Recognition/ Matching/ Segmentation

Informative Views and Sequential Recognition

Tal Arbel and Frank P. Ferrie

Centre for Intelligent Machines, McGill University, Montréal, Québec, CANADA H3A 2A7

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. 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.

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 [2]. 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 [3]. 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

[1] Strategies for gaze planning are operationally defined [4, 5].

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 [6]. A good strategy here would be to select the n best views of an object ranked according to its belief distribution.

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 needed. This leads to a sequential recognition strategy that seeks to improve uncertain interpretations by accumulating evidence over several views. 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 [7]. 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 [8, 9, 10, 11], deformable solids [12], and algebraic surfaces [13]. In our context, models are constructed through a process of *autonomous exploration* [4, 5] in which a part-oriented, articulated description of an object is inferred through successive probes with a laser range-finding system. The set-up used to perform experiments consist of 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 1a, a process of bottom-up shape analysis leads to an articulated model of the object's shape (Figure 1b) in which each part is represented by a superellipsoid primitive [11]. Associated with each primitive is a covariance matrix **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 [4]. A system which automatically builds object models based on this principle is reported in [5, 14].

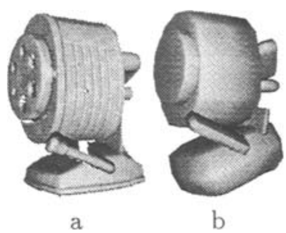


Fig. 1. (a) Laser range finder image of a pencil sharpener rendered as a shaded image. (b) An articulated, part-oriented model of the sharpener using superellipsoid primitives; 8 superellipsoids are used, one corresponding to each of the parts of the object.

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 [15], dot product [12] 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 [4]. 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 [2]. Earlier work has shown how this theory can be used to methodically synthesize belief distributions corresponding to each model hypothesis, $\mathcal{H}_\rangle$, given the parameters corresponding to the unknown model, $\mathcal{M}$, computed from the current measurement D_j, i.e. $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|})$ [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. [13] - 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 [7, 16]) 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 $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|})$ and $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_{|+\infty}})$ that is equivalent to $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|+\mathcal{D}_{|+\infty}})$. 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. In Section 3 these results are combined to form an incremental recognition scheme, and in Section 4 we show how it can be applied to the recognition of complex, multi-part objects. Finally, we conclude in Section 5 with a summary of the results and a pointer to future applications.

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

2 Determining Which Viewpoints are Informative

In order to be able to determine whether a viewpoint is 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. 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 [2]. 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.

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 a simple matter of combining the sources of information. Tarantola defines the logical *conjunction* of states of information such that the solution to the inverse problem is given by the theory AND the measurements AND any a priori information about the models. 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. 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 the classical Bayesian solution under the aforementioned conditions. The a posteriori information about the model parameters is given by the marginal probability density function:

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

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:

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

   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 [5], 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:

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

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

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

   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

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

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.

2.2 Determining Which Viewpoints are Informative using the Inverse Theory

Figure 2 shows by example how the resulting belief distribution can be used to differentiate between informative and uninformative viewpoints. In this case, 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.

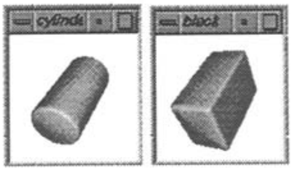
The problem of distinguishing between the two kinds of states becomes one of determining the threshold, below which one can safely state that the beliefs are in fact insignificant. It is obvious that cases where the beliefs in all the models are zero are uninformative. However, this threshold depends on the numerical precision of the system. In this sense, it is chosen externally. 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 3 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.

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

The Database



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	

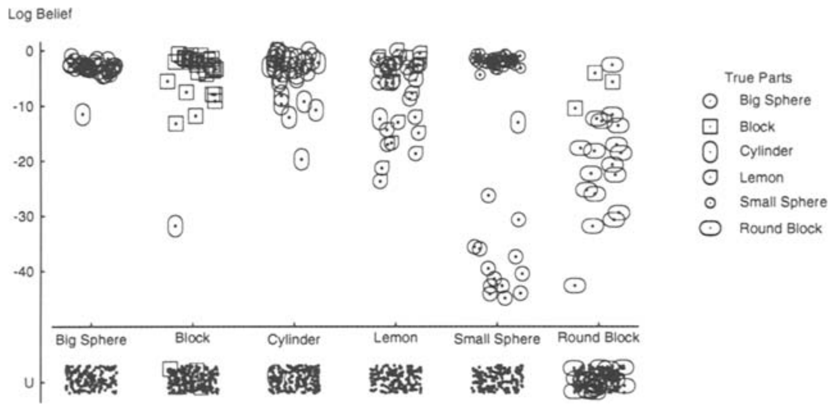
At the left of this figure are the two reference models in the data base: the cylinder and the square block. To their right are measured models of the *cylinder* obtained after scanning its surface from 4 different viewing positions. Below each model one can find the unnormalized belief distributions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

Fig. 2. (a) Informative and (b) Uninformative Views of a Cylinder.

