Stability and Likelihood of Views of Three Dimensional Objects

Daphna Weinshall[1], Michael Werman[1] and Naftali Tishby[1*]

Institute of Computer Science
The Hebrew University of Jerusalem
91904 Jerusalem, Israel
contact email: daphna@cs.huji.ac.il

Abstract. Can we say anything general about the distribution of two dimensional views of general three dimensional objects? In this paper we present a first formal analysis of the stability and likelihood of two dimensional views (under weak perspective projection) of three dimensional objects. This analysis is useful for various aspects of object recognition and database indexing. Examples are Bayesian recognition; indexing to a three dimensional database by invariants of two dimensional images; the selection of "good" templates that may reduce the complexity of correspondence between images and three dimensional objects; and ambiguity resolution using generic views.
We show the following results: (1) Both the stability and likelihood of views do not depend on the particular distribution of points inside the object; they both depend on **only** three numbers, the three second moments of the object. (2) The most stable and the most likely views are the **same** view, which is the "flattest" view of the object. Under orthographic projection, we also show: (3) the distance between one image to another does not depend on the position of its viewpoint with respect to the object, but **only** on the (geodesic) distance between the viewpoints on the viewing sphere. We demonstrate these results with real and simulated data.

1 Introduction

Model-based object recognition is often described as a two stage process, where indexing from the image into the database is followed by verification. However, using noisy images and large databases, the indexing stage rarely provides a single candidate, and the verification stage only reduces the ambiguity but cannot eliminate is altogether. Typically, therefore, we are left with a list of candidate objects, from which we should choose the best interpretation. This problem is demonstrated in Fig. 1, which could be the image of many different objects, all

* This research was sponsored by the U.S. Office of Naval Research under grant N00014-93-1-1202, R&T Project Code 4424341—01, and by the Israeli Science Foundation grant 202/92-2.

of which could possibly be retrieved by the recognition system. How do we decide which object this really is? is it a bagel? maybe a plate? neither. The task is easier when using more likely views of the object, such as those in Fig. 2.

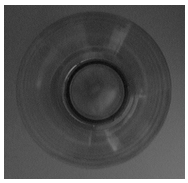

Fig. 1. Non generic (not probable) view of an object.

A plausible strategy is to select the model which obtains the highest confidence, or the highest conditional probability Prob(model/image). To accomplish this, we first rewrite the conditional probability as

$$\text{Prob(model/image)} = \text{Prob(image/model)}\frac{P_m}{P_i}$$

where P_m and P_i denote the prior probabilities of the model and image respectively. From this we see that optimal object recognition requires knowledge of the conditional distribution of images given models. Thus, for example, this likelihood measure is very small for the image of the water bottle shown in Fig. 1, and we therefore interpret the image as something else, such as a bagel or a plate.

Surprisingly, this important question of image likelihood has been (almost) totally neglected. There is a single exception, a study of the distribution of views of simple "objects", specifically planar angles, reported by Ben-Arie [2] and later expanded by Burns et al. [3]. These papers analyzed (via simulations) the changes in the appearance of angles from different points of views, and numerically identified stable images.

Can we say anything general about the distribution of two dimensional views of general three dimensional objects? In this paper we present a first formal analysis of this question. We first define the problem generally, connecting the concepts of stability and likelihood in the same framework. We then concentrate on geometry, to obtain (analytically) some simple and elegant characterizations, as well as some surprising properties, of image stability and likelihood of objects composed of localized features. These results are summarized below, in Section 1.1. Similar analysis should be done for sources of image variation other then geometry, such as lighting.

The theory developed here has many applications and can be used for object recognition in various ways, as described in Section 1.2. One result, where we show that the most stable view of an object is also its most likely view, has the following practical application: it implies that if we want to find and store the *most stable* view(s) of an object, we do NOT need to know the three-dimensional structure of the object; rather, we can expect to find this view by random sampling of the views of the object. This theory is also motivated by

and related to human perception, and some of the results reported here can be used to reinterpret psychophysical findings, as discussed below.

