On the Sequential Determination of Model Misfit

Peter Whaite and Frank P. Ferrie

TR-CIM-94-11 September 1994 (Modified August 1995)

Artificial Perception Laboratory
Centre for Intelligent Machines
McGill University
Montréal, Québec, Canada

Submitted to *IEEE Trans. Pattern Analysis and Machine Intelligence*, September 1994

Postal Address: 3480 University Street, Montréal, Québec, Canada H3A 2A7
Telephone: (514) 398-6319 Telex: 05 268510 FAX: (514) 398-7348
Email: cim@cim.mcgill.ca

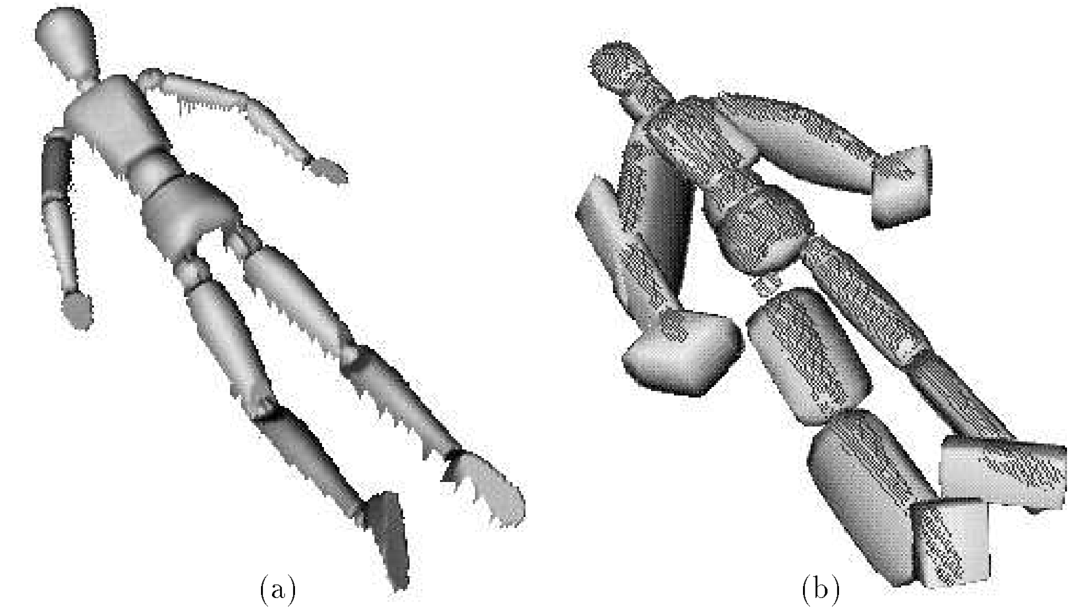
On the Sequential Determination of Model Misfit

Peter Whaite and Frank P. Ferrie

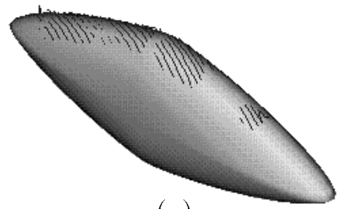
Abstract

Many strategies in computer vision assume the existence of general purpose models that can be used to characterize a scene or environment at various levels of abstraction. The usual assumptions are that a selected model is competent to describe a particular attribute, and that the parameters of this model can be estimated by interpreting the input data in an appropriate manner (e.g. location of lines and edges, segmentation into parts or regions, etc.). This paper considers the problem of determining when these assumptions break down so that an alternate model may be considered or further interpretation of data performed. Specifically, we consider how this can be accomplished with a minimum of a-priori knowledge using an approach that actively builds a description of the environment (i.e. structure and noise) over several different viewpoints. We show that a gaze planning strategy used to minimize model parameter variance can also be used to decide whether the model itself provides an adequate description of the environment.

1. Introduction



(a) (b)



(c)

a) A shaded range image scanned from above a wooden mannequin lying face down.
b) Superellipsoids fitted to segmented data from (a). The dark dots are the range data points. Note that the mannequin's right arm has failed to segment and only a single model has been fitted where two would have been preferable
c) Detail of the superellipsoid fitted to the mannequin's right arm in (b). The dark lines on the surface show the position of the range scans. Although the model doesn't match our perceptual notions of what the arm should look like, it does fit the data well.

Figure 1. Superellipsoid models fitted to a range data from a Wooden Mannequin

Many strategies in computer vision assume the existence of general purpose models that can be used to characterize a scene or environment at various levels of abstraction. They span the range from local characterizations of orientation and curvature [3, 24], to intermediate level representations involving splines and parametric surfaces [1, 7, 8, 24], to still more global representations for solid shape [5, 14, 19]. The usual assumptions are that a selected model is competent to describe a particular attribute, and that the parameters of this model can be estimated by appropriate interpretation of input data. But many of these estimation problems are ill-conditioned inverse problems that cannot be solved without additional constraints derived from knowledge about the environment [15]. This leads to a classical chicken and egg problem where model selection and parameter estimation must be dealt with concurrently, a problem difficult to solve given a single static view of the world.

In this paper we describe an active strategy that permits solution of both problems, i.e. model parameter estimation and model validation. The context is a system for computing an articulated, 3-D geometric model of an object's shape from a sequence of views obtained by a mobile sensor (laser rangefinder) that is free to select its viewpoint [23]. Shape is characterized by general purpose models consisting of conjunctions of volumetric primitives [5]. An active approach is used where the current state of the model, determined from a bottom-up analysis, is used to predict the locations of surfaces not visible in the current view. Gaze is directed to surfaces where the prediction is least certain (maximum variance), and from there additional measurements are made and used to update the model parameters. The validity of the model is tested against its ability to correctly predict the locations of hidden surfaces. Initially both the applicability of the model and estimates of its parameters are uncertain, but as the process unfolds with each successive planning cycle (calculation of new gaze point, measurement, updating of model parameters), such assessments become increasingly clear.

The emphasis of this paper is the model validation problem. Knowing when a particular model fails can provide at least two significant pieces of information. First, it can indicate when assumptions about the scene are wrong and trigger the search for other models that provide a plausible alternative, that is, it can initiate a model selection process. Second, it can indicate when the processes leading to the determination of model parameters have gone awry. This can be used to initiate a backtracking procedure to re-interpret the data, particularly if the validation procedure is also able to indicate the location of *where* the model breaks down. Such would be the case if a model is known to be valid but insufficient data are available from which to correctly apply the model or estimate its parameters.

The example shown in Figure 1 is a case in point and part of the motivation for this research. Figure 1a shows a laser rangefinder image of a wooden mannequin rendered as a shaded image. Based on analysis of surface features, the image is partitioned into regions corresponding to the different parts of the mannequin [6]. A further abstraction is computed by fitting superquadric primitives to each region with the result shown in Figure 1b [5]. At first glance the result appears to capture each of the parts of the mannequin. However, on closer examination (Figure 1c), it can be seen that the partitioning algorithm has missed the cues separating the arm at the right elbow. Superquadrics are appropriate shape descriptors provided that parts are convex, but as Figure 1c shows, do not fit the data well otherwise.

Without contextual knowledge, it is difficult to detect such an error given a single view of the object because there is little basis from which to reject the resulting fit. One would have to know the loci of the occluded surfaces in order assess the model's true fit to the data. However such knowledge is often not possible, e.g. inaccessible viewpoints; or is expensive to obtain, e.g. time required to acquire measurements. The compromise advocated in this paper is a sequential process that incrementally builds its descriptions by optimizing measurement to maximize the certainty of each model, then tests them by verifying their consistency from view to view.

The remainder of the paper is as follows. Section 2 begins with a brief overview of optimization strategy used to plan gaze and estimate model parameters. It provides the necessary background for Section 3 which describes the model validation process, and presents the results of experiments which demonstrate the resulting algorithms at different noise levels and for different noise models. Section 4 shows how the situation shown in Figure 1c can be identified using the gaze planning strategy and model validation procedures. Finally, we conclude with some observations and briefly outline remaining work.

2. Estimating Parameters and Planning Gaze

In earlier work we have considered the problem of how to best direct the gaze of a laser range scanner in order to improve estimates of model parameters and knowledge of object surface position over a sequence of views [20, 21, 23].

The laser scanner is capable (after appropriate transformations) of providing the 3-D coordinates of points sampled from surfaces in the scene. In this scenario it is assumed that the scene is well represented by a conjunction of parametric volumetric primitives, and that data is collected by moving the scanner around on the end-effector of a robot arm (Figure 2). Using methods described in [5] the data are partitioned into sets corresponding to the parts of each visible object. It is assumed that each data set corresponds to a sample of the surface of a single model[1].

Given one of these data sets $\{\mathbf{x}_i | i = 1, \ldots, N\}$ we wish to infer those parameters $\hat{\mathbf{a}}$ that best estimate the true parameters $\mathbf{a}$ of the model in the scene from which the data was collected. In general an exact solution cannot be found because the

[1]In this paper superellipsoids are used to represent parts, but the approach generalizes to other parametric models.

