Combining Geometric and Probabilistic Structure for Active Recognition of 3D Objects

Stéphane HERBIN

Ecole Normale Supérieure de Cachan
Centre de Mathématiques et de Leurs Applications
CNRS, URA 1611
61, Av. du Président Wilson
94235 Cachan Cedex, France
Stephane.Herbin@cmla.ens-cachan.fr

Abstract. Direct perception is incomplete: objects may show ambiguous appearances, and sensors have a limited sensitivity. Consequently, the recognition of complex 3D objects necessitates an exploratory phase to be able to deal with complex scenes or objects.
The variation of object appearance when the viewpoint is modified or when the sensor parameters are changed is an idiosyncratic feature which can be organized in the form of an *aspect graph.*
Standard geometric aspect graphs are difficult to build. This article presents a generalized probabilistic version of this concept. When fitted with a Markov chain dependance, the aspect graph acquires a quantitative predictive power. Tri-dimensional object recognition becomes translated into a problem of Markov chain discrimination. The asymptotic theory of hypothesis testing, in its relation to the theory of large deviations, gives then a global evaluation of the statistical complexity of the recognition problem.

Keywords: 3D object recognition, active vision, aspect graphs, Markov chains, statistical hypothesis testing.

Introduction

Aspect graphs have been extensively studied in a theoretical way. Their practical use, however, has not been clearly demonstrated. Furthermore, their original dynamic nature has not been used directly, except in a few projects.

This article aims at showing that the dynamic nature of aspect transitions can be genuinely exploited for the recognition of 3D objects when fitted with a probabilistic model.

The general organization of the paper will be the following. In a first section, the possible application of aspect graphs as a modelling tool for object recognition will be described. Their limitations will lead us to define in a second section a probabilistic extension of the aspect graph by embedding it in a Markov chain representation. A third section will present the mathematical results stemming from the theory of large deviations in its application to hypothesis testing. In a fourth section, the mathematical features will be computed and tested on a few examples. A small review of related work is presented in the fifth section.

1 Classical aspect graphs

One of the main difficulties of vision, and presumably its foremost idiosyncratic feature, is the dimensional discrepancy between the sensors — usually 2D retinas — and the observed world of 3D objects. This characteristic generates two kinds of problems: 1) the objects are only accessible as appearences which may be *ambiguous*; 2) since objects cannot be apprehended as a whole, visual systems have to deal with a *multiplicity* of appearences related to a same object.

The final objective of visual systems description, and especially those dedicated to recognition, is to find practical or theoretical means of dealing with the appearences of objects. Many approaches have been studied in the literature: among them, the theory of *aspect graphs* has proposed an original way of reducing the multiplicity of appearances down to a global combinatorial structure. This section presents the conceptual fundation of this approach, its possible application to object recognition and its limitations.

1.1 Transitions between stable views

The multiplicity of appearences can be indexed by a state in a finite-dimension control space. The collection of a tri-dimensional object geometric appearences, for instance, may be referenced by a position in a two-dimensional manifold — the view sphere — when the viewing model is an othographic projection.

In general, the mapping between the control space and the set of appearences is locally continuous, except for some very specific areas which segment the control space in connected regions, usually called view cells. In each view cell the object appearences are considered similar or equivalent. The similarity class of appearences is called an *aspect.*

The set of aspects can be organized as a *graph* structure which represents the dissimilarity relations between view cells. The standard aspect graph example is provided by the occluding contour variation of a piece-wise smooth object. The apparent contour, which is the set of points on the object surface that have a high order contact with the viewing direction or belong to an edge, projects onto the retina as a differentiable curve except on a finite number of points. The occluding contour varies abruptly, "catastrophically", for some view directions, revealing a discontinuity in the set of appearences [19, 18]. These catastrophic transitions are called visual events.

Two contours are declared similar or of same qualitative type if they can be transformed one into another by a one to one differentiable mapping. This similarity relation defines equivalence classes of *structurally stable* views, corresponding to the segmentation of the view sphere by the catastrophe loci. Rigorous theoretic description of the aspect graphs for piece-wise smooth objects have been given [26], and some algorithms for their construction in the case of objects described as polyhedron, solids of revolution, parametric and algebraic surfaces.

The significance of the geometric approach to aspect graphs is essentially theoretic. It provides a mathematical justification of view categorization through the detection of differentiable mapping singularities. It gives a description of the aspect structure as a graph of connected view cells. It unifies the multiplicity of appearences into a global object. These three properties indicate the possibility of considering a formal description of vision both viewer-centered — since aspect graphs refer to the distribution of appearences — and object centered — since it organizes globally the appearences in a single formal object.

Conceptually, therefore, aspect graphs are appealing mathematical objects. Besides the complexity of their construction, an important question remains to state how useful

they can be in practice. This paper will concentrate thereafter on the problem of visual object recognition and show that some extensions of the standard aspect graph concept are necessary in order to make it tractable.

1.2 Aspect graphs for object recognition

The long term objective of the work presented in this paper is to describe the characteristics of a *genuinely visual* model of object recognition. Since, as it has been briefly mentioned, aspect graphs conciliate in a way the traditionnally antagonistic viewer and object centered approaches, it could be a good idea to see how they could be used to recognize objects.

The origin of aspect graphs lies in the dynamic nature of vision: the variations of aspects can only revealed by moving or modifying something in an observer visual system. An aspect graph, therefore, should not be considered as an object model since it does not consist in an objective substitute. It reveals as much of the object as of the way it is perceived, and points out the intrinsic incompleteness of vision.

A classical or geometric aspect graph is specific of an object both by the nature of its aspects and by the structure of its variations. Therefore, aspect graphs should be exploited *dynamically* in order to use fully all the information it contains, especially the structural repartition of view cells. The natural way of using classical aspect graphs for object recognition would be to move in the control space, *i.e.* change continuously the viewpoint, and detect the aspect transitions that could be considered characteristic of the object observed.

An aspect, in the classical geometric framework, is defined as the maximal set of structurally stable curves, *i.e.* the class of occluding contours that can be transformed by a diffeomorphism. The key feature of an aspect, therefore, is the distribution of singular points along the occluding contour. Transitions between aspect are detected as birth or death of singularities of the object outline.

The aspect transitions are organized in a graph structure where each node refers to an aspect. In order to recognize an object, it is enough to detect empirically what is the graph related to the object observed by collecting a sequence of aspects until a recognition decision can be made. It is assumed in this scheme that a given object can be fully characterized by the graph structure of its aspects.