1.1 Characterization of views

Consider the viewing sphere, which is a sphere around the center of mass of the object. This sphere contains all possible viewing angles, or camera orientations relative to the object. We characterize each view V by two numbers:

ϵ-likelihood: the probability (or the area on the viewing sphere) over which views of the object are within ϵ of V (as pictures).

Υ-stability: the maximal error obtained when view V is compared to neighboring views less than Υ away (in geodesic distance) from V on the viewing sphere.

Υ-stability measures how stable a particular two dimensional view of a three dimensional object is with respect to change of camera position. ϵ-likelihood measures how often one should expect to see a particular view of a general object, if ϵ error is tolerated, and assuming known prior distribution on the viewing sphere (or viewing orientations). Each number provides a somewhat different answer to a similar question: how representative is a two dimensional view of a three dimensional object?

For objects composed of distinct features, this analysis of the viewing sphere can be carried out relatively simply thanks to the following observation, which is true within an aspect of the object[2]: *Given an object composed of any number of features, the three eigenvalues of the auto-correlation scatter matrix of the features' $3D$ coordinates are sufficient to compute the image differences between any two different views of the object.* Therefore, these three numbers fully characterize the stability and likelihood of any viewpoint.

For such objects we give in Section 3 explicit expressions for ϵ-likelihood and Υ-stability. We give expressions for the distance between any two views in terms of the three eigenvalues of the autocorrelation matrix. We show that *the "flattest" view is the most stable and the most likely.* Under orthographic projection we also demonstrate an elegant and surprising property of the viewing sphere: *viewpoints which are at the same geodesic distance from a certain view on the viewing sphere induce (very different) images that are at the same distance in image space.* In other words, if we fix a view V as the pole on the viewing sphere, all the viewpoints that are on the same latitude on the viewing sphere induce images which are at the same distance from the image of V.

1.2 What is it good for?

The characterization of views by stability or likelihood can be useful for various aspects of object recognition and spatial localization:

[2] We define an aspect as the set of views of the object in which the same features are visible.

Bayesian recognition and image understanding: As explained above, in order to select the most likely model from a set of models, each of which is a possible interpretation of an object in the scene, we need the conditional distribution of images given models. More generally, the probabilistic characterization of views, as defined below, measures how generic viewpoints are. In ambiguous cases, the interpretation which involves a more generic view may be preferable (see also [4]).

Indexing by invariants: To finesse correspondence, various algorithms look for indices which can be computed from $2D$ images, and directly point to the object (or a family of objects) in the database [5]. To be useful, such indices are typically invariant to some aspects of the imaging process. However, geometrical invariants for general $3D$ objects do not exist [3]. By identifying a set of "representative" $2D$ views of an object, such that any other image of the object is not too far from at least one image in this set, we can attach to each object a list of invariant indices which will have small errors.

Viewer-based representations: The three dimensional structure of objects can be represented in two fundamentally different ways: a two dimensional viewer-centered description, or a three dimensional object-centered description. In a viewer-centered description three dimensional information is not represented explicitly. In employing this approach, an object is represented by a list of $2D$ views, that were acquired during a familiarization period. A novel view of the object is recognized by comparing it to the stored views. A measure of image stability and likelihood can be used to select a set of "good" views for such a representation.

Correspondence by two dimensional template matching: Various recognition methods of $3D$ objects, such as alignment, require correspondence between a $2D$ image and a library of $3D$ models. Image to model correspondence (or indexing) is computationally difficult, and may require exponential searches. One solution is to use $2D$ templates for the direct matching of $2D$ images, which may reduce the complexity of search considerably from $O(n^3)$ to $O(dn^2)$, where d is the number of templates (see [1] for a discussion of algorithms for finding all such matches). The two dimensional templates are possibly grey-level images of the object, where distinctive features are used to determine stability and likelihood.

