propagation) derive the distribution of the parameters to be estimated.

1. See the discussion section regarding what to do when we do not have a continuous model of a specific module.

5. Generate a sample of (multivariate) input data vectors by Monte Carlo simulations using the distribution parameters used in the theoretical computations. For each sample of noisy input data, find the optimal (multidimensional) solution to the vision problem.
6. Compute the empirical distribution (histogram) of the computed optimal solutions.
7. Test whether the theoretically derived distribution and the empirically distributions are the same.

The Kolmogorov-Smirnov test can be used to check if the empirical distribution and the theoretically derived distributions are same. In fact, the Kolmogorov-Smirnov (K-S) test can be used to 1) test whether an empirical distribution is identical to a theoretical one or 2) test whether two empirical distributions are identical. The K-S test uses the fact that the maximum absolute difference between the two cumulative distributions has a known (null) distribution. For a more detailed discussion on the K-S test, see [29].

Once the means and the covariance of the output parameters have been derived, there are five tests that can be conducted between the empirical and theoretical derived distributions. In the next section, we describe the five tests.

3 STATISTICAL HYPOTHESIS TESTING PRELIMINARIES

In this section, we briefly describe the five test statistics that we use in later sections. We describe the five statistical tests, the corresponding test statistics, and null distributions. For a more detailed treatment of the results presented in this section, see [6], [3], [22], [2].

3.1 The Kinds of Statistical Hypotheses

Let x_1, x_2, \dots, x_n be a sample from a multivariate Gaussian distribution with population mean μ and population covariance Σ . That is, $x_i \in R^p$ and $x_i \sim N(\mu, \Sigma)$, where p is the dimension of the vectors x_i . In this case, we can have five possible hypotheses. Now, we describe each of the five tests, the corresponding test statistics, and the null distributions. For details, see Anderson [2] and Kanungo and Haralick [20]. We use the following definitions of \bar{x} and S .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

and

$$S = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^t,$$

where we have assumed that the data vectors x_i are p -dimensional and the sample size is n .

Test 1: $\mu = \mu_0$ with known $\Sigma = \Sigma_1$

Test statistic:

$$T = n(\bar{x} - \mu_0)^t \Sigma_1^{-1} (\bar{x} - \mu_0). \quad (1)$$

Distribution under null hypothesis is Chi-squared:

$$T \sim \chi_p^2.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution under the alternate hypothesis is noncentral Chi-squared:

$$T \sim \chi_{p,d}^2,$$

where $d = n(\mu - \mu_0)^t \Sigma_1^{-1} (\mu - \mu_0)$ is the noncentrality parameter.

Test 2: $\mu = \mu_0$ with unknown Σ
Hotelling's Test statistic:

$$T = \frac{n(n-p)}{p(n-1)} (\bar{x} - \mu_0)^t S^{-1} (\bar{x} - \mu_0). \quad (2)$$

Under the null hypothesis, the test statistic T is distributed as an F distribution with p and $n-p$ degrees of freedom:

$$T \sim F_{p,n-p}.$$

The alternate hypothesis is $H_A : \mu \neq \mu_0$; the distribution of the test statistic under the alternate hypothesis is noncentral F with p and $n-p$ degrees of freedom:

$$T \sim F_{p,n-p,d},$$

where $d = n(\mu - \mu_0)^t \Sigma^{-1} (\mu - \mu_0)$ is the noncentrality parameter.

Test 3: $\Sigma = \Sigma_0$ with known $\mu = \mu_1$
Let

$$C = \sum_{i=1}^n (x_i - \mu_1)(x_i - \mu_1)^t = (n-1)S + (\bar{x} - \mu_1)(\bar{x} - \mu_1)^t$$

and

$$\lambda = (e/n)^{pn/2} |C \Sigma_0^{-1}|^{n/2} \exp(-tr(C \Sigma_0^{-1})/2).$$

Test statistic:

$$T = -2 \log \lambda. \quad (3)$$

Distribution under null hypothesis is Chi-squared:

$$T \sim \chi_{p(p+1)/2}^2.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution under the alternate hypothesis is unknown.

Test 4: $\Sigma = \Sigma_0$ with unknown μ
Let $B = (n-1)S$, and

$$\lambda = (e/(n-1))^{p(n-1)/2} |B \Sigma_0^{-1}|^{(n-1)/2} \exp(-tr(B \Sigma_0^{-1})/2).$$

Test statistic:

$$T = -2 \log \lambda. \quad (4)$$

Distribution under null hypothesis is Chi-squared:

$$T \sim \chi_{p(p+1)/2}^2.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$; the distribution under the alternate hypothesis is unknown.

Test 5: $\Sigma = \Sigma_0$ and $\mu = \mu_0$
Define $B = (n-1)S$ and

$$\lambda = (e/n)^{pn/2} |B \Sigma_0^{-1}|^{n/2} \exp(-[tr(B \Sigma_0^{-1}) + n(\bar{x} - \mu_0)^t \Sigma_0^{-1} (\bar{x} - \mu_0)]/2).$$

Test statistic:

$$T = -2 \log \lambda. \quad (5)$$

Distribution under true null hypothesis is Chi-squared:

$$T \sim \chi_{p(p+1)/2+p}^2.$$

The alternate hypothesis is $H_A : \Sigma \neq \Sigma_0$, and $\mu \neq \mu_0$; the distribution under the alternate hypothesis is unknown.

4 APPLICATION: 3D PARAMETER ESTIMATION

We applied the hypothesis testing methodology to validate the 3D parameter estimation software used for constructing the ground truth model from the RADIUS model board data set. In this section, we describe the problem and the optimization approach.

4.1 Site Model Construction

The task is to construct 3D object models from the detected 2D image features and the known geometric constraints of the observed perspective projections of the 3D objects. The data set consists of the 78 images from the two RADIUS model boards and the 3D coordinates of some building vertices. Since the purpose of this site model construction was to establish ground truth for automatic site model construction algorithms, the corresponding points of the building vertices that were observable on the images were identified and located manually. Also, 3D positions of a few of the building vertices are known. A simultaneous estimation of the interior parameters and exterior orientation parameters of the cameras was done by setting up and solving a very large photogrammetric resection problem. Then, using these camera parameters, a multi-image triangulation was performed. This yielded the noisy estimates for the building vertices that was the input to the site model construction software whose testing we describe.

The geometric constraint procedure takes the photogrammetrically estimated 3D point positions and their covariance matrices as observations. It uses the partial models of the buildings to generate constraints on the building parameters. To estimate the optimal 3D parameters that satisfy the relations in the partial models, a constrained optimization model is solved. By error propagation, we derive the covariance matrix of the estimated building vertices which are guaranteed to satisfy the given constraints. This is discussed in our paper [25].