FIGURE 2. Mobile scanner setup. A laser rangefinder with a $1m^3$ field of view is mounted on the end-effector of an inverted Puma 560 robot.

scanner measurements are subject to both systematic and random errors, but a good estimate can be obtained by finding the parameters that minimize the squared sum

$$\chi^2(\mathbf{a}) = \sum_{i=1}^{N} D^2(\mathbf{x}_i, \mathbf{a}), \tag{1}$$

of distances $D(\mathbf{x}_i, \mathbf{a})$ of the data points from the surface of the model. Except for very simple models and distance metrics one usually must resort to iterative techniques, e.g. the Levenburg-Marquardt method [12,16], to perform the minimization.

Provided the estimated parameters fall within the region of parameter space around the true parameters where D is reasonably approximated by its first order linear terms, the classic statistical theory of linear models can describe the parameter errors [16]. This theory tells us that when the errors described by the distance metric are randomly sampled from a normal distribution then the error in the estimated parameters $\delta\mathbf{a} = \hat{\mathbf{a}} - \mathbf{a}$ can be described by a p-variate normal distribution dispersed in the different parameter directions by an amount determined from the matrix of covariances $\mathbf{C}$. Furthermore the quadratic form $\delta\mathbf{a}^T\mathbf{C}^{-1}\,\delta\mathbf{a}$ that defines the distribution is itself randomly sampled from a distribution that obeys a chi-square law with p degrees of freedom. In that case we can find the point of the chi-square distribution χ^2_γ and use it to define the *ellipsoid of confidence*

$$\delta\mathbf{a}^T\mathbf{C}^{-1}\,\delta\mathbf{a} < \chi^2_\gamma, \tag{2}$$

that is an ellipsoidal region of parameter space around the estimated parameters and in which there is a probability of γ that the true parameters lie.

Because of the noise in the model, and because the data are often incompletely sampled, e.g. only one side of the model is visible from a single viewpoint, the parameters will often be under constrained and exhibit large estimation errors. These errors can be reduced by collecting more data, but there are liabilities in terms of cost and accessibility; e.g. the time taken to plan and move the scanner, memory and cpu resources consumed to process additional data, and limits on accessible viewpoints. Ideally we would like to minimize the amount of data collected and the complexity of the movements necessary to place the scanner in the correct position. To do so requires the formulation of a precise relationship between the parameters that govern the data acquisition process and those related to the model being fit.

This task is somewhat difficult because the scanner collects data in the 3-D space of the scene, thus making it difficult to predict the effect that newly collected data points will have on the parameter errors in the p-dimensional space of model parameters. The approach that we have taken to solve this problem is to think of the estimated model as a predictor of surfaces in the scene, and to quantify this error in terms of an interval around each point on the predicted surface. We call this the *surface prediction error interval* and have shown [20] that an "error bar" protruding from a point $\mathbf{x}_s$ on the estimated model's surface is given by the quantity

$$\pm\delta_\gamma(\mathbf{x}_s) = \pm\sqrt{\Delta\chi^2_\gamma}\,\sqrt{\frac{\partial D}{\partial \mathbf{a}}^T \hat{\mathbf{C}}\,\frac{\partial D}{\partial \mathbf{a}}} \tag{3}$$

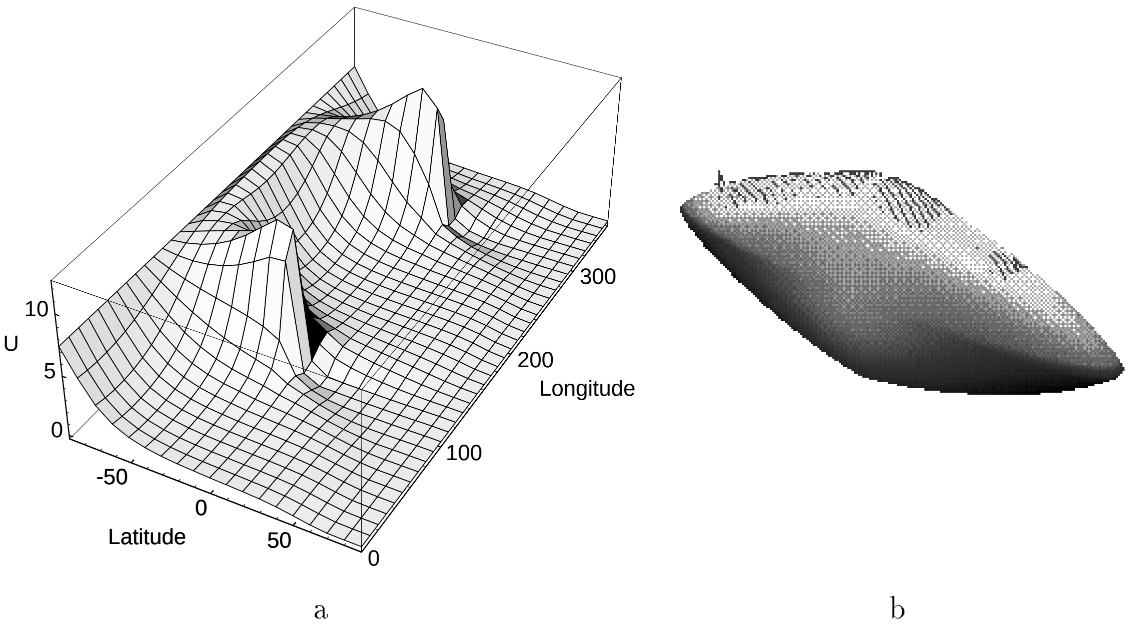
where $(\partial D/\partial\mathbf{a})$ is the gradient of the distance metric evaluated for the point $\mathbf{x}_s$ on the surface of model $\hat{\mathbf{a}}$, and $\Delta\chi^2_\gamma$ is a confidence interval chosen from a chi-square distribution as for the ellipsoid of confidence in (2).

The practical use of this representation for optimizing data collection via gaze planning can be explained with the aid of Figure 3. The figure shows the surface prediction error interval corresponding to the model fit to the arm shown earlier in Figure 1c. In Figure 3a, the interval is coded such that darker shading represents higher uncertainty in surface positions as predicted by the model. Even though the data leading to the model are acquired from a single viewpoint, the resulting prediction extends beyond the visible surfaces and can thus serve as a basis for planning the next gaze direction. An intuitive strategy for doing so would be to direct the scanner to the viewpoint corresponding to the highest uncertainty of prediction. Theoretically we can show that updating model parameters with additional data obtained from this view will minimize the determinant of the parameter covariances [22, 23].

Figure 3a shows a parameterization of the uncertainty surface in the coordinates of a view sphere centered on the the model (uncertainty map), and as can be seen the uncertainty is lowest at the current scanner position, but rises rapidly to a maximum on the opposite side of the view sphere. The optimum strategy here is to move the scanner to the other side of the model, to sample additional data there, and to update the model parameters. However the general problem of gaze planning is much more complex than implied by our example. First, the prediction afforded by the surface error prediction interval is local, so it is unlikely that a complete set of constraining views can be determined on the basis of the model computed from a single viewpoint. In fact the additional data will completely alter the uncertainty map so it must be recomputed after each iteration. Second, the prediction does not take accessibility constraints into account, e.g. certain views may either be unreachable by the scanner, or occluded by surfaces not visible from the current viewpoint. So, as in our example, it is often the case that "the other side" of the model cannot be reached.

In spite of these difficulties, we have found that using uncertainty to plan *incremental* displacements of gaze angle relative to the current viewpoint can result in a successful strategy [21, 23]. We apply a hill climbing algorithm to the changing uncertainty map and use the resulting path to guide the trajectory of the mobile scanner. This can result in a near optimum a data collection strategy with respect to the rate of convergence of model parameters. Also, lack of accessibility is often not that great a problem. For example when representing convex surfaces with superellipsoid primitives we have observed that well constrained parameter estimates can be obtained by taking data with the scanner displaced approximately $30°$ either side of the initial gaze position. This is because the model can interpolate across large "holes" in the data set.

However the success of the exploration strategy hinges on the central assumption that the model fits the data. If this is not the case the parameter covariances, and therefore the surface prediction error, do not accurately reflect the constraints



The figure shows predicted surface uncertainty $U = \sqrt{(\partial D/\partial \mathbf{a})^T \hat{\mathbf{C}} (\partial D/\partial \mathbf{a})} = \delta_\gamma(\mathbf{x}_s)/\sqrt{\Delta\chi^2_\gamma}$ as an **uncertainty map** (a) where U is plotted as a function of view sphere latitude and longitude, and as an **uncertainty surface** (b) where the surface of the model is shaded such that darker shading corresponds to higher values of U.

The lines on the top of the uncertainty surface show the data collected when the scanner was positioned at the north pole of the view sphere (latitude=90°), and to which the model was fitted. As can be seen U is low where data exists, but increases as the model attempts to extrapolate away from the data. The maximum uncertainty lies under the sharp ends of the model, and is marked by the tall peaks on the uncertainty map.

