A computational theory of canonical views *

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

Abstract

When computing 3D interpretations of noisy 2D images, many interpretations are often plausible. We describe a general framework for resolving such ambiguities in object recognition and reconstruction using maximum likelihood estimation. To this end we define measures of likelihood and stability of interpretations. We use these measures to select the "best" interpretation from the set of possible interpretations in object recognition or in reconstruction tasks. These measures also give a practical way to evaluate how "generic" views are, and identify "characteristic" views for the purpose of building multi-view representations. To demonstrate the usefulness and generality of this framework, we computed the proposed stability and likelihood measures using 4 different kinds of image matching algorithms, matching: (1) feature points, (2) angles, (3) occluding contours of smooth surfaces, and (4) shaded images of smooth surfaces.

1 Introduction

Consider the three images shown in the top row of Fig. 1. Given 3 matching objects in the database (shown in the bottom row of Fig. 1), a recognition system is asked to match an object to each image. The images were produced so that each one perfectly matches a single object, a match shown with wide white arrows in Fig. 1. If geometrical consistency was the only cue used by the recognition system, then this would be its output. However, to a human looking at those images, the above interpretation does not seem right. Rather, a human would prefer the interpretation shown with thin black arrows in Fig. 1.

What is the source of the difference? Should a recognition system, which is not concerned with humans judging its performance, pay any attention to this discrepancy? This computational issue is addressed in this paper. We show that if a factor other than geometrical feasibility is taken into consideration during the recognition process, in particular if the robustness of the interpretation is measured, then the output of the recognition system will match human performance in such cases as illustrated in Fig. 1. Thus, human judgment is justified by a valid computational argument. We propose to use this computational argument to augment any recognition system, in order to achieve better (and more robust) overall performance.

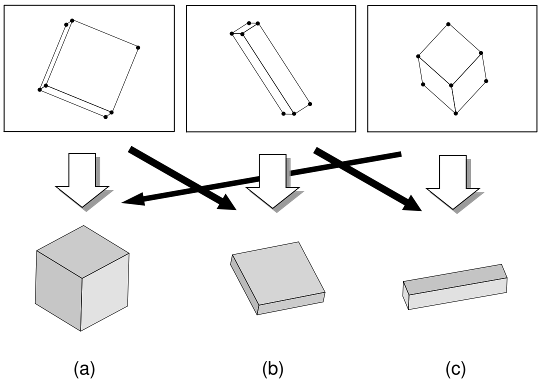

Figure 1: Three polyhedral objects: (a) a cube of dimensions $1 \times 1 \times 1$ cm., (b) a flat box of dimensions $5 \times 5 \times 1$ cm., (c) an elongated box of dimensions $5 \times 1 \times 1$ cm. Top: three images of these objects, obtained from special viewpoints; the extracted features are shown with dark circles. Bottom: the same polyhedral objects viewed from a "typical" viewpoint. A geometry-based recognition system would "correctly" recognize each image in the top row as an instance of the object illustrated in the bottom row (this matching is shown by thick white arrows). A maximum likelihood recognition system would recognize the images in the top row as: left = flat box, middle = elongated box, right = cube (this matching is shown by thin black arrows.)

In comparison, most of the recent efforts towards reconstruction and 3D object recognition address the geometrical issue: which 3D interpretation (or which object) is described in a given 2D image. Once this problem is solved by a reconstruction or a recognition algorithm, the algorithm would typically provide a set of possible 3D interpretations (or objects) of the scene, all sufficiently consistent with the image.

In the present paper we address the problem of defining criteria to choose between interpretations that are equally consistent with the data. Our first measure of the quality of a 3D interpretation is the *view likelihood* function, measuring the conditional probability of an image given the 3D interpretation. Our second measure of

*This research was sponsored by ARPA through the U.S. Office of Naval Research under Grant N00014-93-1-1202, R&T Project Code 4424341—01.

the quality of a $3D$ interpretation is the *view stability* function, measuring how much the image changes as the viewpoint of the object is perturbed (see review of previous work next). These measures quantify the notions of "generic" and "characteristic" views.