4.2 Constrained Optimization

Partial building models are used to constrain the 3D parameters to be estimated. For polyhedron buildings, partial models represent geometric constraints between building entities: lines, planes, and points (vertices). Each entity associates with a set of parameters.

point: A point is specified by its 3D coordinates, denoted by (x, y, z) or, in vector form, \mathbf{x} .

line: A line is defined by a direction cosines $e = (e_x, e_y, e_z)$ and a reference point $\mathbf{b} = (b_x, b_y, b_z)$ on the line. We choose the unique \mathbf{b} that satisfies $\mathbf{e} \cdot \mathbf{b} = \mathbf{e}^T \mathbf{b} = 0$.

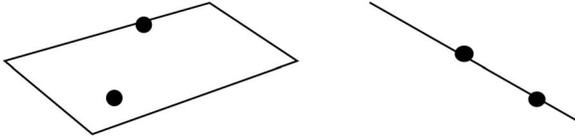


Fig. 1. Position relations in a partial model.

plane: A plane is specified by a normal vector $\mathbf{v} = (\alpha, \beta, \gamma)$ and a directed distance constant d .

The specific relations we used to constrain these parameters are: point-on-plane, point-on-line, plane-angle-plane, line-angle-line, and plane-angle-line.

Each *point-on-plane* relation associates with a planar equation:

$$x^T \mathbf{v} + d = x\alpha + y\beta + z\gamma + d = 0. \quad (6)$$

Each *point-on-line* relation associates with a line equation:

$$\mathbf{x} = \eta \mathbf{e} + \mathbf{b}, \quad \eta \in \mathcal{R}. \quad (7)$$

Fig. 1 illustrates the *point-on-plane* and *point-on-line* relations.

A *plane-angle-plane* relation is specified with the inner product of the normal vectors of two planar surfaces. A zero inner product means two perpendicular planes, while 1 or -1 means two parallel planes. Similarly, a *line-angle-line* relation is specified with the inner product of the direction cosines of two lines and a *plane-angle-line* relation is specified with the inner product of the normal vector of a plane and the direction cosine of a line. Fig. 2 illustrates these angle relations.

For a more detailed description, please see [23].

The observed 3D points and the associated covariance matrix Σ are obtained from triangulation. The perturbation model that we used for the observation was a zero mean Gaussian noise with unknown covariance. Having the partial object model and the perturbation model, we can define the estimation problem. Let $\Theta \in \mathbb{R}^m$ denote the parameters, $X' \in \mathbb{R}^m$ denote the observations, and $p(X' | \Theta)$ denote the likelihood function. In the building estimation problem, the parameters are the coordinates of the points, the normal vectors and distance constants of the planes, and the direction cosines and reference points of the lines.

Assuming that the given optimality criterion is the maximum posterior probability, a Bayesian approach can be used to transform the problem into a maximum likelihood problem with constraints. Let the constraints be denoted by $\Theta \in C_\Theta \subset \mathbb{R}^m$. The problem can be expressed as a constrained optimization problem.

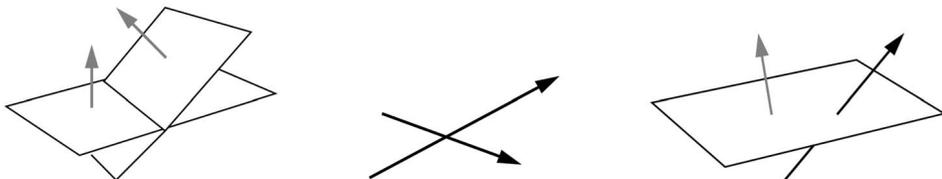


Fig. 2. Angle relations in a partial model.

$$\min\{-p(X' | \Theta) \mid \Theta \in C_\Theta\}.$$

The problem can be reformulated by taking the logarithm of the probability function. Under the assumption of Gaussian noise, we obtain a least squares model. The objective function is the weighted sum of squared errors between the estimated point positions and the observed points, which is also known as the Mahalanobis distance:

$$\min_{\Theta} \left\{ f(\Theta) := (X' - X)^T \Sigma^{-1} (X' - X) \right\} \quad (8)$$

subject to $\Theta \in C_\Theta$,

where X denotes the unknown 3D points and the feasible set C_Θ is determined by the partial model and the unit length constraint for the directional vectors.

If the noise effecting different 3D points is independent, the objective function can be rewritten as

$$f(\Theta) = \sum_{i=1}^K (x'_i - x_i)^T \Sigma_i^{-1} (x'_i - x_i),$$

where Σ_i is the covariance matrix of the i th point and K is the number of observed points.

The constraints can be incorporated into the optimization problem as follows (see [23]):

$$\min_{\Theta} \left\{ f(\Theta) := (X' - X)^T \Sigma^{-1} (X' - X) \right\} \quad (9)$$

subject to $h_i(\Theta) = 0, \quad i = 1, \dots, r$.

In the above, the equation $h_i(\Theta) = 0$ represents all the constraints derived from partial models. The numerical solution to this optimization problem can be achieved by various methods such as: the reduced gradient method [16], the sequential quadratic programming [14], or the augmented Lagrangian method [4].

5 ERROR PROPAGATION

Once the constrained optimization produces a result, we use the error propagation approach [18], [24] to transform the input error covariance matrix to the output covariance matrix. In the building estimation problem, we have the optimization model

$$\min_{\Theta} f(\Theta)$$

subject to $\mathbf{h}(\Theta) = 0$,

where f is the sum of squared errors between the estimated 3D points and the observed 3D points.

The Lagrangian function is

$$L(X', \Theta, \Lambda) = f(X', \Theta) + \Lambda^T \mathbf{h}(\Theta)$$

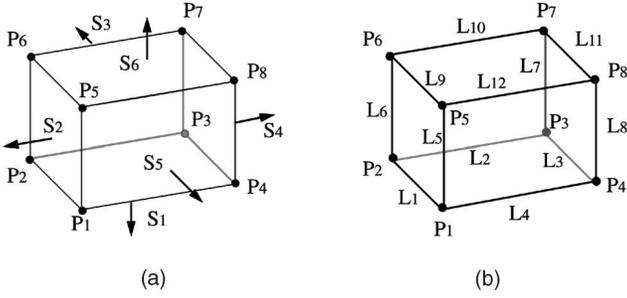


Fig. 3. Angle relations in the flat roof model: (a) point-plane relations, (b) point-line relations.