Our characterization will make it possible to select the "best" templates, which can be matched to the largest amount of different views with the smallest amount of error. Moreover, we will be able to identify local configurations which are particularly stable and therefore should be relied on more heavily during the initial stage of correspondence.

The rest of this paper is organized as follows: in Section 2 we define the above concepts more precisely. In Section 3 we show a simple computational scheme to compute viewpoint characterizations for the case of an object composed of a set of $3D$ features, and describe the basic results. In Section 4 we demonstrate these results with real and simulated data.

Fig. 2. Left: a not very likely view of an object; right: a likely view of a water bottle.

2 Definitions

2.1 The viewing sphere

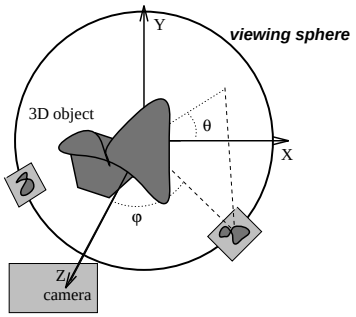

Fig. 3. The viewing sphere of a $3D$ object. Two views on the viewing sphere, obtained by some combination of rotations ϑ and φ, are illustrated.

We fix a coordinate system attached to the camera, where Z is the optical axis of the camera (assumed orthogonal to the image plane). The object is assumed fixed, and the camera (with the coordinate system) rotates around it on the viewing sphere. The viewing sphere is an imaginary sphere around the centroid of the object (see Fig. 3), representing all the possible different viewpoints of the object.

The viewing sphere takes into account deformations in the appearance of an object which are due solely to its $3D$ structure and orientation in space, when the camera is allowed to translate and rotate relative to the object. We assume weak perspective projection and therefore translations of the camera can be ignored if by default images are centered around the center of mass of an object. Therefore without loss of generality, the center of rotation is assumed to be the centroid of the object. With this convention the viewing sphere describes all the possible different images of an object, since there is a 1-1 mapping between a viewpoint and a feasible image.

With this definition, a view of the object corresponds to a point on the viewing sphere, which is completely defined by two angles of rotation. If we assume that all viewing angles are equally likely, areas on the viewing sphere correspond to probability or likelihood. In the following (see also Fig. 3), the viewing sphere is parameterized by two angles: rotation φ around the Z axis followed by a rotation ϑ around the X axis. This defines a spherical coordinate

system whose pole is the optical axis of the camera (the Z axis), and where φ is the azimuth (longitude) and ϑ is the elevation (colatitude).

2.2 Stability and likelihood of views

Consider an object $\mathcal{O}$ and a point on the viewing sphere of $\mathcal{O}$ denoted by $V_{\vartheta,\varphi}$. The range $\vartheta \in [0, \frac{\pi}{2}]$, $\varphi \in [0, 2\pi]$ gives a parameterization of half the viewing sphere in spherical coordinates whose pole is the Z-axis, where φ is the azimuth and ϑ is the elevation.

Let $V'_{\vartheta,\varphi,\alpha,\beta}$ denote another view, corresponding to a rotation in spherical coordinates on the viewing sphere, where the point ϑ, φ is now the pole, $\alpha \in [0, \frac{\pi}{2}]$ the elevation, and $\beta \in [0, 2\pi]$ the azimuth. The distance from $V_{\vartheta,\varphi}$ to $V'_{\vartheta,\varphi,\alpha,\beta}$ on the viewing sphere is parameterized by the elevation angle α. Let $d(\vartheta, \varphi, \alpha, \beta)$ denote the image distance, as defined in Section 2.4, between the images obtained from view V and view V'.

For each view $V = [\vartheta, \varphi]$ we measure the following:

Υ-stability: the maximal error (difference) d, when compared to other views on the viewing sphere separated from it by an elevation $\alpha < \Upsilon$: $\max_{\alpha \leq \Upsilon, \beta} d(\vartheta, \varphi, \alpha, \beta)$

ϵ-likelihood: the measure (on the viewing sphere) of the set $\{(\alpha, \beta) |$ such that $d(\vartheta, \varphi, \alpha, \beta) \leq \epsilon\}$.