The scanner is initially located at the right edge of the uncertainty map. When it moves to the next view position it will follow the *local* uncertainty gradient, and will therefore move up the center of the broad ridge extending out between the two peaks, i.e. towards the south pole along a longitude of approximately 220°. This corresponds to a path that samples the side of the model facing the viewer in (b).

FIGURE 3. Two different representations for the surface prediction error interval

placed on the model by the data. Thus to ensure a meaningful sensor trajectory it is necessary to test the validity of the model at each iteration.

3. The Detection of Misfits

Implicit in our "bottom-up" approach to vision is the notion of "increasing specificity" as processing moves from the lower to the higher layers. By doing things this way we can build computationally intensive lower layers that operate very generally, yet still provide usable data to higher layers designed for specific tasks. However specialized algorithms are usually tuned to a set of assumptions more restrictive than can be truthfully applied to input data processed by the lower layers. Consequently it is necessary to check the validity of the data before proceeding.

Such a necessity becomes apparent when we fit volumetric models to segmented range data. The segmentation algorithm we use [5] deliberately avoids detailed assumptions about the exact shape of the primitives (e.g. that they be symmetric) and requires only that they be convex. To this segmented data we fit models designed to represent the kinds of shapes expected in the world. In our case, because they can economically portray a wide range of symmetrical shapes, we use superellipsoids. The problem is that not all convex shapes are superellipsoids, so while the segmentation algorithm may have correctly processed its input data, there is no guarantee that a valid superellipsoid model can be made to fit it.

The most straightforward means of evaluating the validity of the data is simply to fit and see. If the model fits well then all of the data should lie on or close to its surface. If not there will be significant residual errors, the model can be declared a misfit, and the flow of processing altered to take remedial action. Because the data are subject to random fluctuations it is not possible to conclude that there has been misfit (or that there has not) with complete certainty. We show how to deal with this problem using methods found in the statistical field of decision theory [4, 13].

In the theory that follows we develop three lack-of-fit statistics, each one useful in different situations. The first of these (L_1) requires an accurate model of the data noise, and knowledge of the parameters of that model, in particular the value of the noise level. When the noise level is not known but the noise model is, then the second lack-of-fit statistic (L_2) can be used. It requires repeat measurements of the data in order to provide an independent estimate of data noise, and therefore incurs additional time and processing costs (e.g. it takes about 12 seconds to scan a 256×256 image with the McGill-NRC scanner). In situations where a rapid response is required, and where the noise not known, we propose an incremental lack-of-fit statistic (L_3) which "learns" the local noise level as the scanner moves through the scene. Our experimental results suggest that the measure is able to detect model misfit even if the real noise is not well modeled by the theory.

3.1. Theory. In the discussion that follows we will assume that we have at our disposal a sequence of n_s data sets $S_0, S_1, \ldots S_{n_s}$ of 3-D coordinates $S_j = \{\mathbf{x}_i, i = 1 \ldots n_j\}$ obtained by moving a laser range scanner along some trajectory through the scene. The S_j are not necessarily the original data scans, but are subsets picked out by a segmentation algorithm as having come from the same convex surface. There is also a finite chance that the segmentation algorithm has incorrectly partitioned the data.

3.1.1. *Known sensor noise.* We will first consider the case for which the data noise meets the conditions assumed by the fitting procedure, i.e. that the data is normally distributed in a direction radially about the surface of the true model with zero mean and *known variance* σ^2.

For each step j in the sequence of views we find the model parameters $\hat{\mathbf{a}}_j$ that minimize the least squared error of the combined data sets $S_j^T = S_0 \cup S_1 \cup \ldots S_j$. We do this by iteratively minimizing the following functional,

$$D_4 = |\mathbf{x}_i| \left(1 - \frac{1}{f^{\epsilon_1/2}(\mathbf{x}_i, \hat{\mathbf{a}}_j)}\right), \tag{4}$$

where $f(\mathbf{x}_i, \hat{\mathbf{a}}_j)$ is the implicit equation of the surface of a superellipsoid [18, 20]. Despite the nonlinearity of the model we will assume that a global minimum error has been found and that the errors are small enough so we can linearize the model and apply the well-known result from linear least squares theory — that an unbiased estimate $\hat{\sigma}_j^2$ of the true variance σ^2 can be found from the squared sum of the residuals (which are measured by the D_4 metric),

$$\hat{\sigma}_j^2 = \frac{\sum_{\mathbf{x}_i \in S_j^T}^{n} D_4(\mathbf{x}_i, \hat{\mathbf{a}}_j)^2}{N_j - p}, \tag{5}$$

where N_j is the total number of data points and p is the number of parameters used to fit the model ($p = 11$ for superellipsoids).

Unexpectedly large values of $\hat{\sigma}_j^2$ indicate that the residual errors are not solely due to noise, and therefore give us grounds for believing that the model fits the data badly. A simple strategy to detect misfit is to find those cases for which

$$\frac{\hat{\sigma}_j^2}{\sigma^2} > k_v, \tag{6}$$

where k_v is a threshold used to decide whether models should be accepted or rejected.

Because of random data noise it is impossible to find a value of k_v that correctly classifies the models in all situations, and we must learn to live with two types of detection errors. The first of these, the Type I error, occurs when a model fits well but chance variations increase the value of $\hat{\sigma}_j^2$ enough that the model is erroneously rejected. The other, the Type II error, is the alternative; that a model fits the data badly but random variations result in a reduction of $\hat{\sigma}_j^2$ large enough to cause the model to be erroneously accepted. In general there is a tradeoff — larger values of k_v decrease the chance of Type I errors but increase the possibility of Type II errors.

It is possible to evaluate the Type I error. When a model fits the data and the residuals are distributed normally, the statistic

$$W = (n-p)\frac{\hat{\sigma}^2}{\sigma^2} \tag{7}$$

is known to be sampled from a chi-squared distribution with $n-p$ degrees of freedom. Thus the probability of a Type I error is

$$P\left(\hat{\sigma}_j^2/\sigma^2 > k_v \text{when the model fits}\right) = P\left(W > (n-p)\ k_v\right). \tag{8}$$

Graphically it is the area under the chi-squared probability distribution to the right of $(n-p)\ k_v$.

However it usually makes more sense to work the other way around; that is from the probability distribution find the value of k_v which gives a tolerable Type I error. The level is often expressed in terms of a *confidence level* γ, or the probability of correctly classifying the good models as good. Knowing that (7) follows a chi-squared law we can find the point $\chi^2_{\gamma,n-p}$ on the distribution for which the probability of a Type I error is $P(W > \chi^2_{\gamma,n-p}) = 1-\gamma$, and reject models as misfits at the γ level of confidence when

$$\mathrm{L}_1 = \frac{\hat{\sigma}_j^2}{\sigma^2} > \frac{\chi^2_{\gamma,N_j-p}}{N_j-p}. \tag{9}$$

In contrast to the Type I error, it is very difficult to find the expected levels of Type II error. The reason for this is that the Type II error is the probability that $\hat{\sigma}_j^2/\sigma^2 < k_v$ given that *the model does not fit* the data. The number of different data configurations that can lead to this situation is so huge, and the interaction of the fitting algorithm to them so unpredictable, that it is impractical to find the probability distribution of $\hat{\sigma}_j^2$ that takes into account all the ways in which a model can be misfitted.

3.1.2. *Unknown sensor noise level.* When the true level of data noise is unknown we can use an estimate of it, provided that estimate is independent of the model fitting process. One way to do this is to exploit repeated measurements. Suppose at some stage during an experiment the laser beam has hit locations $i = 1 \ldots q$ on surfaces of the scene, and at each location we have made m_i measurements. An estimate of the variance, often called the *pure estimate* $\hat{\sigma}_R^2$, is

$$\hat{\sigma}_R^2 = \sum_{i=1}^{q} \sum_{k=1}^{m_i} |\mathbf{x}_{ik} - \overline{\mathbf{x}}_i|^2 / m_R, \tag{10}$$

where $m_R = \sum_{i=1}^{q}(m_i - 1)$ and $\overline{\mathbf{x}}_i = \sum_{k=1}^{m_k} \mathbf{x}_{ik}/m_i$ is the mean value of the measured surface coordinates at location i. If a model fits the data well $\hat{\sigma}_{Rj}^2$ and $\hat{\sigma}_j^2$ computed for the first $j+1$ data sets should be approximately equal[2]. A lack-of-fit statistic that uses the weighted difference of the two estimates relative to the pure estimate is [2]

$$\text{Ł}_2 = \frac{(\hat{\sigma}_j^2(N_j - p) - \hat{\sigma}_{Rj}^2 m_{Rj})/(N_j - p - m_{Rj})}{\hat{\sigma}_{Rj}^2}. \tag{11}$$

Ł_2 can be shown to be sampled from an F ratio distribution with $N_j - p - m_{Rj}$ numerator and m_R denominator degrees of freedom. Models can be rejected at the γ level of confidence when $\text{Ł}_2 > F_{\gamma, N_j - p - m_{Rj}, m_{Rj}}$.