We can now state what could be the most important elements of a recognition system based on aspect graph discrimination. It shoud be characterized by:

- Visual features: occluding contour singularities.
- Data structure: distribution of singular points.
- Control parameters: view point direction.
- Recognition principle: graph structure discrimination.

1.3 Difficulties

The elementary recognition scheme described above, although conceptually appealing, has to face many difficulties. There are mainly two series of obstacles to its practical use: aspect graphs are too complex to build *and* to use, and in many cases, the concept itself of aspect graph do not bring enough reliable discriminating information for recognition.

The complexity of aspect graphs The construction of aspect graphs, although theoretically solved for many classes of objects, is computationnaly expensive. Most of the thorough studies have concentrated on a single object [24, 35], revealing the difficulties of designing an automatic procedure.

An evaluation of the intrinsic complexity of aspect graphs has been performed by S. Petitjean [23] for piecewise-smooth algebraic surfaces. For an object made of n quadrics, for instance, the number of aspects exact upper bound is a polynome of degree n^{12} when an orthographic projection is used. This huge bound may explain why aspect graphs are difficult to build in the case of complex objects, and intractable in practice.

Besides this intrinsic complexity, one may suspect that the direct use of the graph structure is also a difficult problem. The question of deciding whether two graphs are isomorphic has never been proved to be NP-complete, but no algorithm in polynomial time has ever been found either [12, 17]. In the othographic case, however, the aspect graph is planar, which implies that a faster algorithm can be used. In the general case, and since the number of aspects is huge, even for a simple object, the recognition principle cannot lie only in a graph structure discrimination.

The limitations of classical aspect graphs The "classical" concept of aspect graph itself may not be adequate for object recognition because of three reasons.

Firstly, the notion of singular point on the contour comes from a continuous, and even differentiable, approach of vision processes. In practice, most of the available signals are in the form of pixel arrays, where the notion of singular point is ill defined. If aspects can be built from a theoretical object model, it seems difficult to detect them empirically on images obtained from a digital camera retina, for instance.

Secondly, the control space contains a unique parameter, although in general multi-dimensional: the viewpoint direction and position. Classical aspect graphs are based essentially on geometric features of the objects. Information such as color, texture or photometry, are not used and may have in some occasions a greater discriminating power.

Thirdly, the graph structure may not hold enough discriminating information. **Fig.** 1 illustrates this problem.

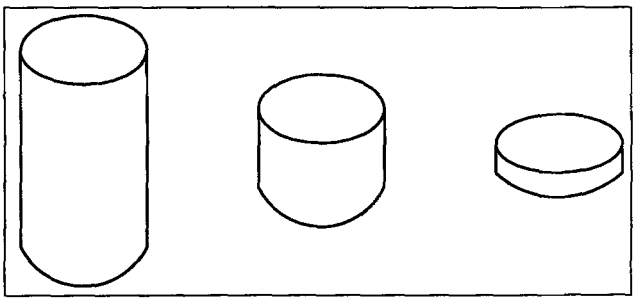


Fig. 1. Three different objects from the same viewpoint having the same geometric aspect graph.

The three objects have the same simple aspect graph but are clearly distinguishable since, for instance, they would not be naturally grasped in the same manner. This fact implies that something more must be added to the classical geometric approach to aspect graphs to efficiently and reliably use them for recognition purposes.

2 Probabilistic aspect graphs

The first section has proposed a way of using aspect graphs for object recognition based on its original dynamic nature, and pointed out several limitations of the approach.

This section intends to show that some of the previously exposed difficulties can be removed if a probabilistic model is fitted to the graph structure in the form of a Markov dependence.

2.1 Probable aspects

Research in computer vision have been mostly interested until now in a geometric or topological description of phenomena: this step was necessary in order to make the intuitions more formal and rigorous, and allowed the practical use of rather abstract mathematical fields. The recent development of geometric invariants is one of the most prominent examples.

Geometric theories are essentially qualitative: they do not get along very well with metric measurements. Recently, however, several studies dealing with quantitative concepts have been conducted. They concern the definition of geometric robustness and utility. thanks to a concept of canonical or generic object view.

A canonical view can be characterized according to two dimensions: stability and likelihood. The definition of a view likelihood relies on a very simple and rather intuitive phenomenon: the probability distribution of an angle defined by two intersecting lines has a maximum when the viewing direction is perpendicular to the plane defined by the the two lines, if the view sphere is uniformly sampled [2, 5]. This "peaking effect" defines therefore a most probable view and has been exploited in [31] to recognize very simple polyhedrons. Another definition of generic view based on a notion of bayesian stability was proposed in [10].

The two concepts of likelihood and stability of views have been generalized in [37] and used in [13] to determine the canonical views of some simple smooth objects from their occluding contour. It was shown it these papers using some formal definitions that stable views are the most likely observed ones. More generally, canonical viewpoints are orthogonal to the first two 3D object inertia axes.

The practical use of these notions of stability and likelihood is difficult: the complexity of analytical results increases drastically with the complexity of the objects. Furthermore, and more fundamentally, the definition of a canonical view using a stability criterion is contradictory with the definition of an aspect as a set of views referenced by a region in the control space delimited by a set of catastrophic hypersurfaces. One can represent an aspect by the most stable view as a prototype, but the stability criterion does not say anything about the overall aspect organization and on their potentially predictive capacity.

The general framework studied in this paper is to allow the system to generate dynamically new object appearences by modifying its viewpoint or other control parameters. The key feature is the structure of the appearence variation. Therefore, what should be embedded in a probabilistic environment is the collection of aspects themselves, and not the set of views or appearences. The origin of an aspect is, of course, of geometric nature, but the only accessible reality is an already categorized view, in other words, the aspect itself.

The reduction of the object geometry to a probability distribution on categorized visual data allows one to integrate roughly some uncertainty on the appearence variations. It also gives a quantification of the utility of views. Some aspects will be more often observed than others depending on the object geometry; the conditional probability of aspect occurence can be converted into a likelihood. The next section will explain in more details how the aspect probability structure can be analyzed statistically.

2.2 Markov representation of aspect graphs

Visual aspects are categories of object appearences. The graph structure that comes with them, deduced from the visual events organizing their variations, can be considered as a primitive predictive tool. The graph connectivity foretells what are the expected visual events, and thus what are the possible observable aspects.