Suppose that $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ is an optimal point. From the necessary conditions of a local minimum point, the linearized model at the optimal point can be obtained by solving [14], [23]

$$\begin{pmatrix} \tilde{Q} & \tilde{H}^T \\ \tilde{H} & 0 \end{pmatrix} \begin{pmatrix} \Delta\Theta \\ \Delta\Lambda \end{pmatrix} = \begin{pmatrix} -\tilde{B}\Delta X \\ 0 \end{pmatrix}, \quad (10)$$

where

$$\begin{aligned} \tilde{Q} &= \nabla_{\theta\theta}^2 \mathcal{L}(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda}) = \nabla_{\theta\theta}^2 f(\tilde{X}, \tilde{\Theta}) + \sum_{j=1}^r \tilde{\lambda}_j \nabla^2 h_j(\tilde{\Theta}), \\ \tilde{B} &= \nabla_{\theta x}^2 \mathcal{L}(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda}) = \nabla_{\theta x}^2 f(\tilde{X}, \tilde{\Theta}), \\ \tilde{H} &= \nabla h(\tilde{\Theta}). \end{aligned}$$

The Lagrangian matrix at the point of $(\tilde{X}, \tilde{\Theta}, \tilde{\Lambda})$ can be approximated by the Lagrangian matrix at the minimum if the error is small. Hence, the linear model can be approximated by

$$\begin{pmatrix} \mathbf{Q}^* & (\mathbf{H}^*)^T \\ \mathbf{H}^* & 0 \end{pmatrix} \begin{pmatrix} \Delta\Theta \\ \Delta\Lambda \end{pmatrix} = \begin{pmatrix} -\mathbf{B}^* \Delta X \\ 0 \end{pmatrix},$$

where

$$\begin{aligned} \mathbf{Q}^* &= \nabla_{\theta\theta}^2 \mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) \\ &= \nabla^2 f(X', \hat{\Theta}) + \sum_{j=1}^r \hat{\lambda}_j \nabla^2 h_j(\hat{\Theta}), \\ \mathbf{B}^* &= \nabla_{\theta x}^2 \mathcal{L}(X', \hat{\Theta}, \hat{\Lambda}) = \nabla_{\theta x}^2 f(X', \hat{\Theta}), \\ \mathbf{H}^* &= \nabla h(\hat{\Theta}). \end{aligned}$$

Assume that the constraints are linearly independent. Then, the row vectors in matrix H^* are linearly independent. We can use the null space method to compute the error propagation matrix J [14], [23].

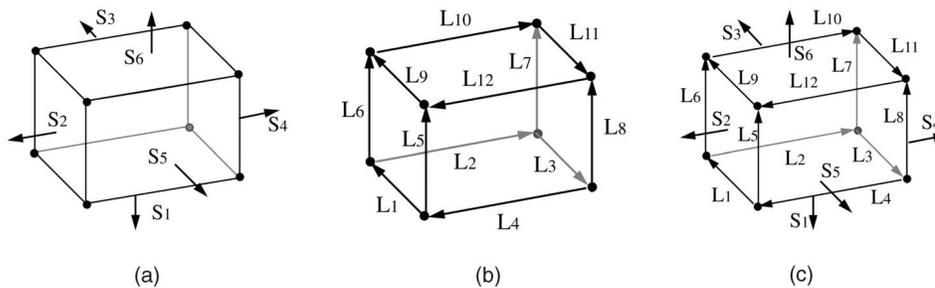


Fig. 4. Angle relations in the flat roof model: (a) plane-angle-plane, (b) line-angle-line, (c) plane-angle-line.

Once the error propagation matrix is obtained, we can propagate the covariance matrix of the observations Σ to the output. The covariance matrix of the estimated parameters, Σ_{Θ} , can be approximated by

$$\Sigma_{\Theta} = J \Sigma J^T. \quad (11)$$

6 EMPIRICAL DISTRIBUTION COMPUTATION AND TEST STATISTICS

To validate the optimization algorithm and the error propagation model, an experiment is needed. This section describes the experimental methodology for this validation.

6.1 Ideal Data Generators and Noise Model

Three building types, the cube box, the peak roof house, and the hip roof house, appear frequently in the given sites. The cube box model contains six planes, 12 lines, and eight points. Its point-on-plane relations and the point-on-line relations are illustrated in Fig. 3. Its plane-angle-plane, line-angle-line, and plane-angle-line relations are illustrated in Fig. 4. The peak roof house model assumes a slated roof on a cube. This model contains seven planes, 15 lines, and 10 points. Its position relations are illustrated in Fig. 5 and its angle relations are illustrated in Fig. 6. The hip roof house model assumes a roof with slated sides and ends. It contains nine planes, 17 lines, and 10 points. Its point-on-plane and point-on-line relations are shown in Fig. 7. Its angle relations are shown in Fig. 8. For a more detailed description of the models and corresponding constraints, see [23]. These three models are chosen as our prototype models with unknown length, location, and orientation parameters.

In the experiment, ideal data generators randomly generate the ideal parameters for the prototype models and produce the ideal 3D points.

Assume that a 3D coordinate system x - y - z is used. To simulate the site model situation, the ground is assigned as the plane $z = 0$. Without losing generality, we assume that the ideal model parameters determining the 3D positions of the building vertices have uniform distributions. The center of the bottom plane of a basic model is in a region defined by $[-x_0, x_0], [-y_0, y_0], [-z_0, z_0]$. The basic model is rotated on the ground with a random angle $\phi \in [\phi_0, \phi_1]$.

The three length parameters for the cube box model are denoted by a, b, c , with $a_0 \leq a < a_1$, $b_0 \leq b < b_1$, and $c_0 \leq c < c_1$.

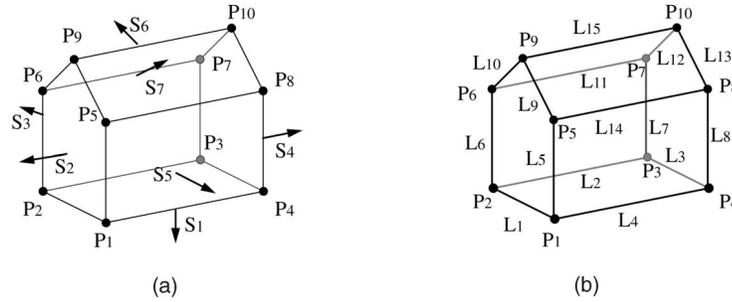


Fig. 5. Position relations in the peaked roof model: (a) point-plane relations, (b) point-line relations.

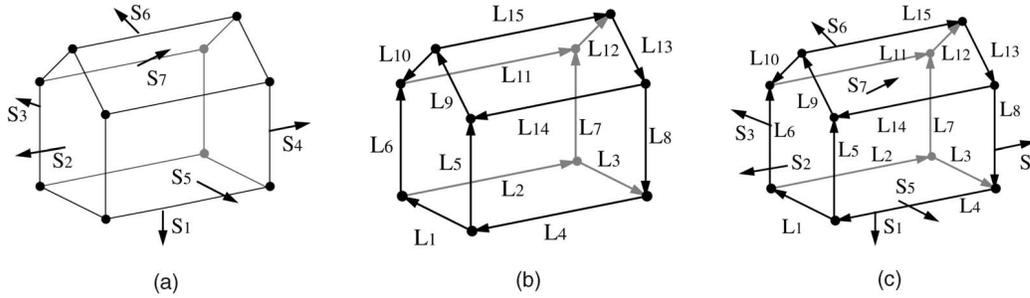


Fig. 6. Angle relations in the peaked roof model: (a) plane-angle-plane, (b) line-angle-line, (c) plane-angle-line.

In this experiment, the ranges of the parameters for the cube model are set as follows:

x_0	y_0	z_0	ϕ_0	ϕ_1	a_0	a_1	b_0	b_1	c_0	c_1
50	50	0	0	2π	30	60	30	60	30	60

The peak roof model uses four length parameters, a, b, c, d , where a, b, c are same as those in cube model and their ranges are the same. The height of the peak roof has parameter d , with $10 \leq d \leq 20$. The hip roof model requires one more parameter e than the peak roof model. The length of the roof edge is $a - 2e$, with $5 \leq e \leq 10$.

