Informative Views and Sequential Recognition

Tal Arbel and Frank P. Ferrie

TR-CIM-95-10 Nov. 1995

3D Vision Group
Centre for Intelligent Machines
McGill University
Montréal, Québec, Canada

Submitted to *Fourth European Conference on Computer Vision* and *IEEE Conference on Computer Vision and Pattern Recognition*, Oct. 1995

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

Informative Views and Sequential Recognition

Tal Arbel and Frank P. Ferrie

Abstract

In this paper we introduce a method for distinguishing between informative and uninformative viewpoints as they pertain to an active observer seeking to identify an object in a known environment. The method is based on a generalized inverse theory using a probabilistic framework where assertions are represented by conditional probability density functions. Consequently, the method also permits the assessment of the beliefs associated with a set of assertions based on data acquired from a particular viewpoint. The importance of this result is that it provides a basis by which an external agent can assess the quality of the information from a particular viewpoint, and make informed decisions as to what action to take using the data at hand.

To illustrate the theory we show how the characteristics of belief distributions can be exploited in a model-based recognition problem, where the task is to identify an unknown model from a database of known objects on the basis of parameter estimates. This leads to a sequential recognition strategy in which evidence is accumulated over successive viewpoints (at the level of the belief distribution) until a definitive assertion can be made. Experimental results are presented showing how the resulting algorithms can be used to distinguish between informative and uninformative viewpoints, rank a sequence of images on the basis of their information (e.g. to generate a set of characteristic views), and sequentially identify an unknown object.

Résumé

Dans cet article, nous pr'esentons une m'ethode qui permet de distinguer les points de vue informatifs et non-informatifs d'un objet tels que perçus par un observateur actif qui cherche à identifier un objet dans un environnement connu. La m'ethode repose sur une généralisation de la théorie inverse utlisée en probabilité où les hypothèses sont représentés par des fonctions de densité de probabilité. Conséquemment, la méthode permet aussi l'estimation de la confiance associée à un ensemble d'hypothèses basés sur les données obtenues d'un certain point de vue. L'importance de ce résultat est qu'il procure une base par laquelle un agent externe peut estimer la qualité de l'information provenant d'un point de vue et en conséquence prendre une décision éclairée quant à l'action à réaliser.

Pour illustrer la théorie, nous montrons comment les caractéristiques des fonctions de distribution de la confiance peuvent être exploitées dans le cadre d'un problème de reconnaissance basée sur un modèle. La tâche consiste à identifier un modèle à partir d'une base de données d'objets représentés paramétriquement. Ceci débouche sur une stratégie de reconnaissance séquentielle par laquelle les évidences sont accumulées sur plusieurs vues (au niveau des distributions de confiance) jusqu'à ce qu'une hypothèse définitive puisse être établie. Des résultats expérimentaux démontrent comment l'algorithme peut être utilisé pour: distinguer entre les vues informatives et non-informatives, classer une séquence d'images sur la base de leur information (i.e. pour générer un ensemble de vues caractéristiques) et identifier séquentiellement un object inconnu.

1. Introduction

Consider an active agent charged with the task of roaming the environment in search of some particular object. It has an idea of what it is looking for, at least at some generic level, but resources are limited so it must act purposefully when carrying out its task [1]. In particular, the agent needs to assess what it sees and quickly determine whether or not the information is useful so that it can evolve alternate strategies (the next place to look for example). Key to this requirement is the ability to make and quantify assertions while taking into account prior expectations about the environment. In this paper we show how the problem be cast in probabilistic terms from the point of view of inverse theory [25]. Assertions are represented by conditional probability density functions, which we refer to as *belief distributions*, that relate the likelihood of a particular hypothesis given a set of measurements. What is particularly important about the methodology is that it yields a precise recipe for generating these distributions, taking into account the different sources of uncertainty that enter into the process. Based on this result we show how the resulting distributions can be used to (i) assess the quality of a viewpoint on the basis of the assertions it generates and (ii) sequentially recognize an unknown object by accumulating evidence at the probabilistic level.

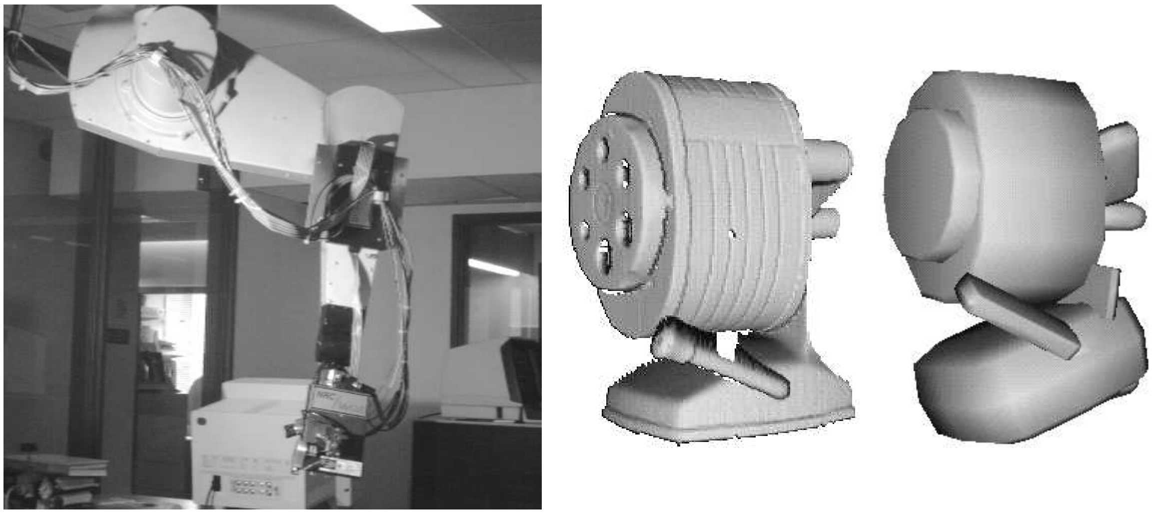
Specifically, we show how uncertainty conditions prior expectations such that the shape of the resulting belief distribution can vary greatly, becoming very delta-like as the interpretation tends towards certainty. In contrast, ambiguous or poor interpretations consistently tend towards very broad or flat distributions [4]. We exploit this characteristic to define the notion of an *informative viewpoint*, i.e. a view which gives rise to assertions that have a high probability according to their associated belief distribution. There are at least two applications for this result. First, in the case of an active observer, viewpoints can be chosen so as to maximize the distribution associated with an object of interest. This does not specify *how* to choose an informative viewpoint[1], but can be used as a figure of merit for a particular choice. Second, in the case of an off-line planner, it is often advantageous to be able to pre-compute a set of characteristic views to aid in recognition [15, 16, 23, 11, 12, 17, 8]. A good strategy here would be to select the n best views of an object ranked according to its belief distribution.

Second, although viewpoints can be labelled as either informative or uninformative, ambiguous cases where there is "reasonable" belief in more than one interpretation still exist. It becomes apparent that evidence from more than one viewpoint is

[1] Strategies for gaze planning are operationally defined [27, 29].

needed. This leads to a sequential recognition strategy that seeks to improve uncertain interpretations by accumulating evidence over several views. But at what level of representation should this evidence be accumulated? The autonomous exploration procedure that we use to generate the set of database models, for example, sequentially constructs a complete 3-D representation at the level of surface geometry [29]. One could follow a similar approach at the recognition phase, i.e. recalculate the belief in each hypothesis as the explorer adds new data to its representation of the unknown object. Unfortunately this would be computationally prohibitive, largely due to the expense of data fusion [22]. A better approach would be to process each view independently and avoid the fusion problem at the data level by seeking instead to combine information at the level of the belief. An active agent would then be able to gather evidence until the composite belief associated with a particular hypothesis exceeds a prescribed figure of merit. We show that such evidence can be accumulated by histogramming votes from each viewpoint and picking the hypothesis with the highest score. This strategy is appropriate provided that clear "winner" hypotheses prevail in a largely view-invariant manner.