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. 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, $\mathcal{H}_\rangle$, given the parameters of the unknown model, $\mathcal{M}$, computed from the measurement, D_j, by $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|})$. Then, given two data sets D_j and D_{j+1} corresponding to different viewpoints we seek a conjunction of $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|})$ and $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_{|+\infty}})$ that is equivalent to $P(\mathcal{H}_\rangle|\mathcal{M}_{\mathcal{D}_|+\mathcal{D}_{|+\infty}})$. 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.(See [17].) As a result, the beliefs are not normalized, making it difficult to compare the values from different viewpoints in a sensible fashion. As well, in situations where the beliefs are "close" in value, 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



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.

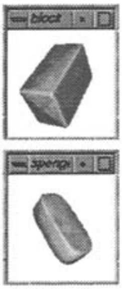
Fig. 3. Log of beliefs in the Big Sphere, Block, Cylinder, Lemon, Small Sphere, and Round Block.

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 4 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 4a indicate that the system is able to identify the block despite wide variations in its three dimensions. The ambiguous cases (Figure 4b) 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 4c), 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.

In order to communicate the validity of all hypotheses above a particular threshold, the beliefs are binarized at a threshold value. By normalizing our confidence values in this manner, combining them from different viewpoints becomes straightforward. Should the maximum likelihood hypothesis prevail in a



Ref. Models

Measured Model						
Belief in Block						
Unnorm.	0.2	0.007	2.0×10^{-13}	3.4×10^{-13}	0	0
Bin.	1	1	1	1	0	0
Belief in Rnd Block						
Unnorm.	0	0	5.8×10^{-6}	0.002	0	0
Bin.	0	0	1	1	0	0
	a) Informative		b) Ambiguous		c) Undetermined	

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

Fig. 4. Informative, Ambiguous, and Undetermined States for the Block.

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.

Figure 5 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.

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

Displayed above are the 9 models resulting from sequentially measuring the square block at 40° increments. From top to bottom, 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.

Fig. 5. Incremental Recognition of a Block.

4 Application to Complex Objects

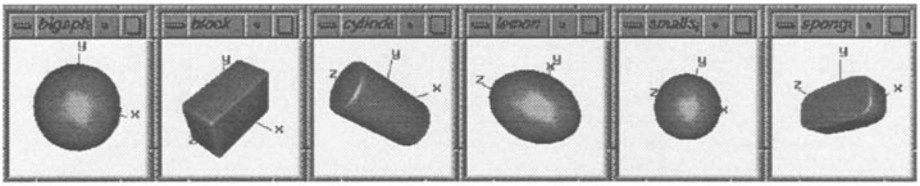
In practical recognition scenarios objects rarely correspond to the simple shapes depicted in Figure 5. A more realistic object model, such as the one shown earlier in Figure 1, accounts for the fact that objects are often comprised of multiple parts that can be articulated in different ways. This suggests a recognition by parts approach using the sequential recognition strategy developed in the preceding sections. However the task becomes much more difficult because of the complex self-occlusions by different parts of the object. Still, the results suggest that the sequential recognition strategy outlined is sufficiently robust to cope with this added complexity.

In these experiments, two articulated models were used: 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. 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 (rad $= 20mm$, 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 [4, 5]. 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 6 illustrates the actual potato-head and alarm used in recognition experiments, and the representative models of each object that result from training.

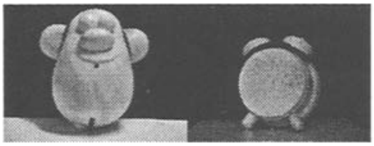
The first result (Figure 7) shows that the system is able to successfully recognize instances of articulated parts of a complex object with only partial information available, even with the added effects of self-occlusion. It also indicates that, for most models, an external threshold retained most of 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 be 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.

Table 1a shows the results of now accumulating this evidence over the sequence of views. A similar result is obtained for the alarm clock object (binarized belief distribution not shown) and shown in Table 1b. The results show that the majority of the evidence in the incorrect models was removed after application of an external 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

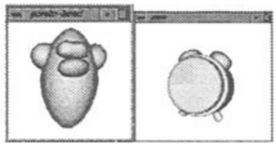


BS B C L SS RB

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).