The probabilistic description of aspects gives a quantitative measure of their empirical utility. In the same way, it is possible to embed the graph structure into a probabilistic environment in order to provide some utility measure of the visual events themselves. Some events will happen more likely than others, will be more stable, will give a better and more robust characterization of the object observed.

The true origin of visual events is vision dynamics. Aspect change can only be detected if something has been actively modified, if the viewpoint has been forced to move. Said informally, it is necessary to act if you want that something happens.

It is possible to be more formal, however. If s_t represents an observed aspect at a given discrete time t, and if an action a_t from a control space is produced, the probability that an event of the form $\mathrm{s}_t \rightarrow \mathrm{s}_{t+1}$ for an object o_k can be described as a probability transition: $\Pr[\mathrm{s}_{t+1} \mid \mathrm{s}_t, a_t, o_k]$.

This last expression simply means that the probability of observing a given visual event depends on the type of object observed, on the type of viewing parameter modification, and on the current aspect observed. This probabilistic interpretation of a visual event leads to a modelling of an aspect graph as a controlled Markov chain, specific to a given object.

This probabilistic aspect graph can be simplified by assuming that the the actions are drawn from a stationnary probability distribution $\mu(a)$. A homogeneous Markov chain can then be derived from this law as:

$$\mathbf{P}_k[\mathrm{s}_{t+1} \mid \mathrm{s}_t] = \int_a \Pr[\mathrm{s}_{t+1} \mid \mathrm{s}_t, a, o_k] \, \mathrm{d}\mu(a)$$

An object can now be modelled as a Markov chain describing the probabilistic evolution of its aspects.

At this point, it should be necessary to be a bit more precise about what I would call a *model* of an object. The probabilistic formulation of an aspect graph, as I have mentioned it above, cannot be considered as an approximate substitute of some reality. It characterizes only the *a priori* structural relations of appearences when some viewing parameters are modified according to a stationnary stochastic action law, and when the views are categorized by a fixed process. An object model must be understood as a prediction or anticipation of a sensory actuality when an observer interacts with it.

The Markovian dependence is able to take into account two different series of features: empirical measures of event utility, leading to the definition of a likelihood, and definition of visual events. The transitions with probability equal to zero and those with positive probability do not have the same interpretation. Indeed, the first ones reveal the structural organization of events, whereas the second ones measure their relative frequencies, *i.e.* their informative capacity. This double nature of the probability transitions will be used more precisely in a next section for the actual recognition of objects and the measure of recognition problem complexity.

2.3 Construction as estimation

If a Markov representation of aspect graph is used, a natural method for the determination of the model is an empirical estimation. The exact aspect graph construction translates into a problem of statistical learning. The complexity to deal with is of

another kind: problem of dimension, learning time, approximation and generalization errors. This new type of complexity is usually more manageable and quantizable.

The dynamic nature of aspect graphs is captured by the notion of stochastic process. The construction of a probabilistic aspect graph will simply consist in sampling the Markov chain by generating a random sequence of actions drawn from the stationnary control law $\mu(a)$. For each object o_k all there is to do is to detect the aspect transitions $s' \rightarrow s$ and collect the frequencies $n_k(s, s')$. The corresponding probability transition will be obtained by using the standard maximum likelihood estimator [3]. From a computing point of view, an estimated aspect graphs will therefore consist in a array of integers, with many coefficients equal to zero.

There are two types of questions related to the estimation of Markov representation of aspect graphs: 1) when the learning time is unbounded, when can it be decided to stop the estimation in order to achieve a given level of confidence; 2) when the learnig time is limited, what to do with the available samples. These two questions have been studied in [15] and will be presented in a forthcoming paper. The rest of the paper is devoted to a mathematical analysis and characterization of a 3D object recognition based on Markov representations of aspect graphs.

2.4 Generalized aspects

So far, the only type of action that has been explicitly mentioned was a modification of the viewpoint, according to the classical geometric approach. It has also been pointed ou that the concept of singular point was ill defined in the context of pixel-based sensory data. Another notion of singular point, however, can be defined by introducing a *scale* of analysis.

The definition of a generalized singular point on a contour, controlled by a scale of analysis, has been proposed in a previous paper [14], following classical work on edge detection. It was emphasized that no optimal intrinsic analysis scale can be found in order to characterize a given digital curve. A shape should be analyzed at all scales. The philosophy of *scale space theory* [9, 20, 21, 11] claims precisely that the global spectrum of local extrema obtained by spatial filtering at all scales brings useful information. This means that — if we relate it to our problem of *active* recognition — scale should be considered as a control parameter, and should be actively modified.

The concept of scale of analysis, which was introduced rather naturally in order to give a a definition of a singular point on a digital contour, is in fact the prototype of a new kind of appearence variety. In other words, the variations of object appearence when the viewpoint is modified and when the scale of analysis is changed are not different in essence. The scale can be added to the control space as another dimension. From the observer, moving the viewpoint or changing the scale of analysis produces aspect transitions which are undistinguishable.

Some notion of scale in relation to aspect graphs has been proposed in [8]. The scale was not used however as a free control feature but as a parameter able to deal with some kind of "geometric noise".

3 Active recognition as a statistical inference

Markov chain representations of aspect graphs are generated by sampling a stationnary action law in a given control space. This control space may not contain only specifications of the viewpoint, but also parameters characterizing the description of appearences such as scales of pattern analysis.

Tri-dimensional object recognition translates now into a problem of Markov chain discrimination. A classical approach to this problem is to generate an empirical trajectory of a given observed chain and produce a statistical inference from the sequence of collected states — in our case a sequence of aspects.

The observation of certain aspect transitions will be rare or even impossible for some objects: they will be *selective*. When the visual features and the categorized appearences are well adapted to a given recognition problem, several aspects will be able to reshape the set of hypothetical objects by rejecting those that could not be associated with some observed aspects or transitions. In the extreme case, some observed aspect transitions will be able to index directly a given object by rejecting all the others.

The general recognition principle will be the following: the system samples the Markov chain corresponding to a given object by generating new random actions. When a selective transition is observed, a set of rejecting decision can be made in order to restrict the set of hypotheses. If it is required to take a decision based on the available data, a test is performed according to a likelihood ratio.

In this recognition scheme, two different statistical regimes compete. The first one is a function of a set of likelihood values which is modified by interacting with the environment, the second uses *a priori* discriminating data to select dynamically the set of potentially observable objects. This section is intended to formalize this scheme and to give several mathematical tools to analyze it.