This brings us to the problem of obtaining the belief distributions. Here we consider the recognition problem itself, focusing on a model-based approach. Specifically, model-based recognition focuses on matching an unknown model, which is computed on-line from sensory data, with a predetermined model computed off-line and residing in a database of known objects [5]. What differentiates approaches is largely a matter of the kinds of models used to represent objects in the scene and how models are matched. Our interest is in three-dimensional object recognition in which objects are represented by parametric shape descriptors (i.e. models) such as superellipsoids [7, 6, 21, 13], deformable solids [10, 20], and algebraic surfaces [24]. In our context, models are constructed through a process of *autonomous exploration* [27, 28, 29] in which a part-oriented, articulated description of an object is inferred through successive probes with a laser range-finding system. Figure 1a shows the set-up used to perform experiments — a two-axis laser range-finder mounted on the end-effector of an inverted PUMA-560 manipulator. For any particular viewpoint, such as the one shown in Figure 1b, a process of bottom-up shape analysis leads to an articulated model of the object's shape (Figure 1c) in which each part is represented by a superellipsoid primitive [13]. Associated with each primitive is a covariance matrix $\mathbf{C}$ which embeds the uncertainty of this representation and which can be used to plan subsequent gaze positions where additional data can be acquired to reduce this uncertainty further [27, 28]. A system which automatically builds object models based on this principle is reported in [29, 18].



(a) (b) (c)

FIGURE 1. (a) Mobile laser range-finding system used to construct object models. (b) Laser range-finder image of a pencil sharpener rendered as a shaded image. (c) An articulated, part-oriented model of the sharpener using superellipsoid primitives; 8 superellipsoids are used, one corresponding to each of the parts of the object.

Many approaches have been advocated for the problem of *matching* models. The majority of these employ various metrics to measure the distance between models in the appropriate parameter spaces, e.g., Mahalanobis distance [14], dot product [20] to mention but a few. These strategies rarely include both the uncertainties in the parameters of the measured models and the ambiguities of the representations in the database. However, when fitting a model to data that are noisy, there is an inherent lack of uniqueness in the parameters that describe the model. In these cases it is impossible to make a definitive statement as to which model fits the data best [27]. For this reason, rather than choose external constraints that would force the choice of one model over another, it would be more instructive to embed the uncertainty in the chosen description into the representation. This is precisely the approach that we have taken in computing the belief distribution.

Our methodology is based on a probabilistic inverse theory first introduced by Tarantola in [25]. Earlier work has shown how this theory can be used to methodically synthesize belief distributions corresponding to each model hypothesis, $\mathcal{H}_i$, given the parameters corresponding to the unknown model, $\mathcal{M}$, computed from the current measurement D_j, i.e. $P(\mathcal{H}_i|\mathcal{M}_{D_j})$ [4, 3]. This procedure explicitly accounts for uncertainties arising from the estimation of the unknown model parameters, database model parameters, and prior expectations on the frequency of occurrence for each of the database entries. In this case, the solution reduces to the classical Bayesian solution, similar to the result obtained by Subrahmonia et al. [24] - the primary difference being in the techniques used to obtain the solution. The inverse solution forces all sources of knowledge to be made explicit prior to the experiment. The method provides a more general recipe for combining information in a formal and structured fashion. In addition, they (and many others [5, 9]) are interested in constructing a discriminant that makes an absolute identification of the measured object. We argue that making assessments about identity from single measurements can be erroneous. We are more interested in assessing the *quality* of the identification from a particular viewpoint and to communicate this belief to other processes to determine whether further sampling is required.

The sequential recognition strategy therefore seeks to combine information at the level of the belief distribution. That is, given two data sets D_j and D_{j+1} corresponding to different viewpoints we seek a conjunction of $P(\mathcal{H}_i|\mathcal{M}_{D_j})$ and $P(\mathcal{H}_i|\mathcal{M}_{D_{j+1}})$ that is equivalent to $P(\mathcal{H}_i|\mathcal{M}_{D_j+D_{j+1}})$. Although the theory formally defines conjunction, such an operation requires knowing how a change in viewpoint conditions the respective belief distributions. Later on, we will show that if the maximum likelihood hypothesis [2] is largely invariant over a sequence of trials, then a robust interpretation can be made by tabulating the votes for each one and picking the hypothesis with the highest score. We also show that this invariance can be maximized by using the structure of the belief distribution to filter out uninformative hypotheses.

The remainder of the paper is organized as follows. We begin in Section 2 by describing how to distinguish between informative and uninformative viewpoints. We then introduce the general inverse theory in Section 2.1 and explain how to apply the theory to the problem of recognizing parametric models. We then indicate how the theory can be used to label viewpoints as informative or uninformative in Section 2.2. Finally, the results lend themselves to an incremental recognition scheme described in Section 3. In Section 4, we describe a series of experiments illustrating the application of the method to characterize informative views within

[2]This refers to the hypothesis that the correct answer is the one with the highest belief.

the context of recognizing parts of multiple-part objects. As well, we perform a series of incremental recognition experiments that test the accumulation of evidence from sequential viewpoints. Finally, we conclude in Section 5 with a summary of the results and a pointer to future applications.

2. Determining Which Viewpoints are Informative

Prior to using information obtained from a particular viewpoint, an agent must be able to assess the quality of the information in order to determine its usefulness for the task at hand. Assertions that are made at each viewpoint must therefore be quantified based on prior expectations about the scene. In this section, we will illustrate how the problem can be cast into a probabilistic framework from the vantage point of the inverse theory. Assertions are made in the form of conditional probability density functions (*belief* distributions) which indicate the likelihood of the measurement matching each of the models in the database. The importance of the methodology is that it incorporates the various sources of uncertainty into the solution.

The information assessment can be cast into the context of model recognition. In [4, 3], it has been shown that recognition based on complete information produced very reliable results. However, complete information is not always available, and potentially expensive to acquire. Assessments can be made from a single-viewpoint, but the degree of reliability depends on the amount of information available. For example, some viewpoints capture enough of the unique characteristics of the object to sufficiently distinguish it from the others in the database. We will refer to these viewpoints as *informative viewpoints*. Other viewing positions, where it is impossible to say which object in the database the unknown is closest to, are called *uninformative viewpoints*. By determining if a viewpoint is informative or not, we can establish if further sampling is necessary to be able to recognize the object reliably.

In order to be able to determine whether the viewpoint was informative or not, the recognition engine should quantify the identification by producing a degree of confidence in the hypotheses, rather than establish an absolute identity for the unknown object. In this fashion, views with stronger hypotheses in terms of a significantly higher degree of confidence in one model than the others, can be considered informative. Viewpoints associated with low confidence levels in the hypotheses are considered uninformative. In the next sections, we will illustrate how the inverse theory can be used to generate confidence in various hypotheses, and illustrate how it can be used to distinguish between informative and uninformative viewpoints within the context of model-based object recognition.

2.1. The Inverse Problem Theory. In lieu of a single maximum likelihood solution, we seek a method that generates a measure of confidence in various hypotheses within the context of an object recognition problem. The recognition problem we wish to address requires us to infer from measurements of an unknown object that model which most closely represents it in a database of known objects. Like all inverse problems, the recognition problem is ill posed in that, i) several models can give rise to identical measurements and, ii) experimental uncertainty gives rise to uncertain measurements. As a result it is not possible to identify the unknown object uniquely. There are various ways of conditioning ill posed problems, but these all require strong, and often implicit, a priori assumptions about the nature of the world. As a result a method may work well only in specific cases and because of the hidden implicit nature of the conditioning assumptions, cannot be easily modified to work elsewhere.

For this reason we have adopted the very general inverse problem theory of Tarantola [25]. In it the sources of knowledge used to obtain inverse solutions are made explicit, so if conditioning is required, the necessary assumptions about that knowledge are apparent and can be examined to see if they are realistic. The theory uses probability density functions to represent the following sources of knowledge:

1. Knowledge given by a theory which describes the physical interaction between models $\mathbf{m}$ and measurements $\mathbf{d}$, denoted $\theta(\mathbf{d}, \mathbf{m})$,
2. Knowledge about the model from measurements, denoted $\rho_D(\mathbf{d})$.
3. Information from unspecified sources about the kinds of models which exist in the world (namely that there are a discrete number of them). We denote this knowledge $\rho_M(\mathbf{m})$. Knowledge like this is a powerful constraint and can be used to eliminate many of the unconstrained solutions.

Given this knowledge the theory tells us how it should be combined, but leaves any decision about its usefulness up to the tasks that require it. For example, when attempting to recognize objects we would ideally want the unknown model be identified correctly all the time. Because of experimental uncertainties this can never happen, and there is always the possibility that an object will be identified incorrectly. Only the task can know if the likelihood of errors is acceptable.