We first define and compute these functions for general objects, showing several interesting and rather general properties. The functions depends on the particular $2D$ distance measure used to compare two images with each other, and can be readily evaluated given such a distance. We show that the proposed *view stability* function is similar to the *view likelihood* function, up to scaling by a term reflecting the prior distribution of viewpoints. Using these measures in the example of Fig. 1, the interpretations preferred by humans (shown with thin black arrows in Fig. 1) are by far the best (our measure of robustness is at least twice as high for each of these interpretations). The work described here was reported in [10].

2 Previous work

Bayesian image understanding: Freeman [4] suggested using a measure of view stability in the interpretation of ambiguous scenes. This scheme differs from ours in two important ways: First, it is Bayesian (thus requiring prior distributions) whereas our scheme is based on maximum likelihood. Second, Freeman did not discuss how to compute measures of view "genericity" directly from images and models. Thus the approaches can be easily put together, where our likelihood measures take the role of prior distributions in Freeman's Bayesian scheme.

View likelihood: Image likelihood of angles was computed using numerical simulations by Ben-Arie [1] and Burns et al. [2]. Dickinson et al. [3] empirically found the more likely views of particular objects decomposed into geons. In this earlier work, the analysis of likelihood was carried out for specific objects only. In comparison, our analysis directly gives the stability and likelihood profiles of general objects. A more significant problem with this earlier work is that they ignore the distortion created by the transformation from the viewing sphere to the image space, and thus give the wrong results. For example, as discussed in Section 4.2, the real peaking effect is much stronger than that obtained by the simulations reported in [1, 2].

View stability: Some studies proposed to measure stability via the Lie derivatives of the group of transformations describing the motion of the camera [6]. In most of this earlier work, the analysis of stability rarely went beyond the basic definitions (which were different from our definition).

3 General theory

The viewing sphere is an imaginary sphere around the centroid of the object, representing all the possible different viewpoints of the object. Under weak perspective projection, a view (or image) of the object corresponds to a point on the viewing sphere, which is completely defined by two angles. Let the angles ϑ, φ (elevation and azimuth respectively) parameterize the viewing sphere for the first view $V_{\vartheta,\varphi}$. Let the angles δ, ψ parameterize the viewing sphere for the second view $V_{\delta,\psi}$. The second parameterization is chosen relative to the first view, so that δ is the geodesic distance on the viewing sphere between the first and second viewpoints.

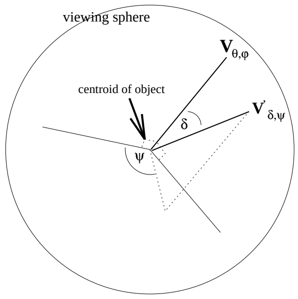

Figure 2: Two views on the viewing sphere, $V_{\vartheta,\varphi}$ and $V_{\delta,\psi}$.

3.1 Likelihood and Stability of views

We first define view likelihood and stability:

Definition 1: *The likelihood $l(\vartheta,\varphi)$ of view $V_{\vartheta,\varphi}$ is:*
$l(\vartheta,\varphi) = \lim_{\varepsilon\to 0} \mathrm{Prob}(V_{\vartheta,\varphi} \text{ observed}/_{\text{measurement error } \varepsilon})$

Definition 2: *The stability $s(\vartheta,\varphi)$ of view $V_{\vartheta,\varphi}$ is:*

$$s^{-1}(\vartheta,\varphi) = \lim_{\|S\|\to 0} \frac{< \text{average image difference over S} >}{\|S\|}$$

where S is an area on the viewing sphere around $V_{\vartheta,\varphi}$.

Let $d(\vartheta,\varphi,\delta,\psi)$ denote the *image distance* between the images obtained from view $V_{\vartheta,\varphi}$ and view $V_{\delta,\psi}$; $d(\vartheta,\varphi,\delta,\psi)$ may be any $2D$ difference function. Let $d(\vartheta,\varphi,\delta,\psi) = D(\vartheta,\varphi,\psi)\delta + O(\delta^2)$ ($D(\vartheta,\varphi,\psi)$ is the derivative of $d()$ with respect to δ in the direction ψ), and assume $D(\vartheta,\varphi,\psi) > 0$. The likelihood $l(\vartheta,\varphi)$ and stability $s(\vartheta,\varphi)$ of view $V_{\vartheta,\varphi}$ of a given object can be shown to be (the derivation is omitted):