3.1.3. *Consecutive Estimates of Variance.* When repeated measurements cannot be taken we propose that misfit can be detected by comparing consecutive estimates of variance. If the model $\hat{\mathbf{a}}_{j-1}$ fitted to the first j data sets $S_0 \cup S_1 \ldots S_{j-1}$ is valid then the estimated variance $\hat{\sigma}_{j-1}^2$ should be a valid estimate of data noise. If on the next iteration the variance $\hat{\sigma}_j^2$ found after adding S_j is significantly greater than $\hat{\sigma}_{j-1}^2$ we have grounds for believing that the model cannot account for the additional data and that it is therefore unacceptable. It is difficult however to evaluate the Type I errors, and therefore to design a test at the appropriate level of confidence. One might think that because $\hat{\sigma}_{j-1}^2$ and $\hat{\sigma}_j^2$ are sampled from chi-squared distributions an F distribution would correctly account for their ratio. Unfortunately, this relationship is true only if the chi-squared distributions are independent. Because they share coordinates from the first j data sets, such is obviously not the case.

[2] To avoid any confusion, the index j is added to variable subscripts to indicate the sequential order of data samples and their statistics, e.g. $\hat{\sigma}_j^2$ is the sample variance computed over the first $j+1$ data sets.

The approach we have taken is to minimize the dependency by using only the residuals of the newly added points to estimate the data noise. First we compute $\hat{\sigma}^2_{j-1}$ in the usual way (5) from *all* of the data in the first j data sets $S^T_{j-1} = S_0 \cup S_1 \cup \ldots S_{j-1}$. The other estimate of variance $\hat{\sigma}^2_{Pj}$ is computed using *only* the data in S_j, that is

$$\hat{\sigma}^2_{Pj} = \frac{\sum_{\mathbf{x}_i \in S_j} D_4(\mathbf{x}_i, \hat{\mathbf{a}}_j)^2}{n_j}, \tag{12}$$

where in this case n_j is the number of data in S_j, but the model $\hat{\mathbf{a}}_j$ is the least squares fit to *all* of the data S^T_j. Because the residuals are distributed normally then $\hat{\sigma}^2_{j-1}$ and $\hat{\sigma}^2_{Pj}$ are sampled from a chi-squared distributions with $N_{j-1} - p$ and n_j degrees of freedom respectively. Therefore when the two variance estimates are independent the incremental lack-of-fit statistic

$$\mathrm{L}_3 = \frac{\hat{\sigma}^2_{Pj}}{\hat{\sigma}^2_{j-1}} \tag{13}$$

is sampled from an F ratio distribution with n_j numerator, and $N_{j-1} - p$ denominator, degrees of freedom. Models can be rejected as misfits at the γ level of confidence when $\mathrm{L}_3 > F_{\gamma, n_j, N_{j-1}-p}$.

However the L_3 lack-of-fit statistic should be used with caution because the estimate of $\hat{\sigma}^2_{Pj}$ is biased. In effect some of the data variability is used to compute the model parameters, and this loss results in an estimate of variance lower than it should be. We compensate for the loss in (5) by dividing by $N_{j-1} - p$, that is p points have been used up fitting the p model parameters. When we take only a subset of the data as in (12) it is hard to arrive at an appropriate compensatory figure; mainly because it is difficult to evaluate the relative influence exerted by the subset on the fit. The lower bias of $\hat{\sigma}^2_{Pj}$ will be compensated to some degree by a narrower confidence interval in an F distribution with a higher number of degrees of freedom so the overall effect is probably minor and will in any case decrease as the number of data points increases.

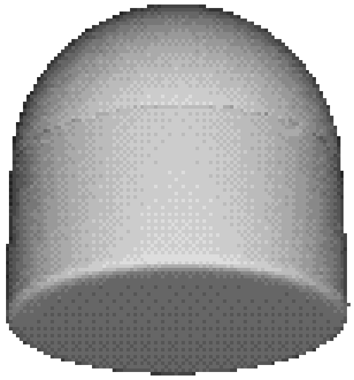
With the L_3 metric, $\hat{\sigma}^2_{Pj}$ is calculated over a more localized region of the surface. Given that surface features causing misfit are most likely to be in the newly scanned region then the mean squared residual error here will be higher than if it were computed over the entire region so far scanned. The result is an apparent increase in the metric's sensitivity to misfit error. However this sensitivity is offset by a higher confidence threshold due to a lower number of degrees of freedom in the chi-squared distribution of $\hat{\sigma}^2_{Pj}$ calculated from fewer data points.

An implicit assumption when using the L_3 statistic is that $\hat{\mathbf{a}}_{j-1}$ is a valid fit, and that $\hat{\sigma}^2_{j-1}$ is a valid estimate of the data noise level. By induction it must also be true that the initial estimate $\hat{\mathbf{a}}_0$ be a valid fit, so in practice it is up to us to select the appropriate initial conditions which make sure that this is the case. Generally this can be done without great difficulty, and with only a rough a-priori knowledge of the scene being explored. For example by knowing the minimum size of objects in the scene one could limit the initial scan to a small region, and validly fit it to the surface of almost any large model (even a planar patch).

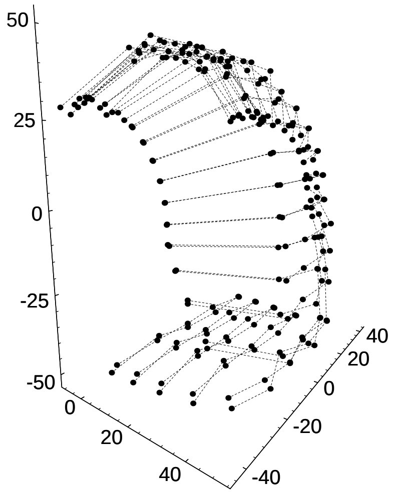
3.2. Simulation Experiments. In the experiments that follow we used a scene synthesized from two superellipsoid models, a sphere and a cylinder both of $50mm$ radius, but joined so as to blend smoothly and form an squat cylindric shape with a spherically domed top (Figure 4). This shape was chosen because the overall convexity of the surface ensures that it will not be partitioned by the segmentation algorithms. Data collected from the top of the scene can be initially modeled with a superellipsoid, but as the scanner moves from the top of the scene to a view of the bottom the misfit increases, at first slowly as more of the cylindrical edge is exposed, then abruptly when the flat bottom surface comes into view.

Range data is sampled from the scene using a computer simulation of the McGill-NRC range scanner we have in our laboratory [17]. The camera is always directed so that its line of sight is towards the origin of the scene coordinate system, but is allowed to move around on the surface of a view sphere of radius ϱ, also centered on the scene origin. Camera position is specified by a latitude and longitude (ϑ, φ) set up with respect to the scene coordinate frame such that the positive Z axis intersects the view sphere at its north pole ($\vartheta = 90°$), and so that the X-Z plane cuts the view sphere around the meridian of zero longitude. Our scanner uses two mirrors to sweep the laser beam over a field of view $36.9°$ by $29.2°$ along the camera's X and Y axes respectively. In both cases the mirror angles are controlled by an index between 0 and 256 that divides the field of view into equi-angular increments. The sampling is specified by two triples of numbers $\{i_{min}, i_{max}, i_{inc}\} \times \{j_{min}, j_{max}, j_{inc}\}$ where the X mirror is moved from index i_{min} to i_{max} in steps of i_{inc}, and likewise for the Y mirror and the j indices. If a mirror is not moved, for example when only a single scan line is taken, then the redundant maximum and incremental values will be dropped. We call the array of data collected by scanning the X and Y mirrors a *range image*.

An advantage of a simulated range scanner is that it is very easy to implement and investigate the effect of different noise models. For these experiments we will use a *radial noise* model in which normally distributed noise (i.e. Gaussian noise) is added so as to displace the data point from the surface in a direction radial to



a) We use a scene composed of two superellipsoid models, a sphere and a cylinder, joined to make a smooth transition. Although the compound model is convex it cannot be described by a single superellipsoid surface. The scene above is as seen from a view sphere latitude of about $-20°$.



b) A typical sequence of data. The dots mark the data points, and the lines show the direction of the scan lines, which are doubled to obtain repeat measurements. A radial noise model ($\sigma = 2.0$mm) was used.

FIGURE 4. The 3-D Scene and Data used in these experiments

the model's center. This noise model matches the assumptions upon which the least squares minimization is based, and therefore those of the tests that detect misfit.