This raises the interesting question of what we should do if the level of errors is not acceptable. Because the sources of knowledge are explicit they are not only visible to the operational tasks, but are also potentially open to manipulation by them. In principal it should be possible for the task to condition or actively acquire the a priori knowledge required to make the solution acceptable. We have already demonstrated that what we call autonomous exploration functions well at the model building level [26, 29] and we now intend, with the aid of this theory, to incorporate feedback from the recognition task as well.

2.1.1. *The Inverse Solution.* The theory postulates that our knowledge about a set of parameters is described by a probability density function over the parameter space. This requires us to devise appropriate density functions in order to represent what we know about the world. The solution to the inverse problem then becomes quite straight forward — it is simply a matter of combining the sources of information. The logical operation of *conjunction* is appropriate, i.e. the solution to the inverse problem is given by the theory AND the measurements AND any a priori information about the models. Tarantola extends the notion of logical conjunction to define the conjunction of two states of information [25, pages 29–31]. With this definition we can therefore combine the information from the joint prior probability density function $\rho(\mathbf{d}, \mathbf{m})$ and the theoretical probability density function $\theta(\mathbf{d}, \mathbf{m})$ to get the a posteriori state of information

$$\sigma(\mathbf{d}, \mathbf{m}) = \frac{\rho(\mathbf{d}, \mathbf{m})\, \theta(\mathbf{d}, \mathbf{m})}{\mu(\mathbf{d}, \mathbf{m})} \tag{1}$$

where $\theta(\mathbf{d}, \mathbf{m}) = \theta(\mathbf{d}|\mathbf{m})\, \mu_M(\mathbf{m})$ and $\rho(\mathbf{d}, \mathbf{m}) = \rho_D(\mathbf{d})\, \rho_M(\mathbf{m})$ over the joint space $M \times D$, where M refers to the *model space* and D, the *data space*. The so called non-informative probability density $\mu(\mathbf{d}, \mathbf{m}) = \mu_D(\mathbf{d})\mu_M(\mathbf{m})$ represents the reference state of information in much the same way that noise is used when measuring information in terms of signal to noise ratios. The formulation of appropriate non-informative densities is a complex issue, but for our purposes we will assume that all the non-informative densities are uniform over their respective spaces.

Accordingly, (1) is more general that the equations obtained through traditional approaches, but degenerates to them in specific cases. Under the conditions mentioned, the solution is identical to the Bayesian solution [25, page 61] where the a posteriori information about the model parameters is given by the marginal probability density function:

$$\sigma(\mathbf{m}) = \rho_M(\mathbf{m}) \int_D \frac{\rho_D(\mathbf{d})\, \theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})}\, d\mathbf{d}. \tag{2}$$

2.1.2. *The Part Recognition Problem.* In the system we have constructed, range measurements are taken, surfaces are reconstructed, segmented into parts, and individual models are fit to each part. We will treat *the whole system as a measuring instrument.* Given some model $\mathbf{m}$ in the scene, range measurements are taken and from these an *estimate* of the model $\mathbf{d}$ is obtained, which we call a *measurement of the model* in the scene.

1. **Information Obtained from Physical Theories** We first formulate an appropriate distribution to represent what is known about the physical theory that predicts estimates of the model parameters given a model in the scene. Such a theory is too difficult to formulate mathematically given the complications of our system. We therefore build an empirical theory through a process called the *training* or learning stage. Here, Monte Carlo experiments are run on measures of a known model exactly as in traditional statistical pattern classification methods. The conditional probability density function $\theta(\mathbf{d}|\mathbf{m})$ is calculated for each model $\mathbf{m}$ by assuming a multivariate normal distribution. Therefore, the equation for $\theta(\mathbf{d}|\mathbf{m})$ is:

$$\theta(\mathbf{d}|\mathbf{m}) = N(\mathbf{d} - \mathbf{m}, \mathbf{C}_T) \tag{3}$$

   where N is the multivariate normal distribution, with a covariance matrix, $\mathbf{C}_T$, describing estimated modelling errors for a model $\mathbf{m}$.
2. **Information Obtained from Measurements**
   Much of the knowledge we have about a problem comes in the form of experimental measurements. In our system [29], we obtain an estimate of the observed model parameters $\mathbf{d}_{obs}$, and also an estimate of their uncertainty in the covariance operator $\mathbf{C}_d$. The assumption we make is that the multivariate normal distribution $N(\mathbf{d} - \mathbf{d}_{obs}, \mathbf{C}_d)$ represents our knowledge of the measurements. The probability density function representing this information is the conditional probability density function $\nu(\mathbf{d}_{obs}|\mathbf{d})$, such that:

$$\nu(\mathbf{d}_{obs}|\mathbf{d}) = \rho_D(\mathbf{d})/\mu_D(\mathbf{d}) = N(\mathbf{d} - \mathbf{d}_{obs}, \mathbf{C}_d) \tag{4}$$

3. **A Priori Information on Model Parameters**
   In the current context, there are a discrete number of reference models, $\mathbf{m}_i, i = 1 \ldots M$. The probability density function used to convey this knowledge is

$$\rho_M(\mathbf{m}) = \sum_i P(\mathbf{m}_i)\, \delta(\mathbf{m} - \mathbf{m}_i), \tag{5}$$

   where $P(\mathbf{m}_i)$ is the a priori probability that the i^{th} model occurs.
4. **Solution to the Inverse Problem**

Substituting the probability density functions (3), (4), and (5) into (2) gives us the final equation for the a posteriori probability density function

$$\sigma(\mathbf{m}) = \sum_i P(\mathbf{m}_i) N(\mathbf{d}_{obs} - \mathbf{m}_i, \mathbf{C}_D)\ \delta(\mathbf{m} - \mathbf{m}_i). \tag{6}$$

where $\mathbf{C}_D = \mathbf{C}_d + \mathbf{C}_T$. This density function is comprised of one delta function for each model in the database. Each delta function is weighted by the *belief* $P(\mathbf{m}_i)N(\mathbf{d}_{obs} - \mathbf{m}_i, \mathbf{C}_D)$ in the model $\mathbf{m}_i$. The final distribution represents the "state of knowledge" of the parameters of $\mathbf{m}_i$. The beliefs in each of the reference models are computed by convolving the normal distributions in (3) and (4). The advantage of the method is that rather than establish a final decision as to the exact identity of the unidentified object, it communicates the degree of confidence in assigning the object to each of the model classes. It is then up to the interpreter to decide what may be inferred from the resulting distribution.

The methodology introduced applies to the recognition of any parametric primitive. For our purposes, superellipsoid models were chosen because of the range of shapes they can represent as well as their computational simplicity. However, representations based on superquadrics pose a number of problems due to degeneracies in shape and orientation. In solving this problem, work has being done in representing objects by multi-modal distributions, where each mode contains information about a possible equivalent form. Discussion of this process is beyond the scope of this paper.

2.1.3. *Example – Recognizing a Sphere.* Figure 2 illustrates the kinds of results we get by applying the theory to a typical recognition problem. Here, the reference models were produced by training on models created with data acquired by scanning the objects all around their surfaces (i.e. complete 3D data). The reference models, consisting of a smaller sphere, a large sphere, and a lemon, can be seen in Figure 2a. The larger sphere was then measured from a single viewpoint, and the resulting model is shown in Figure 2b. The system's ability to distinguish the larger sphere from both the smaller sphere and the lemon was then tested. The result is the belief distribution found in Figure 2c. One can see that the system has a significantly higher degree of confidence in the hypothesis that the measured model was a large sphere.

2.2. Determining Which Viewpoints are Informative using the Inverse Theory. An important result of the inverse solution is that rather than establish an absolute identity for the unknown object, the system produces a belief in each of the models in the database. We can then use the belief distribution to tell us whether a particular viewpoint is informative or not. This is easily accomplished by denoting

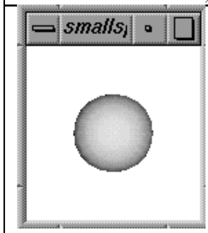
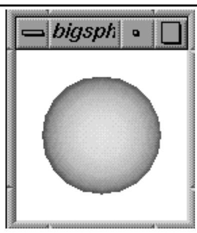
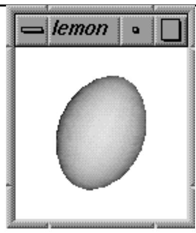
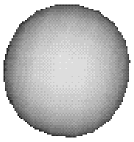
	a) Reference Models		
			
b) Measured Model	c) Beliefs in Reference Models		
	6.12×10^{-43}	0.00273	0

FIGURE 2. Recognizing a sphere. (a) The reference models are: a smaller sphere, a larger sphere, and a lemon. (b) The measured unknown model. (c) The belief distribution.

views with a clear winner, in terms of a significantly higher belief in one model than the others, as informative views. From these positions, the system is able to capture the attributes of the model that distinguish it from the others.