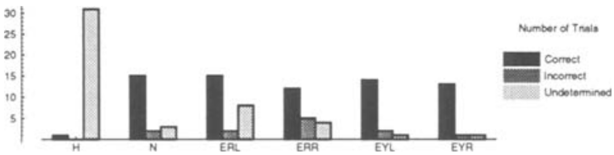


(a) Real potato-head and alarm clock used in experiment.



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

Fig. 6. The reference parts resulting from training.



Threshold = 0.00001

Displayed above is the belief distribution 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 application of a threshold of 0.00001.

Fig. 7. Matching samples of the potato-head taken from single viewpoints.

the choice of threshold was not appropriate for these parts 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 were to choose a winner based on a maximum likelihood scheme of the accumulated evidence, the results would be correct for all models[3].

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

[3] 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.

	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

a) Accumulation of evidence in potato-head, threshold = 0.00001

	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) Accumulation of evidence in alarm clock, 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 a), the measured models include the parts of the potato-head (see Figure 7 for notation.) In b), 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. Zero values are defined by a) b) a threshold of 0.0001.

Table 1. Histogram of binarized belief distributions for the potato-head and the alarm clock after 32 single-view iterations.

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 [2] 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.

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. Ed. Aloimonos, Y., "Purposive, qualitative, active vision", *CVGIP: Image Understanding*, vol. 56, no. 1, pp. 3–129, 1992, special issue.
2. Albert 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.
3. Tal Arbel, Peter Whaite, and Frank P. Ferrie, "Recognizing volumetric objects in the presence of uncertainty", in *Proceedings 12th International Conference on Pattern Recognition*, Jerusalem, Israel, Oct 9-13 1994, pp. 470–476, IEEE Computer Society Press.
4. Peter Whaite and Frank P. Ferrie, "From uncertainty to visual exploration", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 10, pp. 1038–1049, Oct. 1991.
5. Peter Whaite and Frank P. Ferrie, "Autonomous exploration: Driven by uncertainty", in *Proceedings, Conference on Computer Vision and Pattern Recognition*, Seattle, Washington, June 21-23 1994, Computer Society of the IEEE, pp. 339–346, IEEE Computer Society Press.
6. K. Bowyer and C. Dyer, "Aspect graphs: An introduction and survey of recent results", in *Close Range Photogrammetry Meets Machine Vision, Proc. of SPIE*, 1990, vol. 1395, pp. 200–208.
7. Farshid Arman and J.K. Aggarwal, "Model-based object recognition in dense-range images - a review", *ACM Computing Surveys*, vol. 25, no. 1, pp. 5–43, apr 1993.
8. A. H. Barr, "Superquadrics and angle preserving transformations", *IEEE Computer Graphics and Applications*, vol. 1, no. 1, pp. 11–23, Jan. 1981.
9. 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.
10. Narayan S. Raja and Anil K. Jain, "Recognizing geons from superquadrics fitted to range data", *Image and Vision Computing*, April 1992.
11. Frank P. Ferrie, Jean Lagarde, and Peter Whaite, "Darboux frames, snakes, and super-quadrics: Geometry from the bottom up", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 8, pp. 771–784, Aug. 1993.
12. A. Pentland and S. Sclaroff, "Closed form solutions for physically based shape modelling and recognition", in *IEEE Transactions on Pattern Analysis and Machine Intelligence: Special Issue on Physical Modeling in Computer Vision*, T. Kanade and K. Ikeuchi, Eds., July 1991, vol. 13(7), pp. 715–729.
13. Jayashree Subrahmonia, David B. Cooper, and Daniel 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.
14. A. Lejeune and F.P. Ferrie, "Partitioning range images using curvature and scale", in *PROC. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, New York City, New York, June 15-17 1993, pp. 800–801.
15. Daniel Keren, David Cooper, and Jayashree Subrahmonia, "Describing complicated objects by implicit polynomials", Tech. Rep. 102, Brown University LEMS, Laboratory for Engineering Man/Macine Systems, Division of Engineering, Brown University, Providence RI 021912 USA, 1992.
16. Roland T. Chin and Charles R. Dyer, "Model-based recognition in robot vision", *Computing Surveys*, vol. 18, no. 1, pp. 67–108, mar 1986.
17. Tal Arbel and Frank P. Ferrie, "Parametric shape recognition using a probabilistic inverse theory", *Pattern Recognition Letters*, p. TBA, 1996.