For each building type, K ideal buildings are randomly generated. Each one of these K buildings will be associated with n experiments, where Gaussian random noise is added to each of the 3D coordinates of the building and the constrained optimization is used to estimate the building vertices that satisfy the various geometric constraints. As a result of these n experiments, n estimates of the building

parameters are produced. It is these n estimates on which the hypothesis test statistics will be computed. We call the procedure to determine these n test statistics a trial. Since there are K ideal buildings for each ideal building type, we can compute K test statistics. These K statistics can then be used to test the hypothesis that their distribution is as the statistical theory of the test says it should be.

The noise values are independently sampled from a Gaussian distribution $\mathcal{N}(0, \sigma^2 I)$, where σ is the standard deviations of the random variables $\delta x, \delta y, \delta z$. We repeated each experiment with σ being set to 1.0, 2.0, or 3.0. The validation results for all three different standard deviations is similar. So, here, we just discuss the validation for the standard deviation being equal to 3.0. K is set at 100. n is set at 500 for the cube model and 700 for other models.

6.2 Statistical Test

At each trial, a sample of model parameters and corresponding ideal 3D points is produced by the ideal data generator. Let the ideal parameters be denoted by $\hat{\Theta}$. For each ideal building instance having parameters $\hat{\Theta}$, n independent perturbations $\{\Delta X_i, i = 1, \dots, n\}$ are generated from the noise model with distribution $\mathcal{N}(0, \Sigma)$. By adding the perturbations to the ideal points, the perturbed data set $\{X'_1, X'_2, \dots, X'_n\}$ is generated. For each of the perturbed data $\{X'_1, X'_2, \dots, X'_n\}$, an optimal solution $\hat{\Theta}_i$ is computed by solving

$$\begin{aligned} \min_{\Theta_i} \quad & f(X'_i, \Theta_i) \quad i = 1, \dots, n \\ \text{subject to} \quad & \mathbf{h}(\Theta_i) = 0. \end{aligned}$$

Thus, we have n estimates $\{\hat{\Theta}_i, i = 1, \dots, n\}$.

Using (11), we can transform the input covariance matrix through the error propagation matrix to the output. If the linear model is valid, the estimated

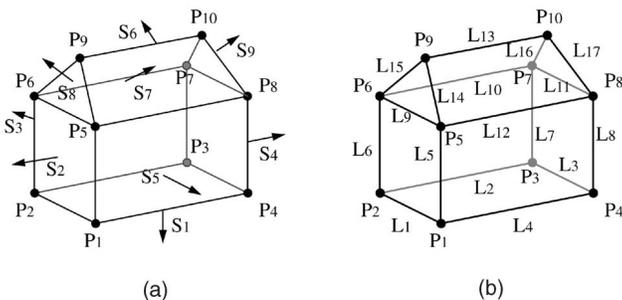


Fig. 7. Position relations in the hip roof model: (a) point-plane relations, (b) point-line relations.

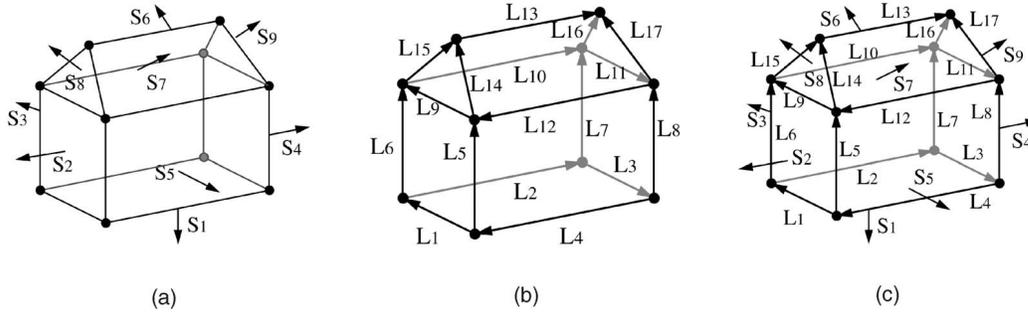


Fig. 8. Angle relations in the hip roof model: (a) plane-angle-plane, (b) line-angle-line, (c) plane-angle-line.

parameters $\{\hat{\Theta}_i, i = 1, \dots, n\}$ should be approximately distributed as $\mathcal{N}(\hat{\Theta}, J\Sigma J^T)$.

Let $\Delta\hat{\Theta}_i$ denote $\hat{\Theta}_i - \Theta$, $i = 1, \dots, n$. Let $\mu_0 = 0$ and $\Sigma_0 = J\Sigma J^T$. Under the linearized model, $\{\Delta\hat{\Theta}_i, i = 1, \dots, n\}$ have distribution $\mathcal{N}(\mu_0, \Sigma_0)$. Considering $\{\Delta\hat{\Theta}_i, i = 1, \dots, n\}$ as a random sample from a Gaussian distribution $\mathcal{N}(\mu, \Sigma)$, we can perform any one of the five hypothesis tests. Here, we just discuss our results on hypothesis $H_5: \mu = \mu_0$ and $\Sigma = \Sigma_0$. Results on the other hypothesis tests are similar.

The significance level α is selected to be 0.05. Under the null hypothesis, the computed statistics of the mean and covariance tests have the null distributions. This can be verified by using the Kolmogorov-Smirnov test (K-S test) on the K test statistics generated from the K trials.

6.3 Range Space Analysis

The standard hypothesis test methods require that the covariance matrix be positive definite. However, the output covariance of a constrained optimization is generally semipositive definite, precisely because of the constraints.

Theorem 1. *Suppose that not all of the derivatives of the constraint equations are equal to zero at the local minimum point, then the propagated error covariance $J\Sigma J^T$ is singular.*

The proof of the above theorem is provided in the Appendix.

To utilize the standard hypothesis technology, we project a semipositive definite matrix onto its range space. Suppose that an $n \times n$ covariance matrix Σ_0 has k nonzero eigenvalues w_1, \dots, w_k and the associated unit eigenvectors v_1, \dots, v_k . A basis of the range space of Σ_0 can be composed by

$$B = (v_1, \dots, v_n).$$

Use B to perform the matrix transform as follows:

$$B^T \Sigma_0 B = \Sigma_B = \begin{pmatrix} w_1 & 0 & \cdot & \cdot & \cdot \\ 0 & w_2 & 0 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & w_k \end{pmatrix}.$$

Let B^\perp be a basis matrix of the null space of matrix Σ_0 . It is obvious that (B, B^\perp) is orthonormal. In the experiment, we check whether $(B^\perp)^T \Delta\hat{\Theta}$ has very small variances (caused by round off errors and nonlinear items). If it is

true, we conduct the hypothesis test on variables $B^T \Delta\hat{\Theta}$ with covariance matrix Σ_B .