We can also use the beliefs to label a viewpoint as uninformative. This would be an indication that results from the current viewing position do not tell us much about the object's identity. Using the inverse solution, one would detect an uninformative viewpoint when the unnormalized belief in each of the models is very low (or zero). In this situation, it is impossible to say which reference model the unknown might correspond to. Such a scenario would arise if the distribution of the unknown model does not significantly overlap with any of the reference distributions. There are two possible reasons for this to occur. The first is the case where the distribution of the measured model is very wide due to large uncertainties in its parameters. The result is low beliefs in all the reference models in the database. This case occurs when the scanning has occurred from a viewpoint where insufficient data were collected. The second case occurs when there is a breakdown in some of the prior assumptions. In this case, the issue is not one of insufficient data. Here, the parameters determined from that particular viewpoint differ significantly from any of the models in the database. The resulting distribution could actually be quite sharp, but simply does not overlap with any of the reference model distributions. In this case, it could be that the linearity assumption breaks down, implying that perhaps the assumption of a normal distribution is not valid. Another possible breakdown of assumptions occurs when the chosen model does not adequately describe the data. Zero belief cases exist when the values of the a posteriori probability density functions are extremely low. Due to numerical underflow, the procedure produces beliefs of zero for each of the reference models.

Figure 3 illustrates the difference between informative and uninformative viewpoints for the case of a cylinder. Here, one can see that the system is able to distinguish the cylinder from a block with great ease, if the cylinder is measured from an informative viewpoint. However, if measured from an uninformative viewpoint, there is little confidence in either model. In this case, the beliefs are in fact below the numerical precision of the system, and therefore become zeros.

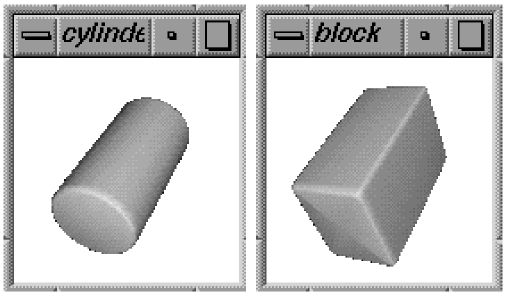
The problem of distinguishing between the two kinds of states becomes one of determining the threshold, below which one can safely state that the beliefs are in fact insignificant. It is obvious that cases where the beliefs in all the models are zero are uninformative. However, this threshold depends on the numerical precision of the system. In this sense, it is chosen externally (and is, therefore, a random cutoff point). We therefore feel justified in raising this threshold to one that excludes other low confidence states. The expectation is that this will eliminate false positive states, as they are generally occur with low belief. One can determine this cutoff point empirically, by observing the belief distributions from different viewpoints, and noting if there is a clear division between the clear winner states and the low confidence states. A bi-modal distribution would indicate that an application of a predefined threshold can easily distinguish between these states.

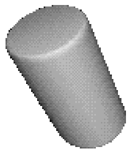
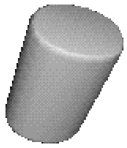
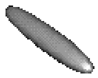
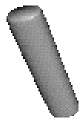
Figure 4 illustrates the logarithm of the beliefs resulting from recognizing 36 different single-view samples of each of six models in a database: a Big Sphere, a Block, a Cylinder, a Lemon, a Small Sphere and a Round Block. The results indicate the bi-modality of the belief distribution. In Section 4, we illustrate the bi-modality of the results of recognizing the parts of articulated models.

3. Sequential Recognition

Provided that the low belief states have been identified, we wish to make a statement about the remaining beliefs. Even though the majority of the cases can be clearly divided into informative and uninformative states, there are still ambiguous cases where a "significant" belief in more than one model exists. Because of these situations, it becomes apparent that evidence from more than one viewpoint is needed.

Database Models

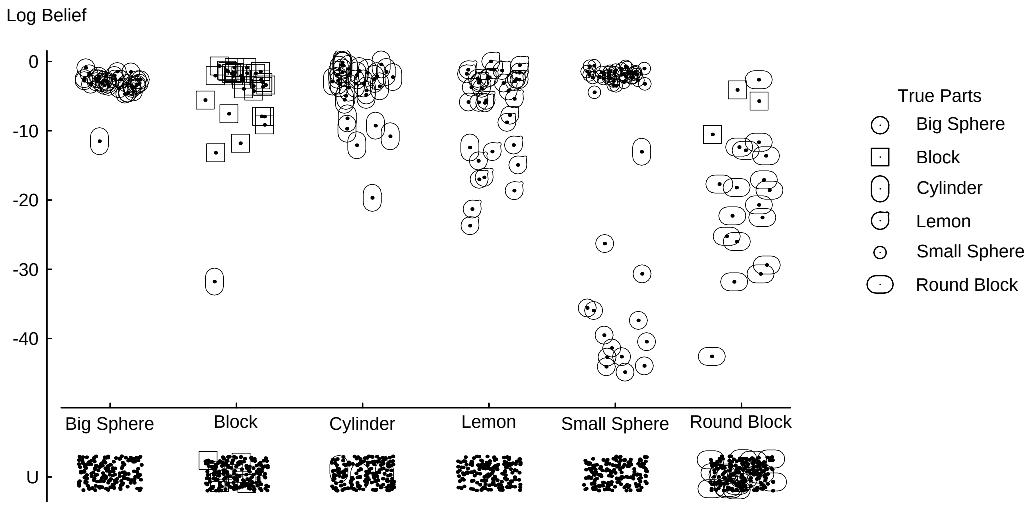


Measured Model	View 1	View 2	View 3	View 4
				
Belief in cylinder	2.237	0.009181	0.0	0.0
Belief in block	0.0	0.0	0.0	0.0
	a) Informative		b) Uninformative	

At the top of this figure are the two reference models in the data base: the cylinder and the square block. Beneath these are measured models of the *cylinder* obtained after scanning its surface from 4 different viewing positions. Below each model one can find the unnormalized belief distributions obtained when attempting to recognize each of the measured models.

FIGURE 3. (a) Informative and (b) Uninformative Views of a Cylinder.

The question becomes: how do we accumulate evidence from different views, when the evidence is in the form of a conditional probability density function? The immediate response is given by the theory (Section 2.1) which formally defines the operation of conjunction of information, i.e. the belief distributions. To state this more formally, we denote belief distributions corresponding to each model hypothesis, $\mathcal{H}_i$, given the parameters of the unknown model, $\mathcal{M}$, computed from the measurement, D_j, by $P(\mathcal{H}_i|\mathcal{M}_{D_j})$. Then, given two data sets D_j and D_{j+1} corresponding to different



Above are the results from attempting to recognize 36 different single-view samples of each of the models in the database. The beliefs in the different models are represented by different symbols, each symbol indicating the true model used during that trial.

The level of numerical underflow of the system is represented by a "U" on the $y - axis$. Because so many trials fall into this category they are marked with a simple point, *except* when the belief is for the true model used in the trial.

By observing the log of the beliefs, one can see the bi-modality in the results.

FIGURE 4. Log of beliefs in the Big Sphere, Block, Cylinder, Lemon, Small Sphere, and Round Block.

viewpoints we seek a conjunction of $P(\mathcal{H}_i|\mathcal{M}_{D_j})$ and $P(\mathcal{H}_i|\mathcal{M}_{D_{j+1}})$ that is equivalent to $P(\mathcal{H}_i|\mathcal{M}_{D_j+D_{j+1}})$. An active agent would then gather sufficient evidence in this fashion until the composite belief distribution associated with a particular hypothesis exceeds a predefined level of acceptability.