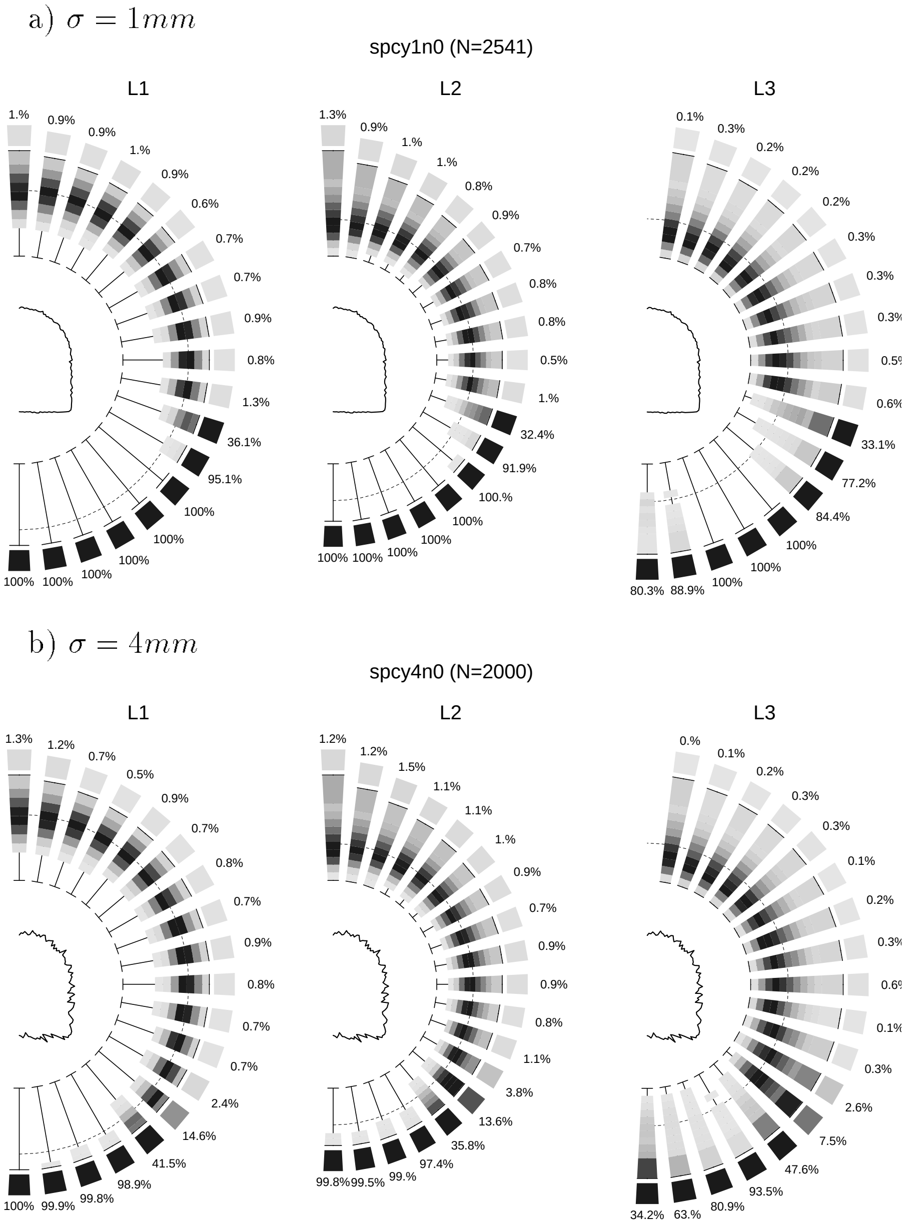
Unless otherwise stated all of the following experiments will be performed with sets of data collected in the following way. Initially the scanner is moved to $\vartheta = 90°$ on a view sphere of radius $\varrho = 150mm$, and the scene is *twice* sampled coarsely ($\{0, 256, 64\} \times \{96, 160, 16\}$) to give 50 points (a repeated 5×5 range image). The field of view is such that the scanner sees only the spherical surface. After an initial scan, additional data from a sequence of views is taken by moving the camera along the meridian of $0°$ longitude in $10°$ increments of latitude until it reaches the south pole. At each position a single line of data ($\{0, 256, 32\} \times \{128\}$) is *twice* scanned to give a set usually containing 10 points (a repeated 5×1 image). The result is a sequence of 19 data sets ordered according to the latitude, so S_0 is the initial set collected from the north pole, and S_{18} is collected from the south pole. In every data set there are repeat samples of each point, so we can evaluate all 3 lack-of-fit statistics under exactly the same conditions. Figure 4 shows a 3-D rendition of a typical sequence of scans with added radial noise ($\sigma = 2.0mm$).

The first set of experiments was designed to evaluate the performance of the three lack-of-fit measures for radially distributed noise, i.e noise in agreement with the assumptions upon which the lack-of-fit statistics are based. A large number of trials were performed at two different noise levels: one about twice that typically observed in our laboratory ($\sigma = 1.0mm$, $N = 2541$), and the other observed when sampling from the limits of the scanner's range ($\sigma = 4.0mm$, $N = 2000$). In each trial a sequence of data was obtained by moving the scanner in $10°$ steps as described above. At each step the three lack-of-fit statistics were evaluated, and accumulated into the corresponding histogram at that step. On completion we obtained a sequence of histograms showing the progression of each lack-of-fit statistic as the scanner discovered the model surface while moving from the top to the bottom of the scene. The results are shown in Figure 5.

We obtain the theoretically expected results for viewsphere latitudes from $90°$ down to $0°$. Here the scanner is just sampling the surface of the sphere, so a valid superellipsoid fit can be obtained. The histograms indicate that for the L_1 and L_2 statistics approximately 1% of the trials exceed the 99% confidence level, and that the histogram value is close to the expected value of 1%. The misfit level is somewhat lower than expected for the L_3 statistic, and the histogram peak is also displaced downward. As mentioned in the theoretical discussion, an effect like this could be due to overestimation of the degrees of freedom when calculating $\hat{\sigma}^2_{Pj}$. Only 5 data points were used in these computations, so an additional degree of freedom would cause a significant decrease in the value of L_3.

For latitudes below $0°$ there is a gradual rise in the rate of misfits, until by $-40°$ almost all of the trials are classified as such. This behaviour also matches that expected, with the slow increase marking the transition region where it becomes increasingly difficult to describe the surface shape as superellipsoid, and the abrupt change indicating gross violations of the assumed symmetry.

The L_3 statistic is not as sensitive as the other two in detecting misfit, and again we would expect that behaviour. Because it compares variance estimates at adjacent latitudes, the L_3 statistic is really detecting incremental increase in misfit, and can therefore be fooled when the misfit is increased slowly. Another way of looking at this is to think of L_3 as adapting to "learn" the noise. Thus we see that unlike L_1 and L_2, the L_3 statistic does not reject fits at latitudes $\vartheta < -60°$ because it has adapted itself to the very high levels of $\hat{\sigma}^2_j$ found at $\vartheta = -60°$.



Histograms are rendered radially at their corresponding view sphere latitude. Each bin is coloured a level of grey in proportion to the number of values falling within it. The number of trials used to compute the histograms is shown in parentheses above each figure.

Each histogram has been computed by dividing the theoretical one-sided 99% confidence interval (shown underneath the histograms) into 11 bins. The first 10 bins split the 95% interval up into equal parts, while the remaining one shows the other 4%. Values exceeding 99% confidence threshold are all accumulated into the outer bin and the percentage falling here is indicated beside it. When the model fits we would expect this figure to be 1%. The dotted circle marks a lack-of-fit statistic value of 1.0. It should coincide with the histogram maximum.

FIGURE 5. Comparison of L_1, L_2, and L_3 misfit tests for radial noise model with noise levels of (a) $1mm$ and (b) $4mm$.

At the higher noise level ($\sigma = 4.0mm$) all three statistics are close to the limits of their ability to discriminate, and only the gross misfit is detected. In fact the noise level is so high that it is starting to obscure the viewer's perception of the corner of the cylinder in the profile data. Noise of this level would not be encountered on our scanner except when measuring surfaces at the limits of its range.

3.3. Real experiments. The simulations confirm the correctness of theory but to what extent is this true when using real scanners for which the theoretical noise models are only an approximation? To test this we have used the apparatus shown in Figure 6 to perform the same experiments but with real data. The apparatus consists of the McGill–NRC scanner mounted in a fixed position with a view of an object clamped to the rotational axis of a small stage. Different parts of object can be scanned by using stepper motors to rotate the stage about two orthogonal axes. Before the experiment begins a calibration procedure is run to determine the orientation and position of the two rotational axes. Once known, the angles of rotation can be used to map scanner range coordinates into a scene frame attached to the rotating object.