$$\begin{aligned}
\tilde{l}(\vartheta,\varphi) &= prior(\vartheta,\varphi)\frac{1}{2}\int_0^{2\pi} \frac{1}{D^2(\vartheta,\varphi,\psi)}d\psi \\
l(\vartheta,\varphi) &= \frac{\tilde{l}(\vartheta,\varphi)}{\int_0^{2\pi}\int_0^{\frac{\pi}{2}} \tilde{l}(\vartheta,\varphi) sin(\vartheta) d\vartheta d\varphi} &\quad (1)\\
s(\vartheta,\varphi) &= \frac{1}{2\pi}\int_0^{2\pi} \frac{1}{D^2(\vartheta,\varphi,\psi)}d\psi &\quad (2)
\end{aligned}$$

Thus:

$$l(\vartheta,\varphi) = \frac{s(\vartheta,\varphi)prior(\vartheta,\varphi)}{\int_0^{2\pi}\int_0^{\frac{\pi}{2}} s(\vartheta,\varphi)prior(\vartheta,\varphi)\sin(\vartheta)d\vartheta d\varphi} \quad (3)$$

We are also interested in 2 qualitative characterizations of views:

Most likely view: $V_{\bar{\vartheta},\bar{\varphi}}$ which satisfies

$$l(\bar{\vartheta},\bar{\varphi}) = \max_{\vartheta,\varphi} l(\vartheta,\varphi) = \max_{\vartheta,\varphi} \tilde{l}(\vartheta,\varphi) \quad (4)$$

Most stable view: $V_{\bar{\vartheta},\bar{\varphi}}$ which satisfies

$$s(\bar{\vartheta},\bar{\varphi}) = \max_{\vartheta,\varphi} s(\vartheta,\varphi) \quad (5)$$

It follows from (3):

Result 1: *If the prior distribution of viewpoints of a given object is uniform, namely, all views are equally likely, then:*

1. *The view stability and likelihood functions are the same up to scaling.*
2. *The* **most stable** *and* **most likely** *views are the same view.*

4 Special cases

Below we give the stability function $s(\vartheta,\varphi)$ for different image metrics, from which the likelihood is obtained using (3):

4.1 Fiducial points:

Given objects composed of feature points, we use a distance function which measures the distance between the projected features up to similarity transformation in the image [11]. Let $a \geq b \geq c$ denote the 3 principal moments of the object (the 3 eigenvalues of the autocorrelation matrix of the feature points). It can be shown that $s_{sim}(\vartheta,\varphi)$ equals:

$$\frac{(a(1-\sin^2\varphi\sin^2\vartheta)+b(1-\cos^2\varphi\sin^2\vartheta)+c\sin^2\vartheta)^2}{bc(1-\sin^2\varphi\sin^2\vartheta)+ac(1-\cos^2\varphi\sin^2\vartheta)+ab\sin^2\vartheta} \quad (6)$$

Similar results are obtained using image distances up to affine transformations, where $s_{aff}(\vartheta,\varphi)$ equals:

$$\frac{(ab\cos^2\vartheta+ac\cos^2\varphi\sin^2\vartheta+bc\sin^2\varphi\sin^2\vartheta)^{\frac{3}{2}}}{abc} \quad (7)$$

The coordinate system on the viewing sphere is chosen so that $V_{\vartheta=0,\varphi=0}$ is the flattest view of the object, when the $3D$ object has its minimal spread along the viewing direction.

It follows from (6) (e.g., by differentiation), that:

Result 2: *[Characteristic views:]*

- *The most stable view is unique, and it is the flattest view $V_{\vartheta=0,\varphi=0}$.*
- *The stability of the flattest view is $\frac{a+b}{c}$.*
- *The stability at viewpoint $V_{\vartheta,\varphi}$ decreases monotonically with its geodesic distance from the flattest view ϑ.*

All the above applies to the most likely view as well, if a uniform prior distribution of viewpoints is assumed.

Corollary 3: *Given a $2D$ image, and without any additional information about the $3D$ structure, the most likely interpretation is the flattest view, implying that the image represents a fronto-parallel flat object.*

This corollary justifies a heuristic, which assigns depth 0 to points whose depth is unknown, as the optimal decision based on the available information. This heuristic is commonly used in the computer vision literature, and apparently by human vision as well.