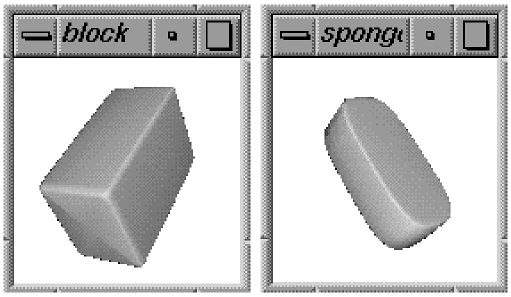
Although the theory formally defines conjunction, such an operation requires knowing how a change in viewpoint conditions the respective belief distributions, as they are normalized with respect to a global frame of reference. As a result, relative values between the views are meaningless. The normalizing factor is some unknown function of viewpoint, and is difficult to obtain analytically. In [2], it is indicated that the reason for the difficulty lies in that this factor can be shown to be a function of $\int_D d\mathbf{d}$, the volume of data space. The issue of how to define this space is a difficult one to address. In order to do so, a commitment to a permissible region of observed parameters must be established *prior* to experimentation. In this sense, this problem is analogous to finding correspondence between range data from different viewpoints where the relationship between the views is unknown. Because of the difficulty of finding this factor, the beliefs are not normalized. As a result, it becomes difficult to match a belief of 500, for example, from one view, with a value of 50 from another. Each of these values may reflect the strongest possible belief from their respective views, however it is difficult to compare them in a sensible fashion. As well, in situations where there is a belief of 50 in one model and 40 in another, it becomes impossible to establish a clear winner.

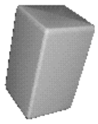
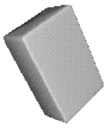
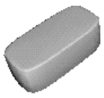
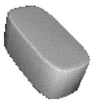
For this reason, we have chosen not to select a "winner" in ambiguous situations, and state that all beliefs above a threshold indicate equally likely hypotheses. We illustrate this philosophy by binarizing the conditional probability density function values at each view, such that all beliefs above the threshold become ones. In this fashion, we have divided the possible results to include:

1. *Informative states*: states with one clear winner (a single positive value).
2. *Uninformative states*: states without a clear winner. These include:
   a) *Ambiguous states*: states with more than one possible winner (more than one single positive value).
   b) *Undetermined states*: states with no winners (all zero values).

It is important to note that ambiguous states are, in fact, undetermined states that lie above the chosen threshold. In theory, careful choice of cutoff level should eliminate these states as well (without eliminating a large number of informative states). Figure 5 illustrates these different states in the case of a square block. Here, the system is asked to identify a square block from different views, and correctly distinguish it from a similar rounder one. This example indicates that the results match human intuition. The clear winners, or informative states, in Figure 5a indicate that the system is able to identify the block despite wide variations in its three dimensions. The ambiguous cases (Figure 5b) occur when the resulting models are rounder in shape. Here, the system has trouble differentiating between the models. In fact, these models resemble the rounded block more than the square one. In the third case (Figure 5c), the system does not have significant belief in any of the models. Intuitively, one can see that these models are not similar to either reference model.

Using this method of representation, rather than base conclusions on maximum likelihood methods from independent viewpoints, the methodology would entail combining evidence from single viewpoints by considering all models whose beliefs are above a threshold to be equally significant. In accordance with Marr's "Principle of Least Commitment" [19], all plausible hypotheses, rather than just one are communicated to the external processes. Furthermore, should the case with a clear winner



Measured Model	Belief in Block		Belief in Round Block	
	Unnormalized	Binarized	Unnormalized	Binarized
a) Informative				
	0.2	1	0	0
	0.007	1	0	0
b) Ambiguous				
	$2.0{\times}10^{-13}$	1	$5.8{\times}10^{-6}$	1
	$3.4{\times}10^{-13}$	1	0.002	1
c) Undetermined				
	0	0	0	0
	0	0	0	0

Above are the two reference models: a block and a rounded block. In the left column of the table are the models of the block measured from (a) informative, (b) ambiguous and (c) undetermined viewpoints. To their right, one can find the unnormalized, and binarized belief distributions (a threshold of 10^{-13}) obtained when attempting to recognize each of the measured models.

FIGURE 5. Informative, Ambiguous, and Undetermined States for the Block.

persist with a significantly higher belief in one model than the rest (referred to as a *bi-modal distribution* - where there is a clear separation between the high confidence and low confidence beliefs), then binarization becomes a perfectly valid method of normalizing the beliefs from one view to the next.

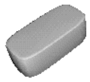
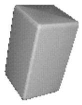
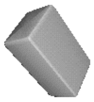
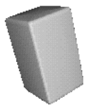
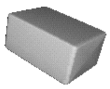
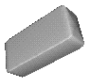
By normalizing our confidence values in this manner, combining them from different viewpoints becomes straightforward. Should the maximum likelihood hypothesis prevail in a largely view-invariant manner, then after a sequence of trials, a robust interpretation can be made by tabulating the votes for each one, represented by the binarized beliefs, and picking the hypothesis with the highest score. In this fashion, a clear winner should emerge. In addition, the confidence in the incorrect models should become insignificant. In Section 4, we will verify this empirically by attempting to recognize a series of real objects from sequential viewpoints. We will also show that the view-invariance is maximized by applying a threshold to filter out the uninformative viewpoints.

Figure 6 illustrates an attempt at sequentially recognizing the square block at $40°$ increments. As in the previous example, the square and round blocks are used as reference models. The raw beliefs are binarized by imposing a threshold of 10^{-13}. Notice that the ambiguous case quickly becomes insignificant with the increase of evidence in the correct model. After only 9 iterations, the clear winner emerges, casting all doubt aside.

4. Experiments and Results

Through a series of experiments, we wish to illustrate how one can use the belief distributions to distinguish between informative and uninformative viewpoints. As well, we wish to show that evidence, in the form of the belief distributions, can be accumulated from sequential viewpoints. The expectation is that a clear winner will emerge rapidly after a small number of views.

4.1. Characterizing Informative Viewpoints by External Threshold. The first question that must be addressed through experimentation is: Can we distinguish between informative and uninformative viewpoints? In previous sections, we have hypothesized that the beliefs are differentiated into two distinct states: the state of high confidence in the correct model and the state of low confidence in the others. Through experimentation on a set of real objects, we will examine the idea that should the distribution be bi-modal, the application of a threshold to distinguish between informative and uninformative states becomes justified.

View Angle	Measured Model	Belief in Block		Belief in Round Block	
		Unnormalized	Binarized	Unnormalized	Binarized
0°		2.0×10^{-13}	1	5.8×10^{-6}	1
40°		0	0	0	0
80°		0.2	1	0	0
120°		0.03	1	0	0
160°		0	0	0	0
200°		0.1	1	0	0
240°		0	0	0	0
280°		0.03	1	0	0
320°		0.001	1	0	0
		Final Score	6		1

Displayed above are the 9 models resulting from sequentially measuring the square block at 40° increments. From left to right, one can see the viewing angle, the measured model, the unnormalized and binarized (threshold of 10^{-13}) belief distribution resulting from attempting to recognize each of the measured models. The final distribution is the histogram of the binarized distributions.

FIGURE 6. Incremental Recognition of a Block.

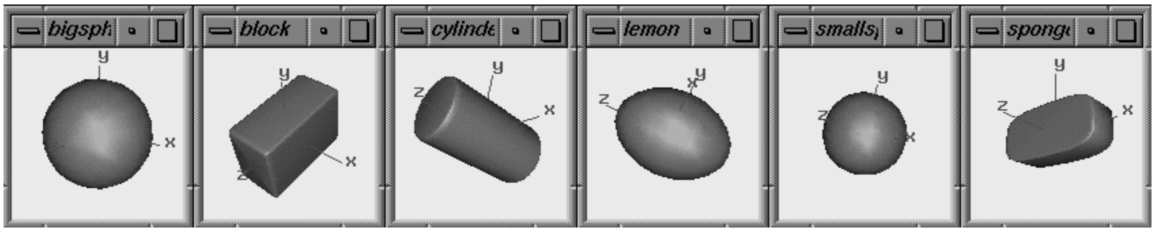
Our current focus is "recognition by parts", whereby measured objects are segmented into their constituent parts, each of which is compared to the parts in the database. Here, objects are seen as collections of independent parts where topological relationships are not yet considered.[3] Recognizing parts of articulated models is challenging due to problems of self-occlusion and segmentation.