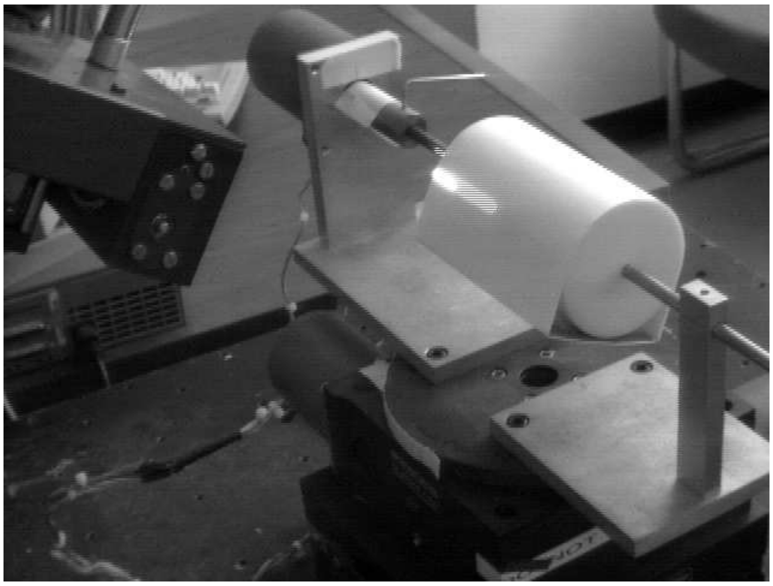


FIGURE 6. Rotary stage used in the real experiments showing the compound model comprised of a smoothly joined cylinder and block

A side effect of the calibration procedure is that it provides us with an estimate of the sensor noise σ. The axes are found by measuring, at several different rotations, the orientation of an inclined plane attached to the stage. For each orientation we can estimate the sensor noise from the residual errors left after fitting a plane to the scanned data. Figure 7 shows that σ varies with orientation, that it depends mainly on the angle the plane makes with the scan direction, and that it is minimum when the surface is normal to the scanner's line of sight. It is well known that σ also varys with the distance to the surface, that it depends on surface properties, and that it can change with time. In general these factors make it very difficult to choose a constant value of σ demanded by the misfit statistics, but for these experiments we have taken the average minimum value over several calibration runs ($\sigma = 0.17$mm). Our choice is motivated by the fact that most of the data are taken from surfaces normal to the scan beam, and that the distance to the surface is approximately that of the calibration plane.

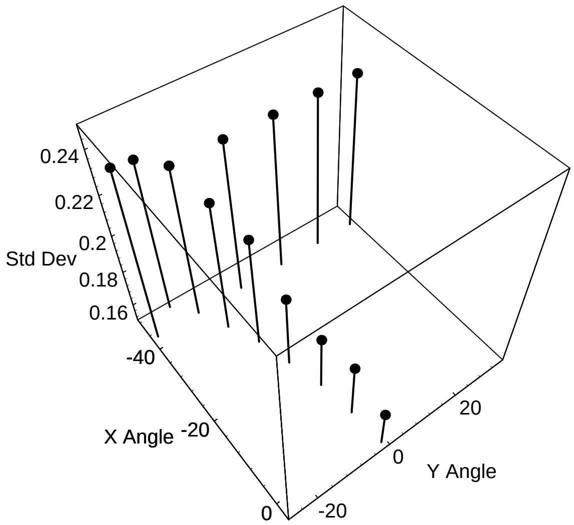


FIGURE 7. Sensor noise as a function of surface orientation. The figure shows σ as a function of the angle between surface to the scanner's X and Y axes. Most of the variation is due to surface slope in the direction of the scan line (the X direction).

The results of the first experiment are shown in Figure 8. The procedure used was essentially the same as that described in Section 3.2, though we used the smoothly

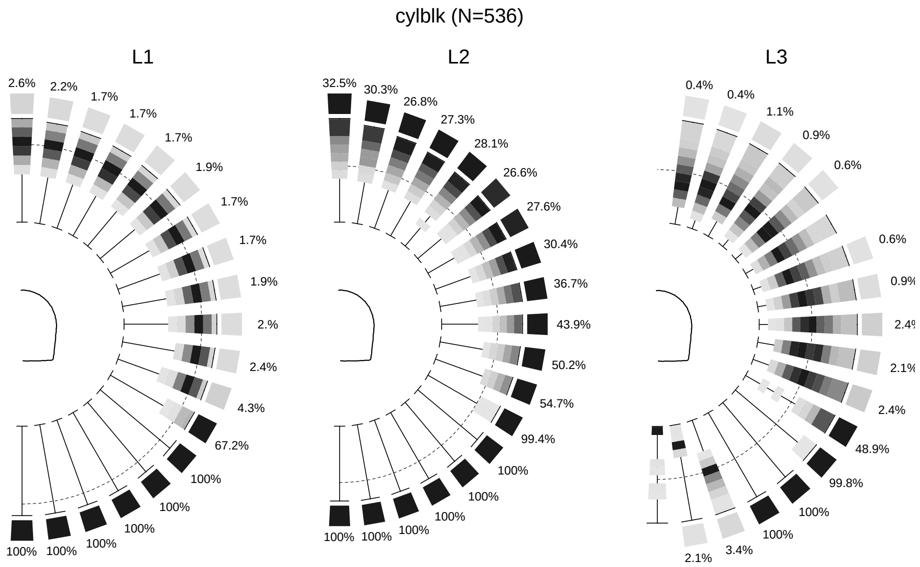


FIGURE 8. Comparison of L_1, L_2, and L_3 misfit tests for real data obtained using McGill-NRC rangefinder.

joined cylinder and block shown in Figure 6 because it was easier to fabricate than the spherically capped cylinder. The sampling was also changed to take into account the different configuration, and to prevent inclusion of points not on the object's surface. The results were accumulated from 536 trials. It took approximately 2 minutes for each trial and around 20 hours to collect the complete data set. In general the results indicate that the L1 statistic overestimates the amount of misfit slightly, that the L2 statistic is in gross error, but that the L3 statistic still behaves very much as predicted by the theory.

The qualitative behaviour of the L1 lack-of-fit statistic matches that in the simulations, except the percentage of trials exceeding the 99% confidence level is about twice that expected (1.7%–2.8% or 9–13 trials). The cause of this discrepancy is indicated in Figure 9 where we show a histogram of the residual errors left after fitting a superellipsoid to a patch of range data scanned from the cylindrical part of the surface (Figure 6). When compared to the normal distribution with standard deviation $\hat{\sigma}$ computed from (5) we observe that the residuals depart from the assumption of normality: there is asymmetry, and the tails of the histogram are somewhat thicker than expected when compared with the width of the peak. One has the impression that the distribution is composed of two or more normal distributions with different variances and offset means, which is the kind of effect expected due to the variation of σ with surface orientation and distance. In addition we observed an overall upward drift in the residual errors over the 20 hour duration of the experiment, indicating that the actual sensor noise worsened during this period. The net result is that the values of $\hat{\sigma}$ obtained from the residual errors are greater than the assumed sensor noise, so the values of the L1 statistic are higher than expected.

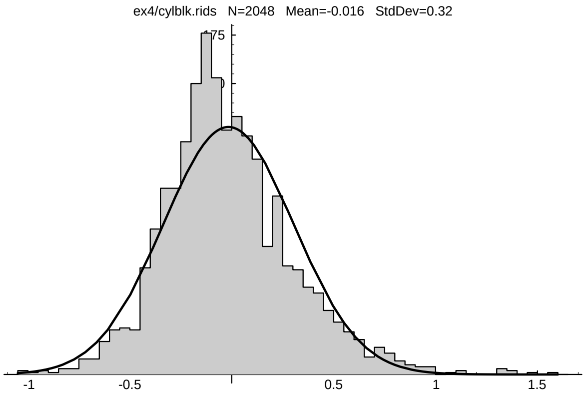


FIGURE 9. Histogram of residual errors left after fitting a superellipsoid to cylindrical data. The solid line shows the normal distribution with the same mean as the residual errors and with a standard deviation $\hat{\sigma}$ computed using equation (5). The dotted line indicates the normal distribution assumed for a sensor noise level of $\sigma = 0.17mm$.

The L2 statistic performs very badly, with the number of trials exceeding the 99% confidence level at around 30 times that expected for a good fit. The reason for the poor performance is that the errors in the repeat data sets are not independent as demanded by the theory. In Figure 10, where we show 8 successive scans of the same patch of surface, it can be seen that there is a noticeable amount of coherency from scan to scan. For example there are similar patterns of variation in scans 4, 5, & 8 for the first 15 mirror positions, and in scans 6, 7, & 8 for the last 12. As a result the noise $\hat{\sigma}_R$ estimated by looking at the differences between successive scans will be significantly less than the variation along a scan, and since the latter is effectively the residual variation left after fitting it will look like misfit to the L3 lack-of-fit statistic. The reason for the repeatability in the "noise" from scan to scan is not exactly known, but we have seen it in other laser range scanners as well. One possibility is that it is caused by speckle interference induced when the laser beam passes through the scanner's optics. However even if this kind of noise was not present in the sensor, exactly the same problem would arise if the surface was roughly textured or patterned. We must conclude that the L2 statistic will only be useful in very specific circumstances.