3.1 Hypothesis testing of positive Markov chains

The decision principle for the recognition scheme that will be presented in this paper consists in testing whether a given object is more likely to be observed than any other. The decision will be based on the computation of all the likelihood ratios for all couples of objects once the sequence of aspects has been collected. We restrict the study in this subsection to the case of Markov representation of aspect graphs described by positive stochastic matrices.

A hypothesis test $D_T^{\Phi}(k,k')$ for a couple of objects o_k and $o_{k'}$ is a random variable taking values in $\{0,1\}$. It takes the value 1 if it is estimated that the object o_k is more likely to be observed than object $o_{k'}$ according to the sequence of aspects $\Phi_T = (S_0, S_1, ..., S_T)$, and 0 otherwise.

Globally, it will be decided that the observed object is o_{k^*} if:

$$\forall k \neq k^*, \quad D_T^{\Phi}(k^*,k) = 1.$$

If not, no decision will be taken (the hypotheses are rejected). This decision scheme implies that the important features to study are the comparisons of two objects.

As it is customary, we define the errors of first and second kind $\alpha_T(k,k')$ et $\beta_T(k,k')$ as:

$$\alpha_T(k,k') = \mathbf{P}_k\left[D_T^{\Phi}(k,k') = 0\right] \text{ and } \beta_T(k,k') = \mathbf{P}_{k'}\left[D_T^{\Phi}(k,k') = 1\right]$$

which measure the mean error of rejecting the true hypothesis, and the mean error of accepting the wrong one.

The Markov representation of an aspect graph will be written as a stochastic matrix $p_k(i,j)$ if it is related to the object o_k. The columns of the matrix will be considered as the conditional probabilities: $p_k(i,j) = \mathbf{P}_k[S_t = i \,|\, S_{t-1} = j]$ if S_t is the random variable describing the observed aspect at time t. We define now the random variables $Y_t(k,k')$ as:

$$Y_t(k,k') = r_{kk'}(S_t, S_{t-1}) = \log\left[\frac{p_{k'}(S_t, S_{t-1})}{p_k(S_t, S_{t-1})}\right]$$

They can be considered as the individual contribution of each new generated action to the log-likelihood ratio between the objects o_k and $o_{k'}$. These contributions are then summed and normalized in:

$$L_T(k,k') = \frac{1}{T}\sum_{t=1}^{T} Y_t(k,k') \tag{1}$$

If a fixed threshold λ is given, the likelihood ratio test consists in deciding that the object o_k is more likely observed than $o_{k'}$ if

$$L_T(k,k') < \lambda$$

The first and second kind errors are then defined as $\alpha_T(k,k') = \mathbf{P}_k\,[L_T(k,k') \in [\lambda, +\infty[]$ and $\beta_T(k,k') = \mathbf{P}_{k'}\,[L_T(k,k') \in]-\infty, \lambda[]$.

The law of large numbers applied to Markov chains allows one to state, if the object o_k is observed, that:

$$L_T(k,k') \underset{T\to\infty}{\longrightarrow} \sum_{j=1}^{|\mathcal{S}|} \mu_k(j) \sum_{i=1}^{|\mathcal{S}|} p_k(i,j) \log\left(\frac{p_{k'}(i,j)}{p_k(i,j)}\right) = -K(\mathbf{P}_k \mid \mathbf{P}_{k'}) < 0 \tag{2}$$

where $\mu_k(j)$ is the invariant measure of the Markov chain $p_k(i,j)$. The coefficient $K(\mathbf{P}_k \mid \mathbf{P}_{k'})$ plays the role of an entropy between the two chains and gives an idea of what should be the useful interval for the threshold λ used to compare the normalized log-likelihood ratio.

A better characterization of the likelihood ratio test will use asymptotic results from the theory of large deviations. Indeed, in a recognition problem, what we are really interested in is the errors behavior. One can prove [22, 16, 6]:

Theorem 1 *The likelihood ratio test for the positive Markov chains $\mathbf{P}_k$ and $\mathbf{P}_{k'}$ and a fixed threshold $\lambda \in]-K(\mathbf{P}_k \mid \mathbf{P}_{k'}), K(\mathbf{P}_{k'} \mid \mathbf{P}_k)[$ has first and second kind errors $\alpha_T(k,k')$ and $\beta_T(k,k')$ verifying:*

$$\lim_{T\to\infty} \frac{1}{T} \log \alpha_T(k,k') = -\Lambda_{kk'}(\lambda) < 0$$

$$\lim_{T\to\infty} \frac{1}{T} \log \beta_T(k,k') = \lambda - \Lambda_{kk'}(\lambda) < 0$$

where the rate function $\Lambda_{kk'}(\lambda)$ is defined as

$$\Lambda_{kk'}(\lambda) = \sup_{x\in[0,1]} (x\lambda - \chi_{kk'}(x)) \tag{3}$$

The function $\chi_{kk'}$ is given by $\chi_{kk'}(x) = \log \rho(\Pi_k(r_{kk'}, x))$ where $\rho(.)$ is the Perron-Frobenius eigen value of the matrix $\Pi_k(r_{kk'}, x) = \{p_k(i,j).\exp[x.r_{kk'}(i,j)]\}_{i,j\in\mathcal{S}}$ and $r_{kk'}(i,j) = \log[p_{k'}(i,j)/p_k(i,j)]$.

This theorem states that the likelihood ratio test generates errors going to zero exponentially fast. The logarithmic speed of convergence is characterized by the rate function $\Lambda_{kk'}(\lambda)$ which can be computed when the stochastic matrices are known.

When the decision threshold λ is zero, the speed of convergence of the two errors are equal. One can show easily that, in this case, the Bayes risk is optimal regarding its converging rate. The rate function corresponding to a zero threshold $\Lambda_{kk'}(0)$ is often called also Chernoff information or bound.

3.2 General case: Non negative stochastic matrices

The asymptotic results presented above require strictly positive stochastic matrices. In practice, the generic structure of aspect graphs has many impossible transitions. The corresponding incidence matrices contain therefore many null coefficients. The stochastic matrices will be in general sparse, although irreducible since all aspects communicate. Likelihood ratios will be defined only for the set of common positive probability transitions.