Two articulated models were used in the recognition experiments: a potato-head toy consisting of two ears, two eyes, a nose and a head, and an alarm clock with two bells, two legs, a cylindrical face and a back were chosen for the purposes of testing the part recognition algorithm on complex objects. In addition, six single-part "distractors" were placed in the database in order to render the recognition task more difficult. These objects consisted of: two spheres ($\text{rad} = 20mm, \text{rad} = 25mm$), a block, a cylinder, a lemon, and a block with rounded edges. The objects were chosen for the experiments because they consisted of parts that generally conformed well to non-deformable superellipsoids, with the exception of the toy head whose shape was tapered. The parts varied in size and shape, so as not to be clustered together too tightly in five-dimensional feature space. However, their distributions overlapped sufficiently enough in several dimensions so that the recognition procedure was challenged in its discrimination task.

Training automatically produced object class representatives, by measuring the object numerous times. Each individual model was created by scanning the object from several views in an exploration sequence [27, 28, 29]. Here, each object was scanned using a laser range-finder, segmented into its constituent parts, a superellipsoid model was fit to each part, and the resulting parameters stored. In order to create the representatives in the database, 24 samples of each single-part object, 10 samples of the potato-head, and 7 samples of the alarm clock were used. Figure 7 illustrates the actual potato-head and alarm used in recognition experiments, and the representative models of each object that result from training.

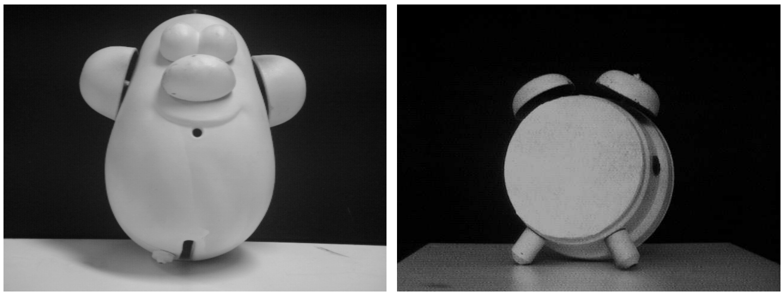
In the first experiment devised, the potato-head toy was measured from 32 independent viewing positions. Recognition was performed on each of these samples in turn, using a database consisting of the parts of the potato-head as well as the single-part objects as distractors. The beliefs were plotted on a logarithmic scale graph. Once again, a bi-modal distribution was anticipated, whereby a clear division between the informative and uninformative states would permit the use of a threshold to distinguish between the two. The results can be found in Figure 8.

[3] Recognition strategies that take topology into account are currently being investigated.

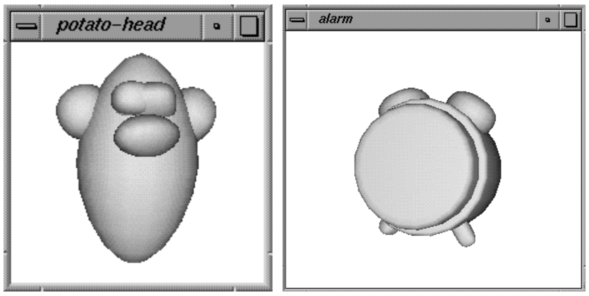


BS B C L SS RB

Displayed above are reference objects that result from training acting as "distractors" for the recognition procedure: a big sphere (BS), a block (B), a cylinder (C), a lemon (L), a smaller sphere (SS), and a rounded block (RB).



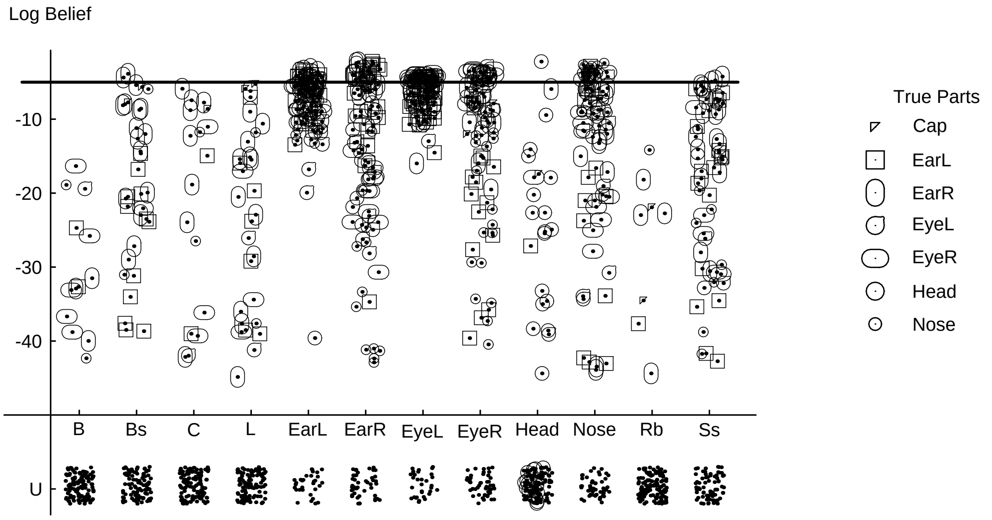
a) Original potato-head and alarm clock.



b) Reference potato-head and alarm clock models created by training.

FIGURE 7. The reference parts resulting from training.

As hypothesized, the results indicate a bi-modal distribution for the beliefs in the potato-head parts. For each of these parts, there is a top cluster, representing relatively high beliefs in the correct models. Beneath this, a thin scatter of beliefs in other models can be seen. Finally, the bottom cluster represents those beliefs that were below the numerical precision of the system (producing zero beliefs). However,



Above are the results from attempting to recognize 32 different single-view samples of each of the parts of the potato-head: the Left Ear (EarL), Right Ear (EarR), Left Eye (EyeL), Right Eye (EyeR), Head (Head), and Nose (Nose). The single-part reference models were also included as distractors for the recognition process. (see Figure 7 for notation.)

One can see the bi-modality in the log of the beliefs in the potato-head models. The beliefs in the distractors appear much more scattered, the majority lying beneath the top cluster of the potato-head parts. The top horizontal line indicates the results that can be achieved if a threshold of 10^{-5} were applied. This would lead to minimal false-positive indications accompanying a high number of correct votes.

FIGURE 8. Log of beliefs in the Potato-Head parts, as well as the Big Sphere, Block, Cylinder, Lemon, Small Sphere, and Round Block.

the majority of the beliefs were concentrated in the the top cluster illustrating that, most of the time, the system had high confidence in the correct part. However, some scattered beliefs in the single-part distractors occurred as well. It is important to note that the majority of these cases lie below the top cluster of correct identifications, indicating that by application of a threshold anywhere from 10^{-10} to 10^{-5} should eliminate the majority of the false-positive cases. The exact value of the cutoff level is not critical. The bold horizontal line illustrates the what can be achieved by applying a threshold of 10^{-5}. The result should lead to minimal false-positive indications accompanying a high number of correct votes. The case of the head, however, emphasizes the possibility of individual threshold levels for maximal efficiency. Here, a much lower threshold would ensure the highest number of correct matches.

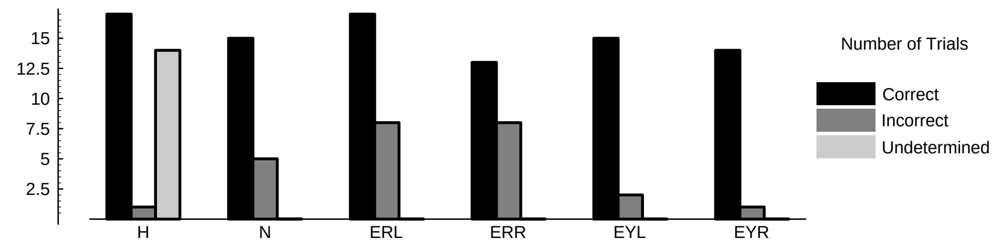
Having established the bi-modality of the belief distribution, the next step is to investigate the application of an external threshold to distinguish between informative and uninformative viewpoints, and consequently, remove the majority of false-positive assertions. To this end, the cutoff point was raised to 0.00001. The results of using maximum likelihood on the beliefs, before and after applying the threshold, can be seen in Figure 9.