As an example, the flattest views - or the most likely views - of 2 non-transparent objects composed of feature points are given in Fig. 3.

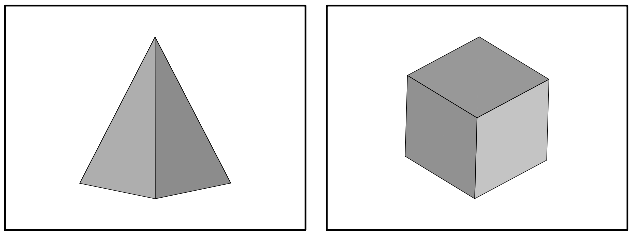

Figure 3: The flattest views of two opaque objects: a square pyramid and a cube.

4.2 Planar angles:

Given an angle α, the $2D$ distance measure we use is the difference between the projected values of the angle in the 2 images. It can be shown that $s_{ang}(\vartheta,\varphi)$ equals:

$$\sqrt{1-\sin^2\theta\sin^2\phi}\,\sin^{-1}\theta\sqrt{1-\sin^2\theta\ (\cos\alpha\sin\phi+\sin\alpha\cos\phi)^2} \quad (8)$$

Here once again we use a coordinate system on the viewing sphere where $V_{\vartheta=0,\varphi=0}$ is the flattest (fronto-parallel) view of the angle.

As with the feature points metric, it is easy to show the following result:

Result 4:

- *The most stable view is unique, and it is the flattest fronto-parallel view $V_{\vartheta=0,\varphi=0}$.*
- *The stability of the flattest view is ∞ (but the probability is finite).*
- *The stability at viewpoint $V_{\vartheta,\varphi}$ decreases monotonically with its geodesic distance from the flattest view ϑ.*

All the above applies to the most likely view as well, if a uniform prior distribution of viewpoints is assumed.

Using (8), we numerically computed the marginal distribution of the appearance of an angle α in the image, by summing over all viewpoints from which the angle looks the same. The distribution of the appearance of the projected angle β is plotted in Fig. 4, for $\alpha = 11.25^o,\ 22.5^o,\ 45^o,\ 90^o$.

[1, 2] counted the number of views from which a given angle α is seen as some $\beta \in [0,\pi]$ by uniformly scanning the viewing sphere. Our analysis shows that this enumeration does not measure correctly how often α is seen as β; to answer this question, the viewing sphere should not be scanned uniformly, but rather each view should be weighted by the view likelihood computed above. This is because the view likelihood, in effect, measures the size change between a small area element on the viewing

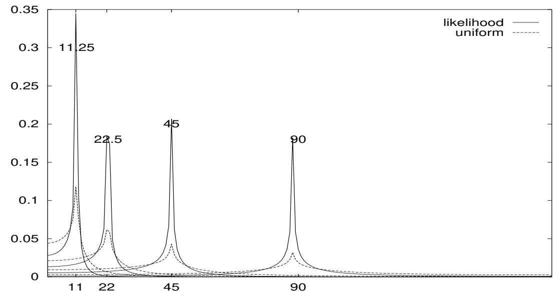

Figure 4: (a) The solid curves describe the likelihood that angle α is seen as β (assuming a uniform prior), plotted as a function of $\beta \in [0..180^o]$. The dotted curves show the same measure obtained by the calculation method of [2]. Four plots are given for each method: $\alpha = 11.25^o$ - the likelihood peaks at 11.25^o, $\alpha = 22.5^o$ - the likelihood peaks at 22.5^o, $\alpha = 45^o$ - the likelihood peaks at 45^o, $\alpha = 90^o$ - the likelihood peaks at 90^o.

sphere to the corresponding area element on the manifold of images. In [1, 2] the "peaking" effect was observed, where an angle α is seen most often as $\beta = \alpha$; repeating this calculation with the correct weight given above, we get the true peaking effect which is much more pronounced (Fig. 4).

4.3 Occluding contours of smooth objects

Given general smooth $3D$ objects, one of the major clues used for recognizing them are their contours, which in some impoverished situations is the only information on the $3D$ object in the picture. Using the method of matching the point distance (least squares up to similarity transformation) between curvature extrema, we can compute the distance between two pictures (note that the points in the two pictures are not necessarily projected from the same $3D$ points).