Due to the round-off error and nonlinear items, the zero eigenvalues of matrix Σ_0 may not be exactly zero. We use a small threshold to distinguish the zero eigenvalues and the nonzero eigenvalues. In all our experiments, the threshold is set to 10^{-6} times the maximum eigenvalue.

For the cube model, the range space of the output error covariance matrix has seven dimensions. This can be understood as follows: Consider a cube house whose faces are all at right angles to each other. Count the number of degrees of freedom. The size of a cube model is defined by three independent parameters. The location of the model is specified by three translation parameters in 3D space. In the experiments, the normal vector of the cube roof is fixed to the vertical direction. The only possible rotation is around the vertical axis of the model. Thus, the total number of independent parameters is seven. For the peak roof model, the above analysis is similar, except that two more parameters are needed to determine the roof height and the ridge position. (In the partial model, we do not fix the horizontal position of the roof ridge to the center of the building.) Thus, the range space of the output covariance matrix for the peak roof model has nine dimensions. The hip roof model inherits all the parameters of the peak roof model. It requires two more parameters to determine how much of the ridge is being cut from each of the two ends (they are assumed to be independent). These parameters can be thought of as the relative starting and ending points for the ridge. Thus, the range space of the output covariance matrix for the hip roof model has 11 dimensions.

7 VALIDATION: COMPARING THEORY WITH EXPERIMENTS

In this section, we show the experimental results on the cube model, the peak roof model, and the hip roof model. An example of the model estimation result is shown in Fig. 12.

7.1 Test of Cube Model with $\sigma = 3.0$

The theoretical and empirical null distribution of the five test statistics for the cube model are shown in Fig. 9. The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

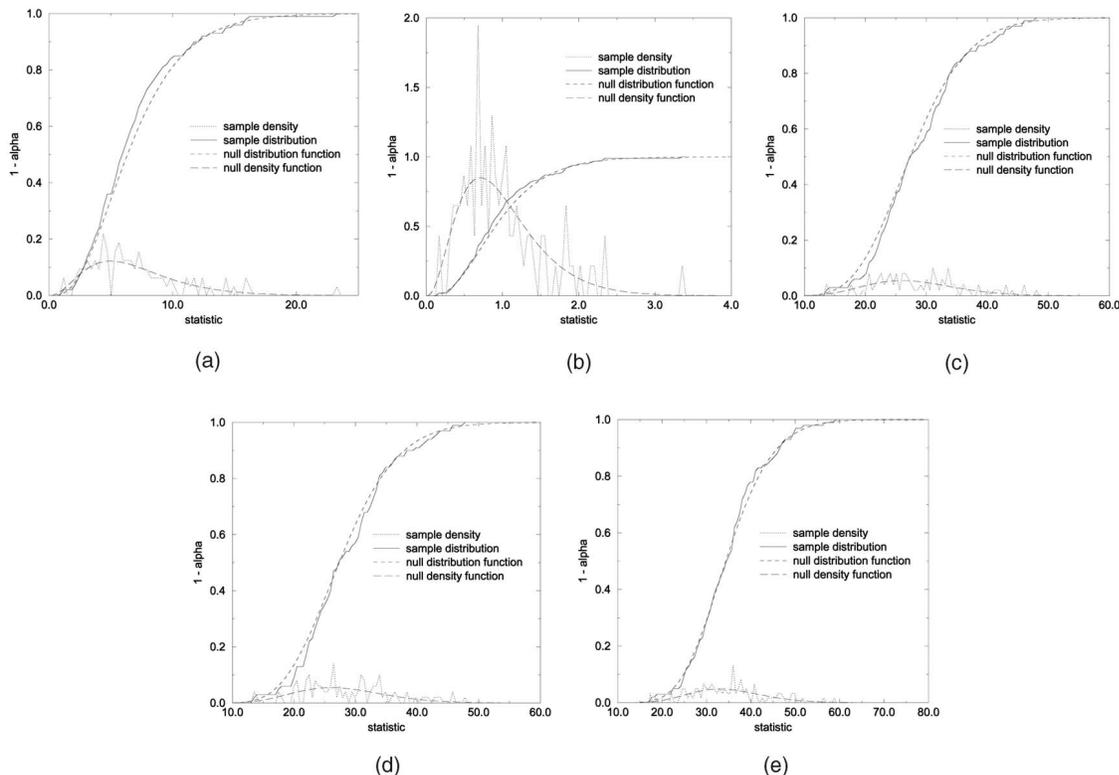


Fig. 9. Null distribution of the five test statistics for the cube model at $\sigma = 3$. (a) Test mean with known variance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.

TABLE 1
Cube Model: (a) Test Statistic Values for the Five Tests, (b) Kolmogorov-Smirnov Distribution Test

test method	number of trials	sample size	significance level	reject rate	test method	null distribution	size n	p-value	K-S test
test μ , known Σ	100	500	0.05	0.06	test μ , known Σ	χ_p^2	100	0.391	pass
test μ , unknown Σ	100	500	0.05	0.05	test μ , unk. Σ	$F_{p,n-p}$	100	0.217	pass
test Σ , known μ	100	500	0.05	0.07	test Σ , known μ	$\chi_{p(p+1)/2}^2$	100	0.866	pass
test Σ , unknown μ	100	500	0.05	0.06	test Σ , unk. μ	$\chi_{p(p+1)/2}^2$	100	0.881	pass
test μ and Σ	100	500	0.05	0.05	test μ and Σ	$\chi_{p(p+1)/2+p}^2$	100	0.700	pass

(a)

(b)

The experimental trials were run multiple times and, each time, the null hypothesis was either rejected or not rejected. The results are summarized in Table 1a. The null hypothesis is not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical distributions are similar. Results are shown in Table 1b. In the K-S test, the value of the degrees of freedom, p , is 7. None of the five test statistics distributions fail the K-S test at a significance level of 0.05. Thus, the optimization model and the error propagation model for the cube model are validated.

7.2 Test of Peak Roof Model with $\sigma = 3.0$

The theoretical and empirical null distribution of the five test statistics for the peak roof model are shown in Fig. 10.

The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

The experimental trials were run multiple times and, each, time the null hypothesis was either rejected or not rejected. The results are summarized in Table 2a. The null hypothesis is not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical null distributions are similar. Results are shown in Table 2b. In the K-S test, the value of the degrees of freedom, p , is 9. None of the five test statistics distributions fail the K-S test at a significance level of 0.05. Thus, the optimization model and the error propagation model for the peak roof model are validated.