We define now the set of positive probability transitions for a given object o_k as:

$$\mathcal{T}_k = \{(i,j) \in \mathcal{S}^2 \,/\, p_k(i,j) > 0\}.$$

where $\mathcal{S}$ is the set of observable aspects. The transitions in this set will be called *admissible* for the object o_k. Using a conditionning by the set of admissible transitions for two objects, the decision can be divided into two terms:

$$D_T^{\Phi}(k,k') = \begin{cases} C_T^{\Phi}(k,k') & \text{if} \quad \Phi_T \in \mathcal{T}_{kk'}^T \\ S_T^{\Phi}(k,k') & \text{if} \quad \Phi_T \notin \mathcal{T}_{kk'}^T \end{cases}$$

where $\mathcal{T}_{kk'} = \mathcal{T}_k \cap \mathcal{T}_{k'}$.

The first term, the *comparative* part, is defined for the admissible transitions of two objects ($\Phi_T \in \mathcal{T}_{kk'}^T$). The second term, *selective*, takes a decision as soon as a selective transition has been observed. The errors can be written as:

$$\alpha_T(k,k') = \alpha_T^C(k,k')\mathbf{P}_k[\Phi_T \in \mathcal{T}_{kk'}^T] + \alpha_T^S(k,k')\mathbf{P}_k[\Phi_T \notin \mathcal{T}_{kk'}^T]$$
$$\beta_T(k,k') = \beta_T^C(k,k')\mathbf{P}_{k'}[\Phi_T \in \mathcal{T}_{kk'}^T] + \beta_T^S(k,k')\mathbf{P}_{k'}[\Phi_T \notin \mathcal{T}_{kk'}^T]$$

The selective part of the decision, $S_T^{\Phi}(k,k')$, do not produce any error: when a selective transition has been observed, a secure rejection decision can be taken if the set of admissible transitions has been perfectly identified. The errors, thus, come from the comparative term of the decision.

The global evaluation of the errors behavior necessitates another mathematical result. It is possible to compute the exit logarithmic speed of a Markov chain from a set of transitions [29]:

Theorem 2 *Let $\mathcal{T} \subset \mathcal{S}^2$ be a transition subset of a given Markov chain $\mathbf{P}_k$ having states in $\mathcal{S}$, and Φ_T a trajectory of length $T+1$. There exists a number $\rho_k(\mathcal{T}) < 1$ such that:*

$$\lim_{T\to\infty} \frac{1}{T} \log \mathbf{P}_k[\Phi_T \in \mathcal{T}^T] = \log \rho_k(\mathcal{T}) < 0$$

This number is the Perron-Frobenius eigenvalue of the matrix $p_k(i,j)$ where the coefficients belonging to $\mathcal{S}^2 \backslash \mathcal{T}$ have been set to zero.

With this result, it is easy to evaluate the global asymptotic behavior of the errors when the trajectory length goes to infinity:

$$\alpha_T(k,k') = \alpha_T^C(k,k')\mathbf{P}_k[\Phi_T \in \mathcal{T}_{kk'}^T] = A_T.e^{-\nu_\alpha T} \tag{4}$$
$$\beta_T(k,k') = \beta_T^C(k,k')\mathbf{P}_{k'}[\Phi_T \in \mathcal{T}_{kk'}^T] = B_T.e^{-\nu_\beta T} \tag{5}$$

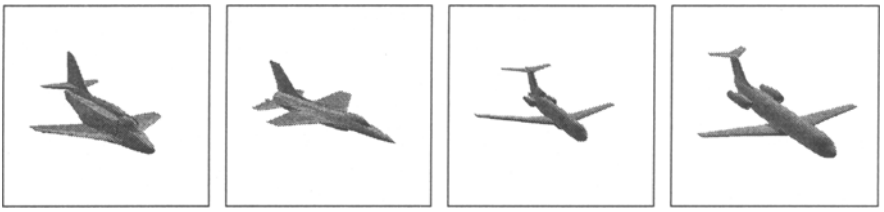


Fig. 2. Example of four 3D objects. The two objects on the left are militar planes, the two on the right are civil planes.

where $\lim_{T\to\infty} \frac{1}{T} \log A_T = \lim_{T\to\infty} \frac{1}{T} \log B_T = 0$. The global logarithmic speeds of convergence can be computed as the sum of two terms:

$$v_\alpha = \Lambda'_{kk'}(\lambda) - \log \rho_k(\mathcal{T}_{kk'}) > 0$$
$$v_\beta = \Lambda'_{kk'}(\lambda) - \lambda - \log \rho_{k'}(\mathcal{T}_{kk'}) > 0$$

where the rate function $\Lambda'_{kk'}(\lambda)$ will be computed using the formula (3) on the Markov chains conditionned by the set of admissible transitions $\mathcal{T}_{kk'}$.

These asymptotic results give a global characterization of a 3D object recognition problem. The complexity of the problem can be measured by a single number: the error logarithmic speed of convergence to zero. This number is the sum of two terms.

The first one, $\Lambda'_{kk'}(\lambda)$, is a quantitative measure of the occurence of aspects. Thanks to this number we are able to quantify the difference between two objects having the same geometric aspect graph, but with different view frequencies. The simple recognition problem presented in the first section in **Fig.** 1 can now be solved and quantified.

The second term, $-\log \rho_k(\mathcal{T}_{kk'})$, characterizes the differences between two sets of visual events. The bigger this term, the more likely recognition will be produced by structural comparisons as the classical aspect graph approach to object recognition would have produced.

4 Experiments

The theoretical elements presented above will be applied to a simulated problem of 3D object recognition. The objects tested are rather complex polyhedric representations of planes (**Fig.** 2). Three different couples will be tested: two militar planes, the civil planes and a couple formed by a militar and a civil planes. Note that the civil planes differ mainly on the relative sizes of their components.

The method used to define an object aspect requires the choice of a total number of aspects. [14, 15] present in more details the algorithms used. This paper concentrates on the conceptual foundation of an active recognition based on sampling randomly an action or control space, and on its the mathematical analysis.

The problem of recognizing 3D objects has been translated into that of discriminating Markov representations of probabilistic aspect graphs. The complexity of the problem has been characterized globally by its asymptotic behavior, which can be quantized by a logarithmic speed of convergence of the error to zero.

The computation of the recognition complexity measure depends only on the stochastic matrices and on their structure. The simulations have tested essentially the values of this measure when the number of aspects for a given problem varies. Several empirical measures of the actual recognition errors have been performed to evaluate the confidence that can be expected from the asymptotic results.