We select the view V which represents an aspect of the object according to one of the following criteria:

Most stable view: the view $V = [\vartheta, \varphi]$ which for all bounded movements of the viewing point from V the image changes the least:

$$\min_{\vartheta,\varphi} \max_{\alpha \leq \Upsilon, \beta} d(\vartheta, \varphi, \alpha, \beta) \tag{1}$$

Most likely view: the view $V = [\vartheta, \varphi]$ that has the largest number of views that as are images close to it:

$$\max_{\vartheta,\varphi} Measure(\{(\alpha, \beta) | d(\vartheta, \varphi, \alpha, \beta) \leq \epsilon\}) \tag{2}$$

2.3 Images of objects with fiducial points

We consider objects composed of n three dimensional fiducial points. Let $\{\hat{\mathbf{p}}_i = (\hat{x}_i, \hat{y}_i, \hat{z}_i)\}_{i=1}^n$ denote the coordinates of the object features in the camera coordinate system in $\mathcal{R}^3$. A three-dimensional representation of the object is the $3 \times n$ matrix $\hat{\mathbf{P}}$, whose i-th column is $\hat{\mathbf{p}}_i$, the vector describing the world coordinates of the i-th feature of the object.

An image of the object is obtained by a rigid transformation (of the object or the camera), followed by weak perspective (or scaled orthographic) projection from three dimensional space to the two dimensional image. An image of the object is therefore the set of n image points $\{\mathbf{p}_i = (x_i, y_i)\}_{i=1}^n$. An equivalent

representation of the image is the $2 \times n$ matrix $\mathbf{P}$, whose i-th column is the image coordinates of the i-th feature of the object. The use of matrix $\mathbf{P}$ to represent an image of the object implies a correspondence between the image features and the object features, where different correspondences lead to permutations of the matrix' columns.

2.4 How to compare two images

Given two images, or the two matrices $\mathbf{P}$ and $\mathbf{Q}$, the question of comparing them is equivalent to matrix comparison. We are using the "usual" metric, which is the Frobenius norm of the difference matrix, and which is the same as the Euclidean distance between points in the images:

$$||\mathbf{P} - \mathbf{Q}||_F^2 = \sum (\mathbf{P}[i,j] - \mathbf{Q}[i,j])^2 = tr[(\mathbf{P} - \mathbf{Q}) \cdot (\mathbf{P} - \mathbf{Q})^T] \tag{3}$$

(tr denotes the trace of a matrix). Henceforth we will omit the subscript F, and a matrix norm will be the Frobenius norm.

Before taking the norm of the difference between the images, we want to remove differences which are due to irrelevant effects, such as the size of the image (which is arbitrary under scaled orthography) or the exact location of the object (e.g., due to an arbitrary translation and rotation of the object in the image). In particular, we may want to consider as equivalent all images obtained from each other by the group of $2D$ **similarity** transformations, which includes $2D$ rotations, translations, and scale. The equivalence under similarity transformation is necessary, since under weak perspective projection, images that differ by image scale, rotation or translation can be obtained from the same object, and should therefore be considered the *same image*.

It can be readily shown that the optimal translation when measuring distance by sum of square distances, under the similarity equivalence, puts the centroid of the object in the origin of the image. We therefore assume w.l.g. that the images are centered on the centroid of the object, so that the first moments of the object are 0. In [7] we define image distance measures, which satisfy all the properties of a metric, and which compare the images $\mathbf{P}$ and $\mathbf{Q}$ while taking into account the desired image equivalence discussed above. We get the following expression:

$$D^2(\mathbf{P}, \mathbf{Q}) = 1 - \frac{||\mathbf{Q}\mathbf{P}^T||^2 + 2det(\mathbf{Q}\mathbf{P}^T)}{||\mathbf{P}||^2||\mathbf{Q}||^2} \tag{4}$$