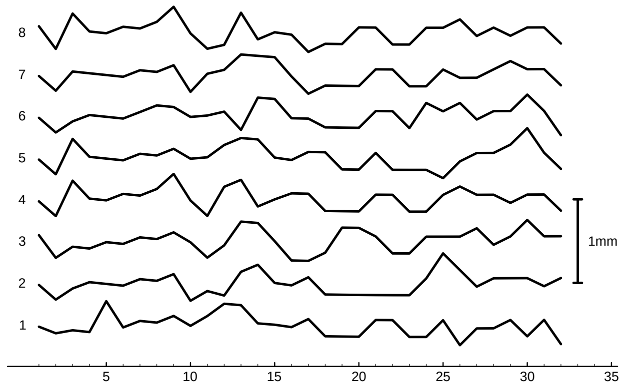


FIGURE 10. Repeated Scans. The figure shows 8 sequential scans taken approximately 1 second apart from exactly the same place on the cylindrical part of the surface used in the experiments. The scans have been offset from each other and a plotted horizontally as a function of the scanner's X mirror index. The vertical scale of each scan is indicated by the 1 mm bar on the left.

In comparison the L3 statistic still behaves as expected, even though the scanner noise characteristics depart from the underlying theoretical assumptions. In fact the results seem to match the theory better than those obtained in the simulations (5) where we observed a lower than expected number of trials exceeding the 99% confidence interval. A possible reason for this is that in the real experiments over twice as many points (12 vs 5) were taken in each scan line so the L3 statistic will be less sensitive to underestimation of degrees of freedom.

A curious point, and one which highlights a limitation of the L3 statistic, is the dip in misfit at a latitude of 30°. If this feature is statistically significant (1% represents only 5 trials in this experiment) the interpretation is that the data obtained at this latitude fits the current model better than the data from all the higher latitudes. That can happen when the measured surface is not exactly superellipsoidal (e.g. because of small errors in the stage calibration). The fitted surface would position itself to minimize the residual error so some parts of it would be inside the measured surface and some parts outside. If the last scan happens to fall near the place the fitted and measured surfaces cross, then the residual errors for it will be lower than average resulting in a low value of L3. We cannot expect the L3 statistic to detect slow departures from the valid class of models, but we can expect it to function well when changes are abrupt (e.g. segmentation errors).

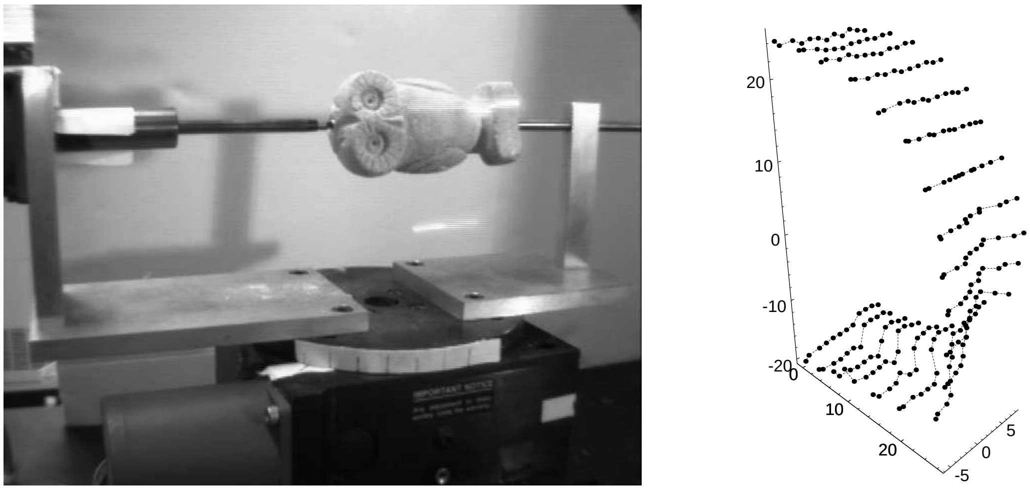
4. An Example

Figure 1 illustrates a scenario which typifies the misfit problem — the jointed right arm of the mannequin has been described using a single model, rather than two as one would expect. In this particular situation segmentation should take place along the local concave creases marking the join between the upper and lower arm. However the discrete sampling of the scanner has "skipped" over the fine detail of the elbow joint. A crease is detected around the elbow, but it is not continuous enough to completely sever the arm data into two surface patches. It can be argued that a more detailed analysis could handle this situation, e.g. [9–11, 24], but there will always be times when it is just not possible to segment smoothly joined, articulated objects at such a low level. Consider the out-stretched human arm — how is the boundary that separates it into the upper and lower arms precisely delineated?

Instead we have to rely on more global models of the surface to provide additional clues as to when data should be partitioned. Model misfit is one such clue, and where it occurs may, under the right conditions, indicate good places to re-partition. However for the arm of the mannequin there is no clue that anything is wrong — the surface and the scanner have conspired to produce unsegmented data that can be fit very well by a superellipsoid model. Only by collecting more data can the structure of the mannequin be correctly inferred and resolved, which brings us back to the gaze planning strategy described in Section 2.

Recall that the strategy operates by directing the scanner to that position on the surface of the current model that exhibits highest uncertainty, or in the case of incremental planning, to a position along the direction of the uncertainty gradient [23]. According to the theory we expect that when data collected at the new sensor position are added to the model $\hat{\sigma}$ will not increase by any significant degree. This can be confirmed by applying an appropriate lack of fit statistic (Table 1). In the event that misfit is detected, further data acquisition can be inhibited until the problem is resolved, e.g. by re-applying the segmentation algorithm to the composite data set.

Another object which can cause the gaze planning strategy to fail is the small owl shown in Figure 11. The problem here is that the crease separating the head of the owl from its body does not completely encircle the neck (Figure 12 top). If the initial data is taken from the back of the owl (Figure 12 bottom) a single model is fit to both the head and the body but the strategy will cause the scanner to move towards



a b

FIGURE 11. The owl. a) View of the owl mounted in the rotary stage. b) A typical sequence of scans collected from a band encircling the region around the owl's neck.

the front of the owl where two models are more appropriate. We investigated the behaviour of the Ł3 statistic in this situation by mounting the owl in the stage so data could be collected from the smooth portion of the back. The initial model fit was cylindrical, though both the Ł1 and Ł2 statistics rejected it outright. The initial misfit is unsurprising given that the soapstone surface is roughly textured, and that the back is slightly concave. A sequence of single line scans was collected by rolling the owl's body over until it faced the scanner — the direction predicted as being the quickest way to improve knowledge of the model surface according to the gaze planning techniques discussed in Section 2.

A typical set of scans is shown in Figure 11b and the Ł3 lack-of-fit histograms in Figure 13. The scale of the histogram has been expanded (the confidence interval is 99.99999%) to reveal the pattern of change even when the misfit is large. Initially the value of the statistic stays below the 99% confidence level but rises rapidly as soon as data are scanned from part of the owl's wing at a latitude of 40°. After this the statistic starts to adapt to the variation exhibited by the wing parts until by 0° the misfit levels have almost dropped back to normal. The abrupt jump at −30° occurs when the scanner encounters the crease around the owl's neck, but the

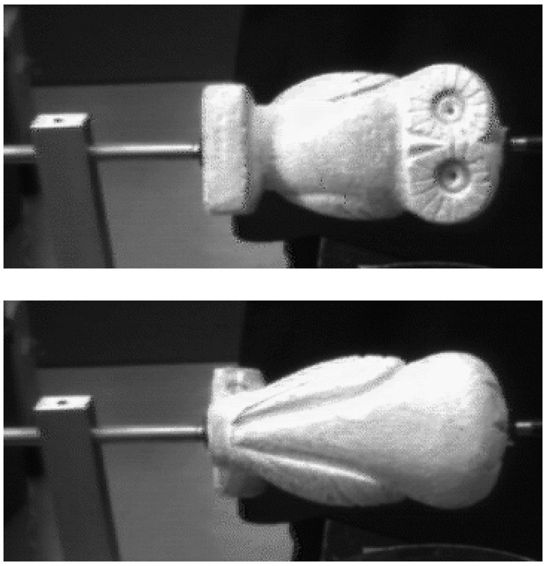


FIGURE 12. Two views of the owl. Depending on whether it is viewed from the front (top view) or from the back (bottom view), the owl can be represented by either two models or a single model respectively.

statistic adapts to this change as well, falling to near normal levels by the time the face is fully in view.

As can be seen by examining the trace of histogram peaks in Figure 13, the Ł3 statistic provides a stable indication of misfit errors associated with the surface boundaries that would normally be determined by segmentation. In practice we have found close agreement between misfit indications based on the Ł3 statistic and empirically determined modeling errors observed in our laboratory system. The assumptions regarding the use of the Ł3 and the other lack-of-fit statistics are summarized below in Table 1.