4.1 Aspect graph structure

The stochastic matrices representing the probabilistic interpretation of aspect graphs will be generally rather sparse. They have a dominant diagonal ($\forall i \neq j, p_k(i,i) \gg p_k(i,j)$) and contain a great ratio of probability transition with a zero value, *i.e.* selective transitions. The distribution of probability transitions can be divided into two modes: one of them concerns the diagonal elements of the matrices and are all above 0.2. This means that the Markov chain will stay in the same state during a rather long time and then move to another state. This should not be surprising since aspects have been defined as structurally stable classes of appearences.

In the theoretical presentation above, the key role of selective transitions has been emphasized since they have an infinite rejecting capacity. It can be observed that, when the number of aspects increases, the number of admissible transitions increases only linearly or sub-linearly: matrices tend to get sparser and their structure becomes more specific to the object they refer to.

Graph connectivity of admissible and selective transitions have interesting features: they exhibit a phenomenon of saturation (**Fig.** 3). When the number of aspects increases, the average graph degree, *i.e.* the average number of connections per node, increases also but with a sub-linear regime, indicating a decreasing connectivity. The average degree of the comparative transition graph, *i.e.* transitions which are common to two objects, reaches a maximum whereas the average degree of the selective transition graph increases more regularly.

The connectivity variation of the stochastic matrices when the number of aspects increases shows that the graph of transitions become more and more structurally discriminating. The comparative transitions increase their number until reaching a limit value after which they get distributed among the newly created aspects without generating new efficient selective transitions.

Several differences are noticeable between the three couples of objects tested. Although they exhibit the same global behavior, their quantitative characteristics show that the number of selective transitions is significantly smaller for the civil planes. This result is not really surprising since, "visually", they appear rather similar. The civil planes are less selectively distinguishable than the two other couples. This result should be confirmed by computing the recognition complexity measures provided by the large deviation theory.

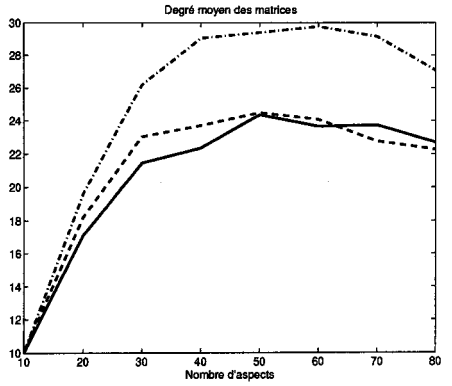
4.2 Active recognition complexity

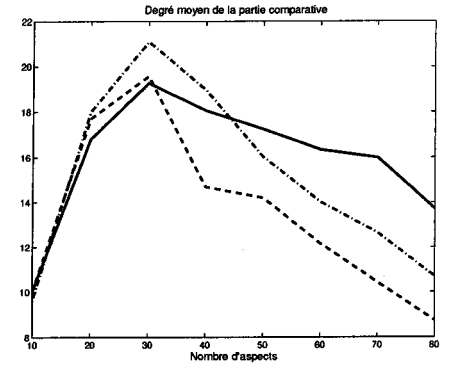
During the recognition phase, two statistical regimes compete: the first one waits until a selective transition is observed, the second one modifies incrementally the log-likelihood and takes a decision based on its final value.

The recognition complexity depends on the type of object observed, on the number of aspects, on the decision threshold and on the strategy of action sampling $\mu(a)$. In the simulations presented here the action law is a uniform random sampling corresponding to a brownian motion on the view sphere, and the decision threshold is fixed to 0.

Two numbers characterize the recognition complexity: the logarithmic speed of exiting the set of comparative transitions $-\log \rho_k(\mathcal{T}_{k'})$ and the Chernoff bound $A'_{kk'}(0)$. The sum of these two terms gives the global logarithmic speed of convergence of both the first and second kind errors $v_\alpha(k,k')$.

The graphics of **Fig.** 4 confirm the expected complexity difference between the problem of discriminating the civil planes and the two other couples of objects. Both the logarithmic exit speed and the Chernoff bound have smaller values. The global logarithmic speed of convergence increases almost linearly with the number of aspects.





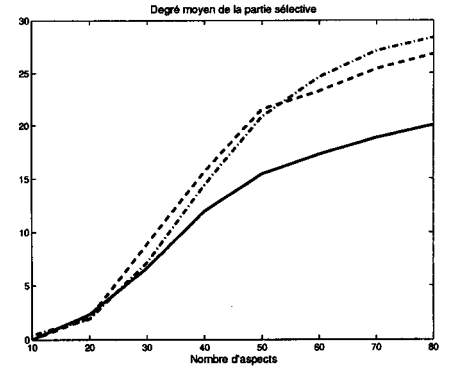
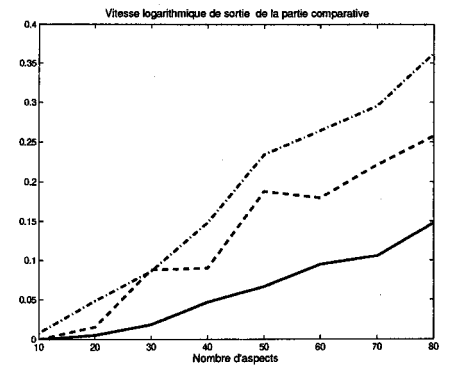
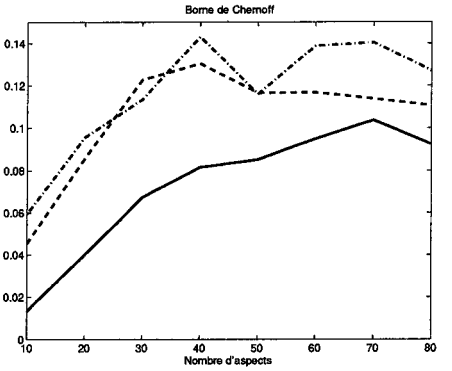


Fig. 3. Average degree of the global (top), comparative (middle) and selective (bottom) transition graphs for the couples of civil (solid line), militar (dashed line) and mixed (dotted line) planes.