2.5 The "flattest" view

Let $R_{\vartheta\varphi}$ denotes a $3D$ rotation in spherical coordinates around the pole $(0,0,1)$, with azimuth φ and elevation ϑ. Consider the 3×3 symmetric autocorrelation scatter matrix of the object:

$$S = \hat{\mathbf{P}}\hat{\mathbf{P}}^T$$

The scatter matrix of the object at view $V_{\vartheta,\varphi}$, obtained by a rotation $R_{\vartheta\varphi}$ on the viewing sphere away from the initial view, is:

$$S(V) = R_{\vartheta\varphi}\hat{\mathbf{P}}\hat{\mathbf{P}}^T R_{\vartheta\varphi}^T = R_{\vartheta\varphi} S R_{\vartheta\varphi}^T$$

Definition 1. The **flattest view** is the view V_f whose scatter matrix $S(V_f)$ is diagonal, and where the eigenvalues (the diagonal elements) are ordered in decreasing order.

It is straightforward to compute the orthogonal matrix L such that $S = L^T D L$, where D is diagonal with diagonal elements in decreasing order (e.g., by computing the SVD of the symmetric matrix S). L is the rotation matrix which rotates the object from its original orientation to V_f . Henceforth we will assume w.l.g. that the coordinate system is initially oriented so that $V_{0,0} = V_f$. Let S_0 denote the diagonal scatter matrix at $V_{0,0}$:

$$S_0 = \hat{\mathbf{P}}_0\hat{\mathbf{P}}_0^T = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix}$$

where $a \geq b \geq c$.

3 Viewpoint characterization

Consider an object which is characterized by n fiducial points in three dimensional space, $\mathcal{O} = \hat{\mathbf{p}}_1, \hat{\mathbf{p}}_2, ..., \hat{\mathbf{p}}_n$. Let $\hat{\mathbf{P}}$ denote the $3 \times n$ matrix whose i-th column is $\hat{\mathbf{p}}_i$.

3.1 Stability and likelihood at each view

As defined in Section 2.2, let $V_{\vartheta,\varphi}$ denote a point on the viewing sphere of object $\mathcal{O}$, and let $V'_{\vartheta,\varphi,\alpha,\beta}$ denote another view of $\mathcal{O}$. (Recall that the distance from $V_{\vartheta,\varphi}$ to $V'_{\vartheta,\varphi,\alpha,\beta}$ on the viewing sphere is parameterized by the elevation angle α.) Let $d(\vartheta,\varphi,\alpha,\beta) = D(V,V')$ denote the image distance, defined in Section 2.4, between the appearance of object $\mathcal{O}$ from view V and its appearance from view V'. We can show that:

Result 1: *$d(\vartheta,\varphi,\alpha,\beta)$ depends only on the diagonal matrix S_0, regardless of the number of features in $\mathcal{O}$ or their distribution in space. We therefore denote the distance by $d_{a,b,c}(\vartheta,\varphi,\alpha,\beta)$.*

We computed $d_{a,b,c}(\vartheta,\varphi,\alpha,\beta)$ by substituting V and V' into Eq (4), to get:

$$D^2 = d^2_{a,b,c}(\vartheta,\varphi,\alpha,\beta) = \frac{(1-\cos(\alpha))\,(abs_1 + acs_2 + bcs_3)}{u(at_1 + bt_2 + ct_3)} \qquad (5)$$