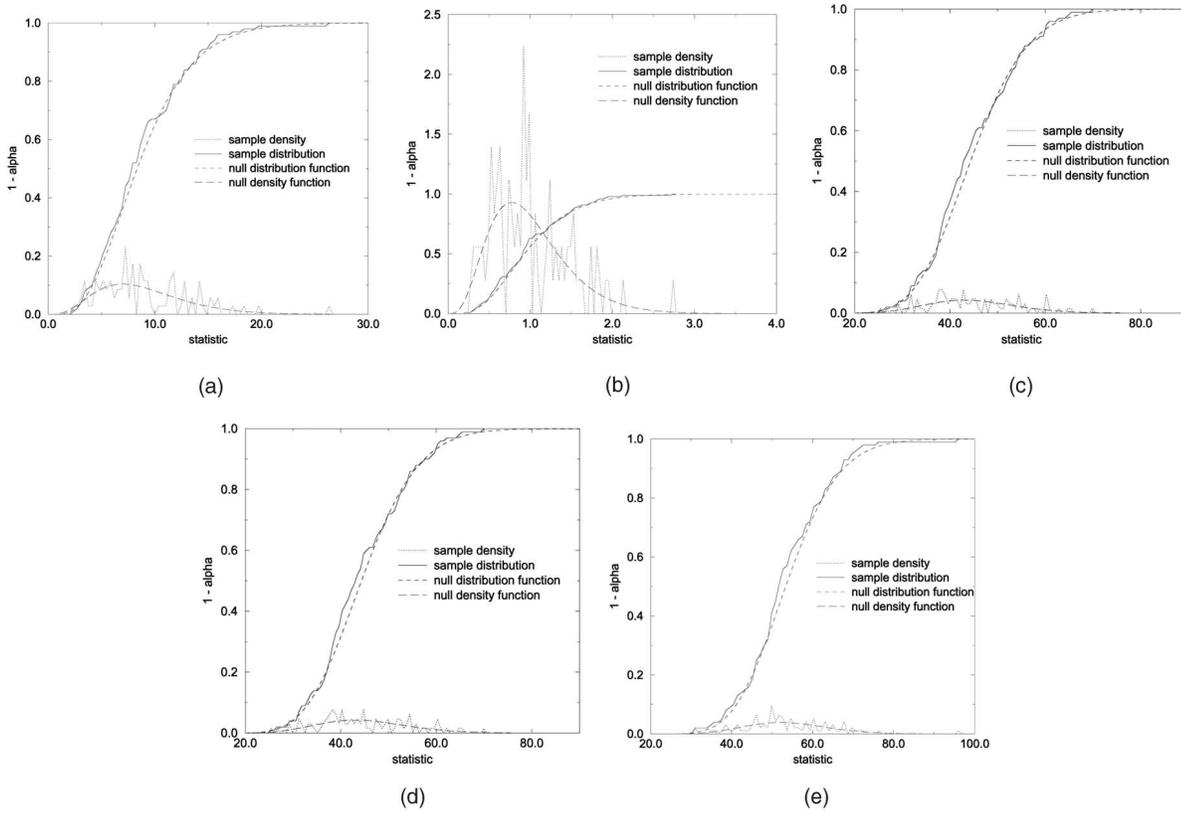


Fig. 10. Null distributions of the five test statistics for the peak roof model at $\sigma = 3$. (a) Test mean with known variance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.

TABLE 2
Peak Roof Model Results: (a) Test Statistic Values for the Five Tests, (b) Kolmogorov-Smirnov Distribution Test

test method	number of trials	sample size	significance level	reject rate	test method	null distribution	size n	p-value	K-S test
test μ , known Σ	100	700	0.05	0.03	test μ , known Σ	χ_p^2	100	0.464509	pass
test μ , unk. Σ	100	700	0.05	0.03	test μ , unknown Σ	$F_{p,n-p}$	100	0.277997	pass
test Σ , known μ	100	700	0.05	0.04	test Σ , known μ	$\chi_{p(p+1)/2}^2$	100	0.450701	pass
test Σ , unk. μ	100	700	0.05	0.04	test Σ , unknown μ	$\chi_{p(p+1)/2}^2$	100	0.402053	pass
test μ and Σ	100	700	0.05	0.02	test μ and Σ	$\chi_{p(p+1)/2+p}^2$	100	0.205762	pass

7.3 Test of Hip Roof Model with $\sigma = 3.0$

The theoretical and empirical null distribution of the five test statistics for the hip roof model are shown in Fig. 11. The standard deviation used was $\sigma = 3.0$. The x axis is the statistic used in the test and the y axis represents $1 - \alpha$, where α is the significance level.

The experimental trials were run multiple times and, each time, the null hypothesis was either rejected or not rejected. The results are summarized in Table 3a. The null hypothesis is not rejected at a 0.05 significance level.

The K-S test was used to test whether the empirical and theoretical null distributions are similar. Results are shown in Table 3b. In the K-S test, the value of the degrees of freedom, p , is 11. None of the five test statistics distributions fail the K-S test at a significance level of 0.05. Thus, the

optimization model and the error propagation model for the hip roof model are validated.

8 DISCUSSION

We have described a methodology for testing software systems that compute results that are statistical in nature. The approach proposed is to theoretically derive the statistical properties of the result and then compare it with empirical distributions. There are several issues that need to be addressed.

Since the error propagation methodology depends on linearization of functions, if the input noise perturbations are “large” with respect to the function behavior at the point where the function is being evaluated, the output distribution

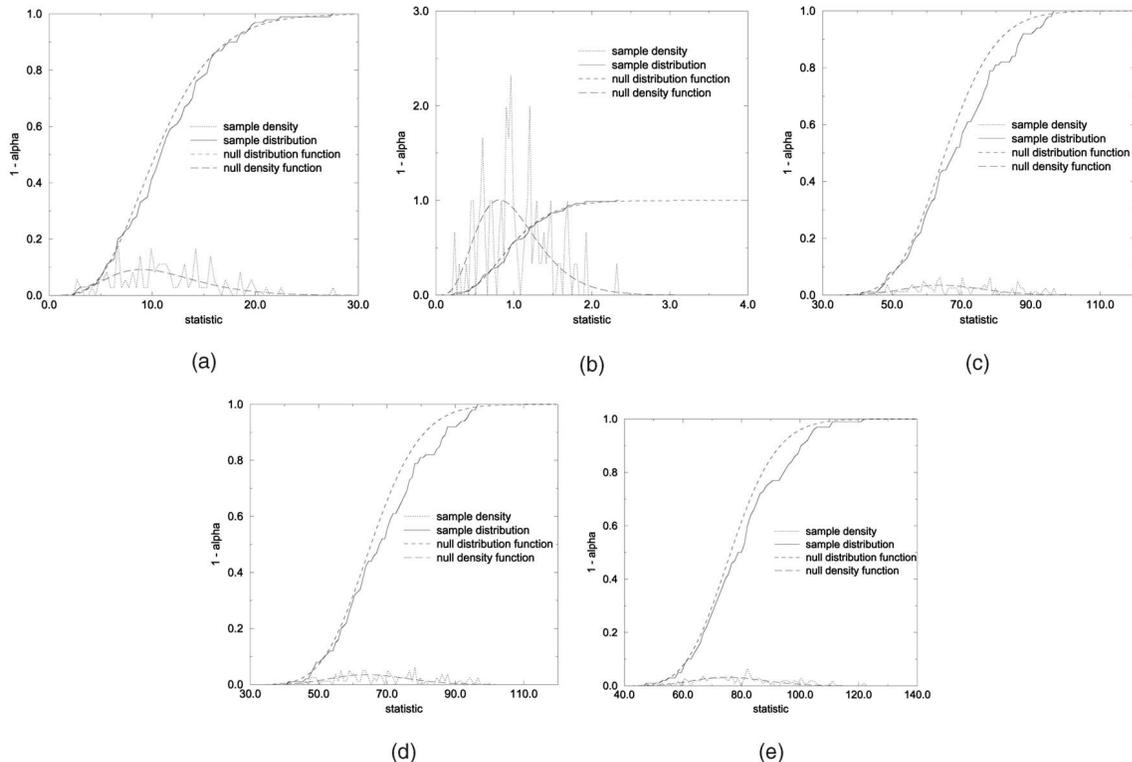


Fig. 11. Null distributions of the five test statistics for the hip roof model at $\sigma = 3$. (a) Test mean with known covariance. (b) Test mean with unknown covariance. (c) Test covariance with known mean. (d) Test covariance with unknown mean. (e) Test mean and covariance.