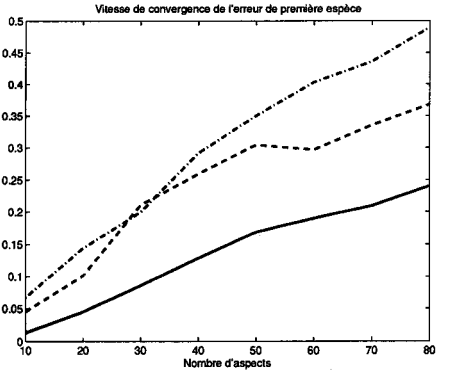


Fig. 4. Logaruthmic speed of exiting from the set of comparative transitions $-\log \rho_k(k')$ (top), Chernoff bound (middle) and logarithmic speed of convergence of the error (bottom) for the couples of civil (solid line), militar (dashed line) and mixed (dotted line) planes.

The relative difference between the civil and the militar is about 40% when 50 aspects are used.

The variation of the two terms forming the global logarithmic convergence speed of the errors are not identical. When the number of aspects reaches about 40, the increase is due mainly to the logarithmic speed of exiting the comparative transitions. The Chernoff bound saturates at a value corresponding approximately to the transition graph average degree saturation of **Fig.** 3: the statistical selective regime becomes the prominent one.

4.3 Recognition performances

The theoretical results presented above gave an asymptotic characterization of the error behavior. In practice, the asymptotic regime, which corresponds to a stationnary one with Markov chains, is seldom reached since the trajectories generated will generally be short.

Fig. 5 shows the error behavior for the different couples of objects tested and for various trajectory lengths. In a logarithmic scale, the errors decrease approximately lineraly towards zero. The expected quantitaive difference between the civil plane problem and the other two is clearly noticeable.

The asymptotic results, however, must be used with care, especially if one wants to employ numerical values. **Fig.** 6 shows the error behavior for the civil plane problem and compares it with its rate function value. On this graphic, the empirical measures show worse performances than would have been expected if the chain had reached its stationnary regime. This difference can be explained by the fact that the rate function only qualifies the logarithmic speed of convergence: the functions A_T and B_T of (4) and (5) are unknown and may be influencial for short recognition trajectories.

Another critical parameter when studying Markov chains are their second largest eigenvalue which characterize the speed of convergence towards the stationnary regime. In the examples presented in this paper, the action law produced trajectories similar to brownian motions, generating Markov chains with a second largest eigenvalue around 0.95, which is a rather large value since the speed of convergence will be proportional to 0.95^T.

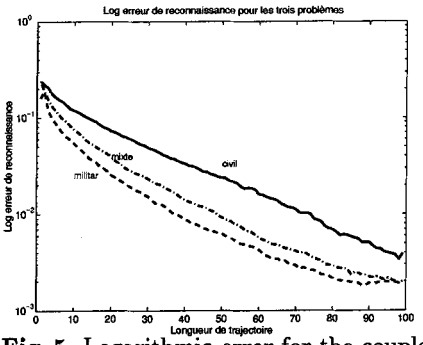


Fig. 5. Logarithmic error for the couples of civil (solid line), militar (dashed line) and mixed (semi-dashed line) planes and 50 aspects.

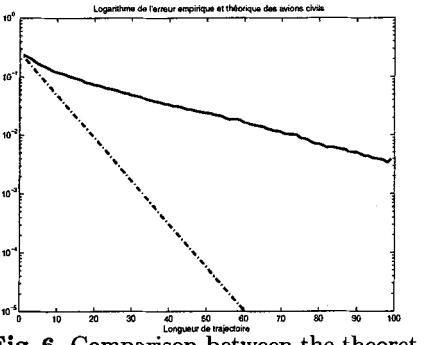


Fig. 6. Comparison between the theoretical (dashed line) and empirical (solid line) error behavior for the couple of civil planes and 50 aspects.

5 Related work

A few studies have been interested in designing recognition systems dealing with an active paradigm. Most of them describe exploratory strategies able to discover the most

discriminating point of view or set of features, and *then* perform recognition. This question can be related to the general problem of sensor planning [32]. Recognition applications have been described in [34, 38, 36, 7, 33].

Recognition based on the active accumulation of pieces of evidence, although it has a long history in the case of 2D patterns, has been less extensively studied in computer vision. [27] describe a system able to solve simple taks by seeking actively and optimally information in a scene. They use a Bayes net to control the way the evidences are handled to perform the task. The actual use of their method for 3D vision is not clear. [25] describes a strategy using two cooperative processes — locating and identifying — to detect the presence of objects in a scene. [30] proposes a general formalism for the specific selection of useful features, and a dynamic procedure to combine them. When a hypothesis is rejected, the system changes its viewpoint or extracts another set of features. The accumulation of pieces of evidence is obtained by restricted the set of potentially observable objects. [1] proposes an evaluation of the information contained in a view by computing a likelihood conditionned to a parametric object model. The likelihood is incrementally updated using a combination of probabilities.

The series of studies most related to the work presented here is [28] and more recently [4]. They describe a system able to learn aspects and to use incrementally a sequence of aspects in a recognition phase. The system is decribed using very complex coupled differential equations with many critical parameters which role seem difficult to analyze precisely. They do not provide either a clear evaluation of recognition performances.

6 Conclusion and future work

This article has shown how active vision can be used for object recognition, and has pointed out that recognition is inherently active.

A general approach combining both "objective" and "subjective" features has been proposed. It is based on the detection of visual events when an agent interacts with the environment and on their statistical purposive accumulation. A set of mathematical tools has been provided in order to analyze and quantify the behavior and the performances of the recognition procedure.

The general stochastic representation of an aspect graph probabilistic interpretation is a *controlled* Markov chain. The recognition procedure presented in this paper used a stationnary action law: one possible extension of the model is to develop more intelligent strategies in order to seek more directly, for instance, the most discriminating aspect transitions. Object recognition becomes translated into a problem of controlled Markov process discrimination, where the action law is purposively controlled by the system.

Acknowledgements

The content of this paper described part of Ph.D work directed by Robert Azencott. The author was supported by a studentship from the CNRS.

References

1. T. Arbel and F.P. Ferrie. Informative views and sequential recognition. In *Proc. Fourth European Conference on Computer Vision, Cambridge, England, April 14-18*, pages I:469–481. 1996.
2. J. Ben-Arie. The probabilistic peaking effect of viewed angles and distances with applications to 3-D object recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **12**(8):760–774, 1990.