where

$$
\begin{aligned}
s_1 &= 1 - 2\,\cos(\vartheta)^2\cos(\alpha) + 2\,\cos(\vartheta)\sin(\alpha)\sin(\vartheta)\cos(\beta) + \cos(\alpha)\\
s_2 &= 1 - 2\,\cos(\alpha)\cos(\varphi)^2\sin(\vartheta)^2 - 2\,\cos(\beta)\cos(\varphi)^2\sin(\alpha)\sin(\vartheta)\cos(\vartheta) +\\
&\quad 2\,\sin(\beta)\sin(\alpha)\sin(\varphi)\cos(\varphi)\sin(\vartheta) + \cos(\alpha)\\
s_3 &= 1 + 2\,\cos(\alpha)\sin(\varphi)^2\cos(\vartheta)^2 - 2\,\sin(\beta)\sin(\alpha)\sin(\varphi)\cos(\varphi)\sin(\vartheta) -\\
&\quad 2\,\cos(\beta)\sin(\varphi)^2\sin(\alpha)\sin(\vartheta)\cos(\vartheta) - \cos(\alpha) + 2\,\cos(\alpha)\cos(\varphi)^2\\
u &= a\left(1 - \sin(\varphi)^2\sin(\vartheta)^2\right) + b\left(1 - \cos(\varphi)^2\sin(\vartheta)^2\right) + c\sin(\vartheta)^2\\
t_1 &= -2\,\cos(\alpha)\cos(\beta)\sin(\alpha)\cos(\vartheta)\sin(\vartheta)\sin(\varphi)^2 - \sin(\vartheta)^2\cos(\alpha)^2\sin(\varphi)^2 -\\
&\quad 2\,\cos(\varphi)\sin(\varphi)\cos(\vartheta)\sin(\alpha)^2\cos(\beta)\sin(\beta) - \cos(\varphi)^2\sin(\alpha)^2\sin(\beta)^2 + 1 -\\
&\quad \cos(\vartheta)^2\sin(\alpha)^2\cos(\beta)^2\sin(\varphi)^2 - 2\,\cos(\varphi)\cos(\alpha)\sin(\beta)\sin(\alpha)\sin(\vartheta)\sin(\varphi)\\
t_2 &= 2\,\cos(\varphi)\sin(\varphi)\cos(\vartheta)\sin(\alpha)^2\cos(\beta)\sin(\beta) + \sin(\varphi)^2\sin(\alpha)^2\cos(\beta)^2 +\\
&\quad \cos(\varphi)^2\sin(\alpha)^2 - 2\,\cos(\alpha)\cos(\beta)\sin(\alpha)\cos(\vartheta)\sin(\vartheta)\cos(\varphi)^2 +\\
&\quad \cos(\vartheta)^2\cos(\alpha)^2\cos(\varphi)^2 + 2\,\cos(\varphi)\cos(\alpha)\sin(\beta)\sin(\alpha)\sin(\vartheta)\sin(\varphi) +\\
&\quad \cos(\alpha)^2\sin(\varphi)^2 - \cos(\vartheta)^2\sin(\alpha)^2\cos(\beta)^2\cos(\varphi)^2\\
t_3 &= 1 + 2\,\cos(\vartheta)\sin(\alpha)\sin(\vartheta)\cos(\alpha)\cos(\beta) - \cos(\vartheta)^2\cos(\alpha)^2 -\\
&\quad \sin(\vartheta)^2\sin(\alpha)^2\cos(\beta)^2
\end{aligned}
$$

3.2 The most stable and likely view

We substituted $d_{a,b,c}(\vartheta, \varphi, \alpha, \beta)$ into Eqs (1),(2), to compute the stability and likelihood measures numerically for various objects, characterized by different parameters a, b, c, and for various likelihood and stability thresholds ϵ and Υ. The simulations lead us to conjecture the following result:

Result 2: *The flattest view $V_f = V_{0,0}$ is both the Υ-stable view and the ϵ-likely view for all Υ and ϵ, and for every object parameterized by $[a, b, c]$.*

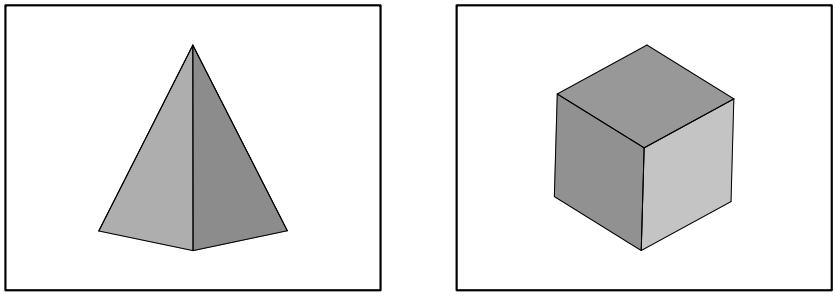

Fig. 4. V_f of a square pyramid and a cube.