Figure 9a illustrates that the system is able to successfully recognize instances of articulated parts of a complex object with only partial information available, even with the added effects of self-occlusion. In fact, even with "complete" data gathered from all around the object surface, most parts were embedded within others and thus part of their surfaces were never visible. This resulted in reference parts being created without complete information. As a result, the training procedure was not able to ensure that model parameters were close to their "true" values.

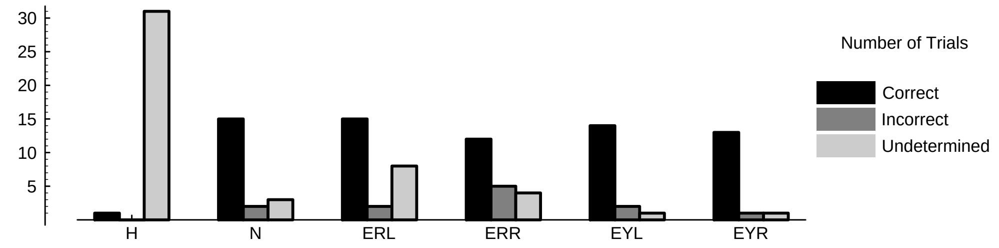
Figure 9b indicates that, for most models, the external threshold retained most of the correct states, confirming that the system had high confidence in the correct identifications. In addition, the majority of the false-positive assertions were eliminated. This confirms the hypothesis that, because the beliefs are bi-modal in nature, the application of an external threshold can used to successfully distinguish between informative and uninformative viewpoints. An active observer can then assess these results from a particular viewpoint and determine if further sampling is necessary.

In the case of the head, however, application of the external threshold caused almost all of the correct identifications to become undetermined states. This is because the beliefs in the correct model were generally low to begin with. The problem is that the head is tapered, breaking the assumption that the objects *can* in fact be accurately modelled by non-deformable superellipsoids. Here, training produced one particular model, and different single-view samples of the head produced very different superellipsoids depending on where the data were collected from. Therefore other equally viable descriptions that result from single view measurements were not recognized correctly. This lead to undetermined states.

Other potential problems occur because the recognition process relies heavily on the accuracy of the segmentation process. Because of this, errors in the segmentation of the range data can lead to errors in recognition. In these experiments, there were several cases where the head was divided into two distinct parts: a "head" and a "cap". Because the database allowed for only one part for the head, the system identified the cap part as being as lemon or some other reference model. This was understandable as the cap was similar in size and shape to these models.



a) Threshold = computational underflow



b) Threshold = 0.00001

Displayed above are the tables describing the belief distributions of the potato-head measured from single view-points. The parts of the potato-head are: a head (H), a nose (N), a left ear (ERL), a right ear (ERR), a left eye (EYL), and a right eye (EYR). Here, labelling one eye as the other, or one ear as the other was considered to be a correct identification. Zero values are defined by a) numerical underflow of system and b) a threshold of 0.00001.

FIGURE 9. Matching samples of the potato-head taken from single viewpoints.

Most of the incorrect states arose due to the similarity of the reference models. For example, the eyes resembled the smaller sphere, the nose and the ears. Similarly, the ears were extremely close to the bigger sphere in size and shape. As a result, their distributions overlapped significantly, making it difficult to distinguish between them using the current modelling scheme. Yet, in the majority of cases, these incorrect identifications occurred with low beliefs. These uninformative viewpoints were identified and eliminated by raising the threshold for undetermined states.

4.2. Accumulating Evidence from Sequential Viewpoints. The experiments described suggest the possibility of an incremental recognition procedure. It is based in the following observations obtained empirically over successive trials:

i) Viewpoints that provide very little information, or *uninformative* views, generally can be detected by their low confidence levels (beliefs). Because of the bi-modality of the belief spread, these can be discovered by application of a threshold. Detection of such events is a clear indicator that further sampling is required.

ii) *Informative* views are generally accompanied by high beliefs, but with the possibility of a false-positive indication. These can also be detected by threshold application.

iii) The likelihood of successive false-positive indications is very small. First, this is a consequence of the high selectivity of the reference distributions which result in low frequencies of false-positive indications in the first place (e.g. Figure 9). Second, it is unusual for observer motion to result in similar viewpoints in two successive views (general position assumption).

In order to explore the possibility of an incremental recognition strategy for complex objects, two experiments were devised. In the first experiment, evidence from single-views of the potato-head toy were accumulated. In the second, the alarm clock parts were added to the database and evidence in the alarm clock was accumulated. In both experiments, the system binarized the beliefs above the predefined threshold at each view. Evidence at each stage was computed by histogramming the binarized beliefs accumulated thus far. The results of accumulating evidence in each object after 32 single-views can be seen in Tables 1 and 2. In Tables 1a and 2a, the cutoff point was determined by the numerical precision of the system. In Tables 1b and 2b, thresholds of 0.00001 and 0.0001 respectively were imposed externally.

Tables 1a and 2a indicate distributions from single-views that were relatively "wide" in that a measured model produced a degree of belief in several reference models at once. The result is that, in most cases, the accumulated binarized evidence points to several models at once. Attempting to choose a single winner after several iterations would therefore be a difficult task. The choice could be limited to a few candidates, as some of the false-positive indications have been rendered insignificant due to insufficient evidence. For example, in the case of recognizing the potato-head's nose, although evidence in the big sphere, block, cylinder, lemon and rounded block existed from particular viewpoints, it was significantly lower than the evidence in the nose after tabulating the votes over several views. This validates the hypothesis that if the belief in the true model was much higher than the beliefs in the other models, by raising the threshold to an appropriate value, one could eliminate the majority of the wrong hypotheses. Over several views, the result would be an overwhelming accumulation of evidence in the true models.

Tables 1b and 2b further validate the hypothesis by illustrating that the majority of the evidence in the incorrect models were removed after application of an external

	H	N	EAR	EYE	BS	B	C	L	SS	RB
H	17	0	1	0	0	0	0	0	0	0
N	0	20	20	20	2	2	2	2	10	1
ERL	1	15	25	25	12	2	2	7	18	1
ERR	1	15	21	21	16	3	8	13	20	4
EYL	1	16	17	17	0	1	1	2	4	0
EYR	1	15	15	15	1	5	3	3	5	0

a) Threshold = computational underflow

	H	N	EAR	EYE	BS	B	C	L	SS	RB
H	1	0	0	0	0	0	0	0	0	0
N	0	16	2	12	0	0	0	0	0	0
ERL	0	1	15	1	0	0	0	0	0	0
ERR	0	1	13	1	2	0	0	0	3	0
EYL	0	3	0	16	0	0	0	0	0	0
EYR	0	3	0	14	0	0	0	0	0	0

b) Threshold = 0.00001

Displayed above are the tables describing the accumulation of evidence from 32 single-view recognition experiments. Each row describes the histogram of the binarized belief distributions for a particular measured model. In this case, the measured models include the parts of the potato-head (see Figure 9 for notation.) The columns refer to the reference models, including the potato-head parts: the head (H), the ears (EAR), the eyes (EYE), and the single-part objects. Zero values are defined by a) numerical underflow of system and b) a threshold of 0.00001.

TABLE 1. Histogram of binarized belief distributions for the potato-head after 32 single-view iterations.

threshold. The exceptions to this rule are the potato-head's head and the face of the alarm clock, where the majority of the evidence in the correct model was eliminated as well. This indicates the possibility that the choice of threshold was not appropriate for these parts (see previous section for discussion). However, even with a uniform threshold, the results indicate that the correct assertion is obtained with the combination of a threshold to remove false assertions and the accumulation of information from a series of views to remove the ambiguous cases. In fact, if one

	F	BA	BELL	LEG	H	N	EAR	EYE	BS	B	C	L	SS	RB
F	20	12	23	9	14	10	12	10	6	5	6	7	7	4
BA	16	25	20	8	8	8	8	10	4	7	5	5	4	2
RBL	2	2	21	12	2	15	23	23	12	6	8	9	14	1
LBL	4	4	21	9	2	14	21	21	10	3	8	6	13	1
RL	2	2	22	22	0	15	22	22	1	6	0	0	3	2
LL	6	5	22	22	0	16	22	22	0	5	1	0	0	1

a) Threshold = computational underflow

	F	BA	BELL	LEG	H	N	EAR	EYE	BS	B	C	L	SS	RB
F	1	0	0	0	0	0	0	0	0	0	0	0	0	0
BA	0	6	0	0	0	0	0	0	0	0	0	0	0	0
RBL	0	0	4	0	0	1	2	1	0	0	1	0	3	0
LBL	0	0	6	0	0	3	0	1	0	0	1	0	2	0
RL	0	0	0	13	0	0	0	0	0	0	0	0	0	0
LL	0	0	0	15	0	0	0	0	0	0	0	0	0	0

b) Threshold = 0.0001

Displayed above are the tables describing the accumulation of evidence from 32 single-view recognition experiments. Each row describes the histogram of the binarized belief distributions for a particular measured model. In this case, the measured models include the parts of the alarm clock: the face (F), the back (BA), the right bell (RBL), the left bell (LBL) the right leg (RL), and the left leg (LL). The columns refer to the reference models, including the alarm clock parts: the face (F), the back (BA), the legs (LEG), and the bells (BELL), the potato-head parts and the single-part objects (see Table 1). Zero values are defined by a) numerical underflow of system and b) a threshold of 0.0001.