3. P. Billingsley. *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago, London, 1961.
4. G. Bradski and S. Grossberg. Fast learning VIEWNET architectures for recognizing three-dimensional objects from multiple two-dimensional views. *Neural Networks*, **8**(7/8):1053–1080, 1995.
5. J.B. Burns, R.S. Weiss, and E.M. Riseman. View variation of point-set and line-segment features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(1):51–68, 1993.
6. A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications.* Jones and Bartlett Publishers, Boston, 1993.
7. S.J. Dickinson, H.I. Christensen, J. Tsotsos, and G. Olofsson. Active object recognition integrating attention and viewpoint control. In J-O. Eklundh, editor, *Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2-6*, number 801 in Lecture Notes in Computer Science, pages 3–14. Springer Verlag, Berlin, 1994.
8. David W. Eggert, Kevin W. Bowyer, Charles R. Dyer, Henrik I. Christensen, and Dmitry B. Goldgof. The scale space aspect graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **15**(11):1114–1130, November 1993.
9. L.M.J. Florack, B.M. ter Haar Romeny, J.J. Koenderink, and M.A. Viergever. Scale and the differential structure of images. *Image and Vision Computing*, **10**:376–388, 1992.
10. W.T. Freeman. Exploiting the generic viewpoint assumption. *International Journal of Computer Vision*, **20**(3):243, 1996.
11. J. Gårding and T. Lindeberg. Direct computation of shape cues using scale-adapted spatial derivative operators. *International Journal of Computer Vision*, **17**(2):163–191, February 1996.
12. M.R. Garey and D.S. Johnson. *Computers and Intractability — A guide to the theory of NP-completeness.* W.H. Freeman, San Francisco, 1979.
13. Y. Gdalyahu and D. Weinshall. Measures for silhouettes resemblance and representative silhouettes of curved objects. In B.F. Buxton and R. Cipolla, editors, *Proc. Fourth European Conference on Computer Vision, Cambridge, England, April 14-18*, volume 1065 of *Lecture Notes in Computer Science*, pages 363–375. Springer Verlag, Berlin, Heidelberg, New York, 1996.
14. S. Herbin. Recognizing 3D objects by generating random actions. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, June 18-20*, pages 35–40, 1996.
15. S. Herbin. *Eléments pour la formalisation d'une reconnaissance active — application à la vision tri-dimensionnelle.* Thèse de mathématiques appliquées, Ecole Normale Supérieure de Cachan, Juillet 1997. In french.
16. D. Kazakos. Asymptotic error probability expressions for multihypothesis testing using multisensor data. *IEEE Transactions on Systems, Man, and Cybernetics*, **21**(5):1101–1114, 1991.
17. J. Köbler, U. Schöning, and J. Torn. *The Graph Isomorphism Problem: Its Structural Complexity.* Birkhauser, Cambridge, MA, 1993.
18. J. Koenderink. *Solid Shape.* MIT Press, Cambridge Ma, 1990.
19. J.J. Koenderink and A.J. van Doorn. The singularities of the visual mapping. *Biological Cybernetics*, **24**:51–59, 1976.
20. T. Lindeberg. Scale-space theory: A basic tool for analysing structures at different scales. *Journal of Applied Statistics*, **21**(2):224–270, 1994.
21. T. Lindeberg. *Scale-Space Theory in Computer Vision.* 1994.

22. S. Natarajan. Large deviations, hypotheses testing, and source coding for finite markov chains. *IEEE Transactions on Information Theory*, **31**(3):360–365, 1985.
23. S. Petitjean. The enumerative geometry of projective algebraic surfaces and the complexity of aspect graphs. *International Journal of Computer Vision*, **19**(3):261–287, 1996.
24. S. Petitjean, J. Ponce, and D. Kriegman. Computing exact aspect graphs of curved objects: algebraic surfaces. *International Journal of Computer Vision*, **9**(3):231–255, 1992.
25. R.P.N. Rao and D.H. Ballard. An active vision architecture based on iconic representations. *Artificial Intelligence*, **78**(1-2):461–505, October 1995.
26. J.H. Rieger. On the complexity and computation of view graphs of piecewise-smooth algebraic surfaces. Technical Report FBI_HH_M_228/93, Hamburg Univerität, 1993.
27. R.D. Rimey and C.M. Brown. Control of selective perception using bayes nets and decision-theory. *International Journal of Computer Vision*, **12**(2-3):173–207, April 1994.
28. M. Seibert and A.M. Waxman. Adapative 3D-object recognition from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **11**:107–124, 1992.
29. E Seneta. *Non-negative matrices and Markov chains*. Springer Verlag, New York, second edition, 1981.
30. L.G. Shapiro and M.S. Costa. Appearance-based 3D object recognition. In M. Hebert, J. Ponce, T.E. Boult, A. Gross, and D. Forsyth, editors, *Proc. International Workshop on 3-D Object Representation in Computer Vision, New York City, NY, USA, 5-7 Dec 1994*, volume 994 of *Lecture Notes in Computer Science*, pages 51–64. Springer Verlag, Berlin, 1995.
31. I. Shimshoni and J. Ponce. Probabilistic 3D object recognition. In *Proc. Fifth International Conference on Computer Vision, Cambridge, MA, USA, June 20-22*, pages 488–493, 1995.
32. K.A. Tarabanis, P.K. Allen, and R.Y. Tsai. A survey of sensor planning in computer vision. *IEEE Transactions in Robotics and Automation*, **11**(1):86–104, February 1995.
33. M. Tistarelli. Active space-variant object recognition. *Image and Vision Computing*, **13**(3):215–226, April 1995.
34. J.K. Tsotsos. On the relative complexity of active vs. passive visual search. *International Journal of Computer Vision*, **7**(2):127–141, January 1992.
35. T. van Effelterre. *Calcul exact du graphe d'aspect de solides de révolution*. Thèse de doctorat, Université de Rennes I, 1995.
36. S. Vinther and R. Cipolla. Active 3D object recognition using 3D affine invariants. In J-O. Eklundh, editor, *Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2-6*, number 801 in Lecture Notes in Computer Science, pages 15–24. Springer Verlag, Berlin, 1994.
37. D. Weinshall, M. Werman, and N. Tishby. Stability and likelihood of views of three dimensional objects. In J-O. Eklundh, editor, *Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2-6*, number 800 in Lecture Notes in Computer Science, pages 24–35. Springer Verlag, Berlin, 1994.
38. D. Wilkes and J.K. Tsotsos. Active object recognition. In *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Champaign, IL, June 15-18*, pages 136–141, 1992.