Examples:

We computed V_f for two simulated familiar objects: a cube and a pyramid. For the cube we consider a certain aspect where 7 vertices are visible. For the pyramid we consider an aspect where 4 vertices are visible. Fig. 4 shows the V_f of each of these objects.

4 Orthographic projection:

In order to match images to the projections of a model, $2D$ similarity normalization was used. There are cases in which it is more appropriate to use 2D affine normalization or orthographic projection. In these cases the two results described in the previous section still hold (see [6]).

Under orthographic projection, the scale of the image is known and we want to avoid normalization, but rather compare the model to the image as is. We therefore take the difference between the given image and the model, where the model is aligned with an affine transformation to the image, and without any manipulation permitted to be applied to the image. If we denote the model $\mathbf{P}$ and the new image $\mathbf{Q}$, we get the following orthographic distance measure, which replaces the similarity metric given in Eq (4):

$$D^2(\mathbf{P}, \mathbf{Q}) = tr[\mathbf{Q}^T \mathbf{Q}(I - \mathbf{P}^+\mathbf{P})]$$

(I denotes the $n \times n$ unity matrix, and $\mathbf{P}^+$ denotes the pseudo-inverse of $\mathbf{P}$).

Eq (5) now becomes surprisingly simple:

$$D^2 = d_{a,b,c}(\vartheta, \varphi, \alpha, \beta) = \frac{abc\sin(\alpha)^2}{ac\sin(\vartheta)^2\cos(\varphi)^2 + bc\sin(\vartheta)^2\sin(\varphi)^2 + ab\cos(\vartheta)^2}$$

and the following result immediately follows:

Result 3: *For all ϑ, φ, the distance between $V_{\vartheta,\varphi}$ and $V_{\vartheta,\varphi,\alpha,\beta}$ depends only on the geodesic distance α and does not depend on the azimuth β, although for different β the views $V_{\vartheta,\varphi,\alpha,\beta}$ are not affine equivalent.*

In other words, if we fix a view V as the pole on the viewing sphere, all the viewpoints that are on the same latitude on the viewing sphere induce images which are at the same distance from the image of V.

Examples:

To demonstrate the above results, we took an image P_{top} of a toy tiger from an arbitrary angle, and then took a sequence of images, (P_1, P_2, Q_1, Q_2, Q_3, S), at other orientations (see Fig. 5). We did not measure the orientations, but we know that the images marked by (P_i) were taken at the same elevation relative to P_{top}, and that the images marked by (Q_i) were taken at the same elevation relative to P_{top}. The elevation of the (P_i) images was smaller than the elevation of the (Q_i) images, which in turn was smaller than the elevation of image S.