ASSUMPTION	L1	L2	L3
Sensor noise is normally distributed[3]	yes	yes	yes
Sensor noise level σ known	yes	no	no
Sensor noise level is constant[4]	yes	yes	weakly
Residual errors due only to sensor noise	yes	yes	no
Residual errors spatially independent[5]	yes	yes	yes
Residual errors temporally independent[6]	yes	yes	no
Repeat measurements available	no	yes	no
No initial misfit	no	no	yes

Table 1: Assumptions used in the different lack-of-fit statistics

5. Discussion and Conclusions

The results we obtain match those our intuition leads us to expect. Perhaps this is better illustrated by considering the analogy of an archaeologist who has discovered a object shaped as above but with only the top of the joint protruding from the sand. So great is the antiquity of this object that the original surface detail has eroded, and the discoverer can only guess at its true nature. Initially it appears to be the top of a container of unusual design, perhaps a burial casket, but only further excavation will tell. From the exposed shape the object looks significantly longer than it is wide, and it will therefore be more economical to begin digging down the objects side. This is done and as the excavation proceeds the initial expectations are confirmed — the object still appears to be a casket. However at some depth further digging suddenly reveals a concavity in the objects surface so pronounced that the archaeologist is forced to drop the casket hypothesis and consider others.

[3]In practice the assumption of normality can be weakened. The factor of real importance is that the cumulative lack-of-fit distribution is accurate at the chosen confidence level, because we can then make accurate predictions about the expected rate of misfit due to random chance.

[4]Constancy of the noise can also be weakened in practice, particularly with the L3 statistic.

[5]By spatially independent we mean that errors at different surface locations do not depend on each other. It is not strictly necessary that this be the case, for example the errors could be Markovian provided the scale of interaction is much smaller than the spatial extent of the measurements.

[6]By temporally independent we mean that the errors from exactly the same surface location at different times are independent. For example, the residual errors resulting from a rough surface are not temporally independent.

Thus it is with the arm of the mannequin and the back of the owl. Initially the laser scanner exposes only a part of the surface so our knowledge of the global shape is extremely uncertain. To resolve this uncertainty we must explore, and to guide our exploration we need an initial hypothesis – that the shape is superellipsoid. However we must always be on guard lest that hypothesis fail. This is the role of the test for misfit — to tell us to reconsider, either by choosing a different hypothesis or by re-examining the data. We are particularly interested in the latter scenario because it is common in an active vision context. Very often we have strong prior knowledge about the appropriate model to use for a given task, but fail because the data used to fit the model is wrong, e.g. segmentation errors.

Can we gain any insight into the nature and location of such errors from the exploration procedure? This would be of obvious advantage to a backtracking procedure. In general the answer appears to be no. While we can determine the exact point at which the model fails, we still cannot ascertain whether this is due to the data already collected or to the data newly acquired. In the case of failures due to partitioning errors, our only alternative thus far is to go back and re-sample the data at higher precision such that the segmentation algorithm [6, 10, 11] has a better chance of detecting the missing boundary.

In this paper we have outlined a framework for this process of what we call *autonomous exploration*. We have shown that by using the current estimate of a model to predict the locations of surfaces in yet to be explored regions of a scene, we can both improve estimates of model parameters as well as validate its ability to describe the scene. Knowing when we are wrong is not sufficient. In an unstructured environment an autonomous system must act to correct that wrong either by selecting a more appropriate model or by re-interpreting the data in light of cues provided by the failure of the model. These topics are currently under investigation in our laboratory.

References

[1] P. J. Besl and R. C. Jain. Segmentation through variable-order surface fitting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-10(2):167–192, Mar. 1988.

[2] D. S. Borowiak. *Model discrimination for nonlinear regression models*. M. Dekker, New York, 1989.

[3] P. Breton, L. Iverson, M. Langer, and S. Zucker. Shading flows and scenel bundles: A new approach to shape form shading. In *Computer Vision – ECCV 92*, pages 135–150, Santa margherita Ligure, Italy, 19–22 May 1992. Springer-Verlag.

[4] E. R. Dougherty. *Probability and statistics for the engineering, computing, and physical sciences*. Prentice Hall, Englewood Cliff, N.J., 1990.

[5] F. P. Ferrie, J. Lagarde, and P. Whaite. Darboux frames, snakes, and super-quadrics: Geometry from the bottom up. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(8):771–784, Aug. 1993.

[6] F. P. Ferrie, A. Lejeune, and D. Baird. Curvature, scale, and segmentation. In *SPIE Applications of Artificial Intelligence X*, pages 240–250, Orlando, Florida, April 20-24 1992.

[7] M. Kass and D. Terzopoulos. SNAKES: active contour models. *International Journal of Computer Vision*, 1:321–332, 1988.

[8] D. Keren, D. Cooper, and J. Subrahmonia. Describing complicated objects by implicit polynomials. Technical Report 102, Brown University LEMS, Laboratory for Engineering Man/Macine Systems, Division of Engineering, Brown University, Providence RI 021912 USA, 1992.

[9] B. B. Kimia, A. Tannebaum, and S. W. Zucker. Toward a computational theory of shape: An overview. In *Proceedings of the First European Conference on Computer Vision*, Antibes, France, 1990.

[10] A. Lejeune and F. Ferrie. Finding the parts of objects in range images. *Computer Vision, Graphics, and Image Processing*, Aug. 1993. McGill Center for Intelligent Machines TR 93-8F, Submitted.

[11] A. Lejeune and F. Ferrie. Partitioning range images using curvature and scale. In *PROC. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 800–801, New York City, New York, June 15-17 1993.

[12] D. G. Luenberger. *Linear and Nonlinear Programming.* Addison–Wesley Publishing Company, Reading, Massachusetts, 1984.

[13] A. M. Mood and F. A. Graybill. *Introduction to the Theory of Statistics.* McGraw-Hill Book Company Inc., New York, N.Y., 1963.

[14] A. Pentland and S. Sclaroff. Closed form solutions for physically based shape modelling and recognition. In T. Kanade and K. Ikeuchi, editors, *IEEE Transactions on Pattern Analysis and Machine Intelligence: Special Issue on Physical Modeling in Computer Vision*, volume 13(7), pages 715–729, July 1991.

[15] T. Poggio, V. Torre, and C. Koch. Computational vision and regularization theory. *Nature*, 317(26):314–319, Sept. 1985.

[16] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling. *Numeric Recipes in C – The Art of Scientific Computing.* Cambridge University Press, Cambridge, U.K., 1988.

[17] M. Rioux. Laser range finder based on synchronized scanners. *Applied Optics*, 23(21):3837–3844, Nov. 1984.

[18] F. Solina and R. Bajcsy. Recovery of parametric models from range images: The case for superquadrics with global deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(2):131–147, Feb. 1990.

[19] D. Terzopoulos and D. Metaxas. Dynamic 3D models with local and global deformations: Deformable superquadrics. *PAMI*, 13(7):703–714, July 1991.

[20] P. Whaite and F. P. Ferrie. From uncertainty to visual exploration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1038–1049, Oct. 1991.

[21] P. Whaite and F. P. Ferrie. Uncertain views. In *PROC. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 3–9, Champaign, Illinois, June 15-18 1992.

[22] P. Whaite and F. P. Ferrie. Autonomous exploration: Driven by uncertainty. Technical Report TR-CIM-93-17, Center for Intelligent Machines, McGill University, Montréal, Québec, Canada, 1993. *Submitted to PAMI, March 1994*. Available via ftp at `ftp.cim.mcgill.ca` in `/pub/3d/papers/tr-cim-93-17.ps.gz`.

[23] P. Whaite and F. P. Ferrie. Autonomous exploration: Driven by uncertainty. In *Proceedings, Conference on Computer Vision and Pattern Recognition*, pages 339–346, Seattle, Washington, June 21-23 1994. Computer Society of the IEEE, IEEE Computer Society Press.

[24] S. Zucker, C. David, A. Dobbins, and L. Iverson. The Organization of Curve Detection: Coarse Tangent Fields and Fine Spline Coverings. In *Proceedings, 2ND International Conference on Computer Vision*, Tampa,Florida,USA, Dec. 1988. IEEE Computer Society, IEEE Computer Society Press.

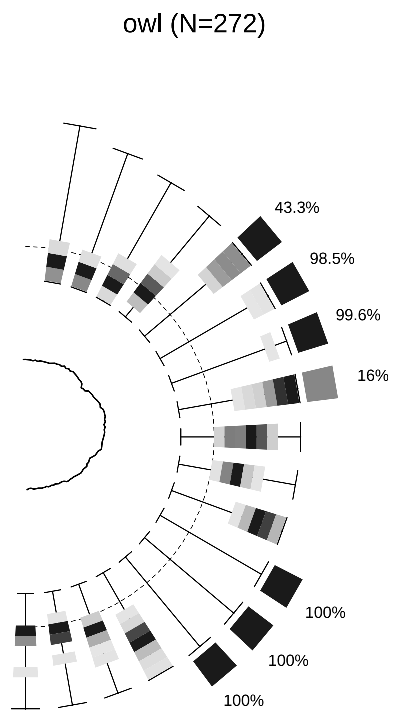


FIGURE 13. L3 misfit histograms for real data obtained from the owl. The scale of the histogram has been expanded to show the pattern of variation when the misfit is very high. In this case the confidence interval is 99.99999% so only 1 in 10^7 trials should fall in the top bin due to random chance.