TABLE 2. Histogram of binarized belief distributions for the alarm clock after 32 single-view iterations.

were to choose a winner based on a maximum likelihood scheme of the accumulated evidence, the results would be correct for all models[4].

[4]We have treated the left and right eyes of the potato-head as being instances of the same class. A similar rule was applied to the left and right ears, as well as the left and right bells and the left and right legs of the alarm clock.

5. Discussion and Conclusions

In this paper, we have introduced a method for distinguishing between informative and uninformative viewpoints and for assessing the beliefs associated with a particular set of assertions based on this data. The importance of this result is that it provides a basis by which an external agent can assess the quality of the information from a particular viewpoint, and make informed decisions as to what action to take using the data at hand. Our approach was based on a generalized inverse theory [25] using a probabilistic framework where assertions are represented by conditional probability density functions (belief distributions). The importance of the method is that it provides a formal recipe for representing and combining all prior knowledge in order to obtain these distributions. We have illustrated how to apply the theory to solve a 3-D model-based recognition problem and have shown how the resulting belief distributions can be used to assess the quality of the interpretation. An important characteristic of the resulting belief distributions is that they are bi-modal, simplifying the problem of determining how to distinguish between informative and uninformative viewpoints.

We have also demonstrated that some viewpoints can give rise to ambiguous information, where the system has confidence in more than one hypothesis. Similar to the motivation behind *autonomous exploration* in the model-building phase [29], ambiguous views have spawned the development of an incremental recognition scheme, where we seek information from a new viewpoint to reduce the overall ambiguity. We have shown how evidence, in the form of the belief distributions, can be accumulated from a sequence of views. The experiments have demonstrated that the maximum likelihood hypothesis is largely viewpoint-invariant, implying that merging votes for the different hypotheses over a sequence of views should lead to a clear winner. Because the beliefs are not normalized, we have given equal weighting to all hypotheses by binarizing the values above a threshold. We have illustrated that by histogramming the binarized beliefs and picking the highest score of the result, we choose the correct winner in all cases.

A major strength of the method is its potential for a wide variety of applications. For example, an active recognition agent can choose viewpoints that will maximize the belief distribution associated with an object of interest. We have not specified *how* to choose this viewpoint, but the method can be used to determine if the particular choice leads to a sufficient level of information. Another important application of the methodology is a strategy for off-line computation of a pre-computed set of characteristic views. One can rank these views according to the belief distributions, and then store the n best views. Predefining these views speeds up on-line computations by directing the active agent's attention to informative views, thereby reducing the search space of viable hypotheses. These and other topics are currently under investigation in our laboratory.

References

[1] E. Aloimonos, Y. Purposive, qualitative, active vision. *CVGIP: Image Understanding*, 56(1):3–129, 1992. special issue.

[2] T. Arbel. Recognizing volumetric objects in the presence of uncertainty. Master's thesis, McGill University, Montréal, Québec, Canada, April 1995.

[3] T. Arbel, F. P. Ferrie, and P. Whaite. Informative views and active recognition. Technical Report TR-CIM-94-16, Center for Intelligent Machines, McGill University, Montréal, Québec, Canada, Dec. 1994. Available via ftp at `ftp.cim.mcgill.ca` in `pub/techrep/1994/CIM-94-16.ps.Z`.

[4] T. Arbel, P. Whaite, and F. P. Ferrie. Recognizing volumetric objects in the presence of uncertainty. In *Proceedings 12th International Conference on Pattern Recognition*, pages 470–476, Jerusalem, Israel, Oct 9-13 1994. IEEE Computer Society Press.

[5] F. Arman and J. Aggarwal. Model-based object recognition in dense-range images - a review. *ACM Computing Surveys*, 25(1):5–43, apr 1993.

[6] R. Bajcsy and F. Solina. Three dimensional object recognition revisited. In *Proceedings, 1ST International Conference on Computer Vision*, London,U.K., June 1987. Computer Society of the IEEE, IEEE Computer Society Press.

[7] A. H. Barr. Superquadrics and angle preserving transformations. *IEEE Computer Graphics and Applications*, 1(1):11–23, Jan. 1981.

[8] K. Bowyer and C. Dyer. Aspect graphs: An introduction and survey of recent results. In *Close Range Photogrammetry Meets Machine Vision, Proc. of SPIE*, volume 1395, pages 200–208, 1990.

[9] R. T. Chin and C. R. Dyer. Model-based recognition in robot vision. *Computing Surveys*, 18(1):67–108, mar 1986.

[10] T. Darrell, S. Sclaroff, and A. Pentland. Segmentation by minimal description. In *Proceedings, 3RD International Conference on Computer Vision*, pages 112–116, Osaka,Japan, Dec. 1990. Computer Society of the IEEE, IEEE Computer Society Press.

[11] D. Eggert and K. Bowyer. Computing the orthographic projection aspect graph of solids of revolution. In *PROC. of IEEE Workshop on the Interpretation of 3-D Scenes*, pages 102–108, Austin, Texas, November 27-29 1989. IEEE.

[12] D. Eggert, K. Bowyer, C. Dyer, H. Christensen, and D. Goldgof. The scale space aspect graph. In *Proceedings, Conference on Computer Vision and Pattern Recognition*, pages 335–340, Champaign, Il., June 15-18 1992. IEEE.

[13] 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.

[14] 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.

[15] J. Koenderink. The singularities of the visual mapping. *Biological Cybernetics*, 24(1):51–59, 1976.

[16] J. Koenderink. Internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32(4):211–216, 1979.

[17] D. Kriegman and J. Ponce. Computing exact aspect graphs of curved objects: Solids of revolution. In *PROC. of IEEE Workshop on the Interpretation of 3-D Scenes*, pages 116–122, Austin, Texas, November 27-29 1989. IEEE.

[18] 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.

[19] D. Marr. *Vision*. W.H. Freeman & Co., San Francisco, 1982.

[20] 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.

[21] N. S. Raja and A. K. Jain. Recognizing geons from superquadrics fitted to range data. *Image and Vision Computing*, April 1992.

[22] G. Soucy. View correspondence using curvature and motion consistency. Master's thesis, Dept. of E.E., McGill Univ., 1992.

[23] T. Sripradisvarakul and R. Jain. Generating aspect graph for curved objects. In *PROC. of IEEE Workshop on the Interpretation of 3-D Scenes*, pages 109–115, Austin, Texas, November 27-29 1989. IEEE.

[24] J. Subrahmonia, D. B. Cooper, and D. Keren. Practical reliable bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants. LEMS 107, Brown University LEMS, Laboratory fo Engineering Man/Machine systems, Division of Engineering, Brown University, Providence, RI 02912, USA, 1992.

[25] A. Tarantola. *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier Science Publishing Company Inc., 52, Vanderbuilt Avenue, NewYork, NY 10017, U.S.A., 1987.

[26] P. Whaite and F. Ferrie. Model building and autonomous exploration. In *SPIE - Intelligent Robots and Computer Vision XII: Active Vision and 3D Methods*, pages 73–85, Boston, Massachusetts, Sept. 8–9 1993.

[27] 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.

[28] P. Whaite and F. P. Ferrie. Active exploration: Knowing when we're wrong. In *PROC. Fourth International Conference on Computer Vision*, pages 41–48, Berlin, Germany, May 11-14 1993. Computer Society of the IEEE, IEEE Computer Society Press.

[29] 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.