TABLE 3
Hip Roof Model: (a) Test Statistic Values for the Five Tests, (b) Kolmogorov-Smirnov Distribution Test

test method	number of trials	sample size	significance level	reject rate
test μ , known Σ	100	700	0.05	0.03
test μ , unknown Σ	100	700	0.05	0.03
test Σ , known μ	100	700	0.05	0.11
test Σ , unknown μ	100	700	0.05	0.11
test μ and Σ	100	700	0.05	0.13

(a)

test method	null distribution	size n	p-value	K-S test
test μ , known Σ	χ_p^2	100	0.771912	pass
test μ , unknown Σ	$F_{p,n-p}$	100	0.588617	pass
test Σ , known μ	$\chi_{p(p+1)/2}^2$	100	0.110815	pass
test Σ , unknown μ	$\chi_{p(p+1)/2}^2$	100	0.103598	pass
test μ and Σ	$\chi_{p(p+1)/2+p}^2$	100	0.096341	pass

(b)

will not be as predicted. This implies that statistical tests of the output distribution will fail. Thus, if the software fails the statistical test, the designer needs to check whether the input noise variances are within reasonable limits.

Another interesting issue in statistical hypothesis testing is that if a test does not fail, one can only say that “we do not have enough statistical evidence to reject the null hypothesis.” That is, even if a test is not rejected, the software implementation could be still faulty—our test just did not detect it. However, if we have two different implementations and if the first test has a higher p-value than the second test, one can say that the first test is a better implementation than the second.

In some cases, there might be a submodule for which we do not have a continuous mathematical characterization.

Thus, the standard error propagation is not possible. However, one can empirically characterize the behavior of such systems and still be able to compute the output distributions of the entire system. In [21], a methodology from the psychophysics literature was adopted to empirically characterize two computer vision algorithms.

9 SUMMARY

We described a statistical methodology to validate the theoretical derivations and software that make up a large system for estimating 3D positions of buildings vertices based on the input data obtained from multi-image photogrammetric resection calculations. Error propagation allowed us to derive the null-distributions of various test statistics of measurable quantities. These theoretically



Fig. 12. This figure shows the estimated houses overlaid on top of the original model-board image.

derived null-distributions allowed us to validate whether the measurements in the implemented system followed the theoretically derived distributions. The Kolmogorov-Smirnov test was performed to check if the empirical distributions and the theoretical distributions were close enough. None of the empirically computed null distributions failed the K-S test. Thus, we have confirmed that the theoretical derivations of the null distributions are correct and the software implementing the theory is also correct. The software to perform the statistical tests has been made publicly available.

APPENDIX

PROOF OF THEOREM 1

Proof. From the given condition, we know that the derivative matrix H is not a zero matrix, i.e.,

$$H = \begin{pmatrix} \frac{\partial h_1}{\partial \Theta} \\ \vdots \\ \frac{\partial h_r}{\partial \Theta} \end{pmatrix} \neq 0.$$

Left multiplying equation $\Delta\Theta = J\Delta X$ with H , we get $H\Delta\Theta = HJ\Delta X$. Because both (X, Θ^*, Λ^*) and $(X + \Delta X, \Theta^* + \Delta, \Lambda^* + \Delta\Lambda)$ are local minimum points of the optimization, $H\Delta\Theta = 0$. Hence, $0 = HJ\Delta X$. Since the formula holds for any ΔX , it implies that

$$0 = HJ. \quad (12)$$

Now, we use this result to prove that $J\Sigma J^T$ is singular. Left multiply $J\Sigma J^T$ with H and right multiply it with H^T . From (12), we have

$$HJ\Sigma J^T H^T = 0\Sigma 0.$$

Since H is not a zero matrix, $J\Sigma J^T$ must be singular. \square

ACKNOWLEDGMENTS

The authors would like to thank Professors Michael Perlman and David Madigan from the Statistics Department and Professors Terry Rockafellar and Jim Burke from the Mathematics Department for providing relevant literature and discussion. This work was funded in part by US Defense Advanced Research Projects Agency contract 92-F1428000-000. The authors would also like to thank the five anonymous reviewers for their comments and suggestions.

REFERENCES

- [1] M. Altman and M. McDonald, "Choosing Reliable Statistical Software," *PS: Political Science and Politics*, vol. 24, pp. 681-688, 2002, http://www.hmdc.harvard.edu/micah_altman/numal/tpm3.pdf.
- [2] T.W. Anderson, *Introduction to Multivariate Statistical Analysis*. New York: John Wiley & Sons, 1984.
- [3] S.F. Arnold, *Mathematical Statistics*. Prentice-Hall, 1990.
- [4] D.P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*. New York: Academic Press, 1982.
- [5] P.A. Bromiley, M. Pokrik, and N.A. Thacker, "Computing Covariances for Mutual Information Co-Registration," *Proc. Medical Image Analysis Conf.*, 2004.
- [6] G. Casella and R.L. Berger, *Statistical Inference*. Wadsworth, 1990.
- [7] A. Chen, R. Drachenberg, and J. Symanzik, "Assessing the Reliability of Web-Based Statistical Software," *Computational Statistics*, vol. 18, pp. 107-122, 2003, http://www.math.usu.edu/~symanzik/papers/2003_computstat.pdf.
- [8] K. Cho, P. Meer, and J. Cabrera, "Performance Assessment through Bootstrap," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, pp. 1185-1198, 1997.
- [9] A.R. Chowdhury and R. Chellappa, "Statistical Error Propagation in 3D Modeling from Monocular Video," *Proc. IEEE Workshop Statistical Analysis in Computer Vision*, 2003, <http://www.cse.lehigh.edu/~rjm2/SACV/papers/chowdhury-chellappa.pdf>.
- [10] P. Courtney and N.A. Thacker, "Performance Characterisation in Computer Vision: Statistics in Testing and Design," *Imaging and Vision Systems: Theory, Assessment and Applications*, pp. 109-128, 2001, <http://peipa.essex.ac.uk/benchmark/tutorials/epsrscs2002/handout/epsrscs%2002-handout.pdf>.
- [11] S. Crossley, N.A. Thacker, and N.L. Seed, "Benchmarking Physical Scene Modelling," *Proc. British Machine Vision Conf. (BMVC)*, pp. 346-355, 1998.