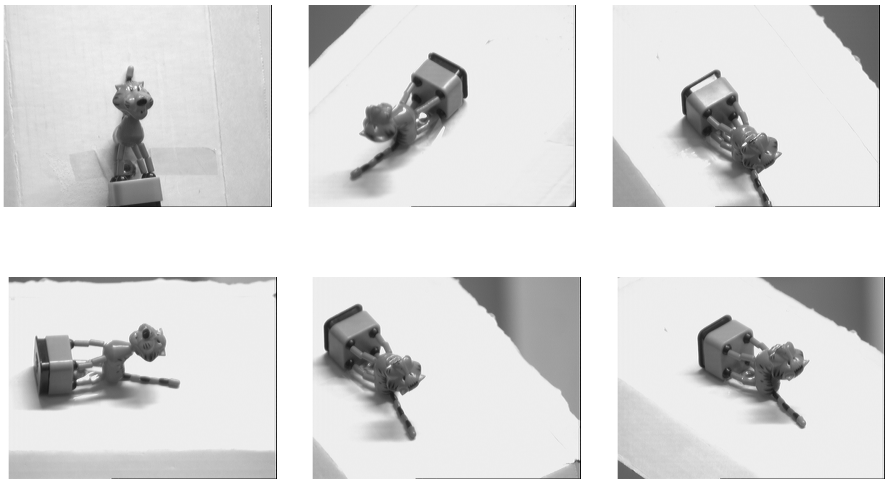

Fig. 5. The pictures used in the experiment: first row, from left to right – P_{top}, P_1, P_2; second row, from left to right – Q_1, Q_2, S.

P_1	P_2	Q_1	Q_2	Q_3	S
0.125	0.163	0.195	0.233	0.205	0.34
6352	7122	11045	10964	10557	15312

Table 1. The distances between the real images shown in Fig. 5. The first row gives the similarity (metric) distance, and the second row gives the orthographic distance.

Table 1 gives the distances of all the pictures from P_{top}, (using ears, eyes, knees, tail and nose as features). As can be seen from the data, the distances between images depend monotonically on the elevation, and the orthographic distance does not depend on the azimuth.

5 Discussion

The analysis and results described above have many applications for geometry-based object recognition, as discussed in the introduction:

– It provides the basic tools for object recognition from noisy images and large databases, giving a measure to select the model that best fits the data from a list of candidates obtained by "traditional" indexing and verification. More generally, it can be used with a general Bayesian image interpretation approach to select the most likely interpretation of a scene.

- It gives the framework within which an invariant recognition scheme such as geometric hashing can be generalized to three-dimensional objects, by storing invariant indices of a list of representative views. This framework also provides a measure for the selection of "good" templates, for the purpose of model to image corresponednce, a computationally hard problem. Here we see the significance of result 2 above, since it tells us that if we want to select the most stable view of an object (say, for a template), we need not necessarily compute the complete three dimensional structure of the object. Rather, since the most stable view is also the most likely view, we may attempt to obtain this view using a learning algorithm that is given a random sample of the views of the object.
- It provides the basic tools to define and analyze *aspect* graphs such that the different aspects are not necessarily topologically distinct, rather they differ metrically. In this way we can choose a representative set of viewpoints so that we cover every possible view of an object upto some error ϵ, and we can use the neighborhood information in the graph in order to track a moving object.

References

1. E. M. Arkin, K. Kedem, J. S. B. Mitchell, J. Sprinzak, and M. Werman. Matching points into pairwise disjoint noise regions: Combinatorial bounds and algorithms. *ORSA Journal on Computing, special issue on computational geometry*, 1992.
2. J. Ben-Arie. The probabilistic peaking effect of viewed angles and distances with application to 3-d object recognition. *T-PAMI*, pages 760–774, 1990.
3. J.B. Burns, R. Weiss, and E. Riseman. View variation of point-set and line segment features. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(1):51–68, 1993.
4. W. T. Freeman. Exploiting the generic view assumption to estimate scene parameters. In *Proceedings of the 4th International Conference on Computer Vision*, pages 347–356, Berlin, Germany, 1993. IEEE, Washington, DC.
5. Y. Lamdan and H. Wolfson. Geometric hashing: a general and efficient recognition scheme. In *Proceedings of the 2nd International Conference on Computer Vision*, pages 238–251, Tarpon Springs, FL, 1988. IEEE, Washington, DC.
6. D. Weinshall, M. Werman, and N. Tishby. Stability and likelihood of views of three dimensional objects. TR 94-1, Hebrew University, 1993.
7. M. Werman and D. Weinshall. Similarity and affine distance between point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):810–814, 1995.

This article was processed using the LaTeX macro package with LLNCS style