- [12] B. Efron and R. Tibshirani, *An Introduction to the Bootstrap*. Chapman and Hall, 1993.
- [13] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. Cambridge, Mass.: MIT Press, 1993.
- [14] R. Fletcher, *Practical Methods of Optimization*. New York: John Wiley & Sons, 1987.
- [15] J. Ghezzi, M. Jazayeri, and D. Mandrioli, *Fundamentals of Software Engineering*. Pearson Education, 2002.
- [16] P.E. Gill, W. Murray, and M.H. Wright, *Practical Optimization*. New York: Academic Press, 1981.
- [17] R.M. Haralick, "Propagating Covariance in Computer Vision," *Proc. IAPR Int'l Conf. Pattern Recognition*, pp. 493-498, Oct. 1994.
- [18] D.A. Hudson, "Inverse Perspective Projection Algorithm: Error Estimation," master's thesis, Univ. of Washington, Seattle, 1990.
- [19] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Elsevier Science, 1996.
- [20] T. Kanungo and R.M. Haralick, "Multivariate Hypothesis Testing for Gaussian Data: Theory and Software," Technical Report ISL: ISL-TR-95-05, Univ. of Washington, Seattle, Oct. 1995. Software available from: <http://lib.stat.cmu.edu/general/mvntest.gz>.
- [21] T. Kanungo, M.Y. Jaisimha, J. Palmer, and R.M. Haralick, "A Methodology for Quantitative Performance Evaluation of Detection Algorithms," *IEEE Trans. Image Processing*, pp. 1667-1674, 1995.
- [22] K. Koch, *Parameter Estimation and Hypothesis Testing in Linear Models*. New York: Springer-Verlag, 1987.
- [23] X. Liu, "Modeling and Performance Characterization of 3D Parameter Estimation Using Perspective Geometry," PhD thesis, Univ. of Washington, Seattle, 1995.
- [24] X. Liu, R.M. Haralick, and A. Bedekar, "Optimization Methods for Estimating 3D Object Parameters," *Proc. 1994 DARPA Image Understanding Workshop*, pp. 1017-1025, 1994.
- [25] X. Liu, R.M. Haralick, and K. Thornton, "Site Model Construction Using Geometric Constrained Optimization," *Proc. 1996 DARPA Image Understanding Workshop*, 1996.
- [26] X. Liu, T. Kanungo, and R.M. Haralick, "Statistical Validation of Computer Vision Software," *Proc. 1996 DARPA Image Understanding Workshop*, pp. 1533-1540, 1996.
- [27] B.D. McCullough, "Assessing the Reliability of Statistical Software: Part I," *The Am. Statistician*, pp. 358-366, 1998, <http://www.amstat.org/publications/tas/mccull-1.pdf>.
- [28] B.D. McCullough, "Assessing the Reliability of Statistical Software: Part II," *The Am. Statistician*, pp. 149-159, 1999, <http://www.amstat.org/publications/tas/mccull.pdf>.
- [29] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *Numerical Recipes in C*. New York: Cambridge Univ. Press, 1990.
- [30] L. Wilkinson, "Practical Guidelines for Testing Statistical Software," *Computational Statistics: Papers Collected on the Occasion of the 25th Conf. Statistical Computing at Schloss Reinsensburg*, Physica-Verlag, 1994, <http://www.spss.com/research/wilkinson/Publications/Accuracy.pdf>.



Xufei Liu received the MS degree in electrical engineering from Tsinghua University, China, in 1985, after which he worked on image processing and computer vision at Tsinghua University for four years. In 1995, he received the PhD degree from the University of Washington, Seattle, with his research work on computer vision algorithm and performance analysis. From 1995 to 2003, he worked on telecommunication systems at AT&T and at Lucent Technologies.

He joined Cisco in 2004 and is working on system development. He is a member of the IEEE.



Tapas Kanungo received the PhD and MS degrees in electrical engineering from the University of Washington, Seattle, in 1996 and 1990, respectively, and the Bachelor of Engineering from the Regional Engineering College, Trichy, India, in 1986. He is a research staff member at the IBM Almaden Research Center, San Jose, California. Prior to joining IBM, he served as a codirector of the Language and Media Processing Lab at the University of Maryland, College Park, where he conducted research in the areas of document image analysis, OCR-based cross-language information retrieval, pattern recognition, and computer vision. From March 1996 to October 1997, he worked at Caere Corporation, Los Gatos, California, on their OmniPage OCR product. During the summer of 1994, he worked at Bell Labs, Murray Hill, New Jersey, and, during the summer of 1993, worked at the IBM Almaden Research Center, San Jose, California. Prior to that, from 1986 to 1988, he worked on speech coding and online handwriting analysis in the Computer Science Group at the Tata Institute for Fundamental Research, Bombay, India. He cochaired the 2001 and 2002 SPIE Conference on Document Recognition and Retrieval, the 1999 IAPR Workshop on Multilingual OCR, and has been a program committee member for several conferences. He was a coguest editor of the *International Journal of Document Analysis and Recognition* special issue on performance evaluation. His current research activities are in text analysis, pattern recognition, and information retrieval. He has published in the areas of document image analysis, computer vision, database query optimization, information retrieval, and performance evaluation. He received the best paper award at the World Wide Web Conference in 2003 and is a senior member of the IEEE.



Robert M. Haralick received the BA degree in mathematics from the University of Kansas in 1964, the BS degree in electrical engineering in 1966, and the MS degree in electrical engineering in 1967. In 1969, after completing the PhD degree at the University of Kansas, he joined the faculty of the Electrical Engineering Department there, where he last served as a professor from 1975 to 1978. In 1979, Dr. Haralick joined the Electrical Engineering Department at Virginia Polytechnic Institute and State University, where he was a professor and director of the Spatial Data Analysis Laboratory. From 1984 to 1986, he served as vice president of research at Machine Vision International, Ann Arbor, Michigan. From 1986 to 2000, he held the Boeing Clairmont Egtvedt Professorship in the Department of Electrical Engineering at the University of Washington. He is currently a distinguished professor in computer science at the graduate center, City University of New York. Professor Haralick has made a series of contributions in computer vision. His most recent work is in clustering high-dimensional data sets and application of pattern recognition to mathematical combinatorial problems. Professor Haralick is a fellow of the IEEE for his contributions in computer vision and image processing and a fellow of the IAPR for his contributions in pattern recognition, image processing, and for service to IAPR. He has served on the Editorial Board of the *IEEE Transactions on Pattern Analysis and Machine Intelligence* and has been the computer vision area editor for *Communications of the ACM* and an associate editor for *Computer Vision, Graphics, and Image Processing*, the *IEEE Transactions on Image Processing*, and *Pattern Recognition*. He currently serves on the editorial board of *Real Time Imaging* and the editorial board of the *Journal of Electronic Imaging*. He has published more than 550 papers and, in August 1998, completed a two-year term as the president of the International Association for Pattern Recognition.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.