For an ellipsoid whose principal axes are u, v, w, and where the main axes of the ellipsoid are along the X, Y, Z axes respectively at view $V_{\vartheta=0,\varphi=0}$, $s_{occ}(\vartheta, \varphi)$ equals:

$$\frac{\frac{\left(uv\sin^2\vartheta + vw\left(1-\sin^2\varphi\sin^2\vartheta\right) + uw\left(1-\cos^2\varphi\sin^2\vartheta\right)\right)}{\sin\vartheta}}{\sqrt{\frac{v\cos^2\varphi\sin^2\vartheta + u\sin^2\varphi\sin^2\vartheta + w\cos^2\vartheta}{\cos^2\varphi\sin^2\vartheta\sin^2\varphi(u-v)^2 + \cos^2\vartheta\sin^2\varphi(u-w)^2 + \cos^2\vartheta\cos^2\varphi(v-w)^2}}}$$

It can be easily seen that $s_{occ}(\vartheta, \varphi)$ has 3 maxima where it equals ∞: $V_{0,*}$, $V_{90^o,0}$, $V_{0,90^o}$. These 3 views are along the 3 main axes of the ellipsoids. Fig. 5 shows the stability of the different views of such an ellipsoid where $w = 3.5,\ v = 2.1,\ u = 1$.

4.4 Gray level images:

We assume a fixed lighting source and the knowledge of the reflectance map of the object. The distance between similarity normalized gray level pictures is taken to be the sum of squared differences of gray levels. This was computed numerically for a Lambertian ellipsoid. The characteristic (most stable) view of the ellipsoid is shown in Fig. 6.

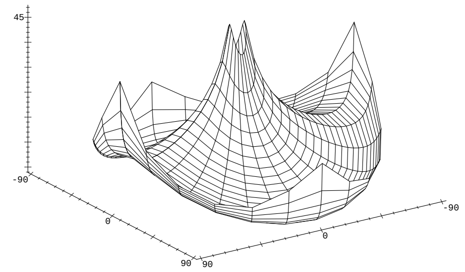

Figure 5: The stability of an ellipsoid, obtained for a matching algorithm that measures the distance between points of curvature extrema on the occluding contour of the ellipsoid. We give a log plot of the stability value, plotted over half the viewing sphere (which is flattened and defines a circle over the axes). The stability is ∞ when the ellipsoid is viewed down one of its axes, corresponding to points $[0, 0]$, $[0, \pm 90]$, and $[\pm 90, 0]$.)

Figure 6: The most likely view of a gray level image of an ellipsoid with axes $1, 2, 5$.

5 Applications

We use view likelihood to enhance existing image recognition (Section 5.1) and $3D$ reconstruction (Section 5.2) approaches. The enhanced paradigm is the following:

1. A geometrical object recognition technique (or a $3D$ reconstruction technique) is first used to compute a list of candidate objects. Each object in the list matches the given image up to a small discrepancy, which is smaller than a fixed error tolerance.
2. The view likelihood is computed for each object in the list, to select the best interpretation - the object whose view is most likely - from the list.

5.1 Maximum likelihood recognition

We start with the example discussed in the introduction (see Fig. 1). We are given a (large) database, which includes the 3 polyhedral objects shown at the bottom row of Fig. 1. Each object in the database is represented by the coordinates of a set of feature points (8 vertices in the case of the 3 polyhedral objects). To recognize the 3 images shown at the top row of Fig. 1, we proceed as follows:

Step 1: geometrical object recognition: It can be easily shown that an affine-based recognition algorithm, such as geometric hashing [7] or linear combination [8], will return the set of all 3 objects as possible matches for each image. It is a bit more complicated to show that with a small error tolerance, a similarity-based algorithm, such as alignment [5], will return the following lists:

- left image $\Longrightarrow$ {object (a), object (b)}
- middle image $\Longrightarrow$ {object (b), object (c)}
- right image $\Longrightarrow$ {object (c), object (a)}

The first object in each list matches the image exactly up to rotation and scale, and corresponds to the matching shown with thick white arrows in Fig. 1. Thus the white arrows show the output of a generic geometrical object recognition technique. Note, however, that the second object in each list predicts the image quite well (the difference is illustrated in Fig. 7).

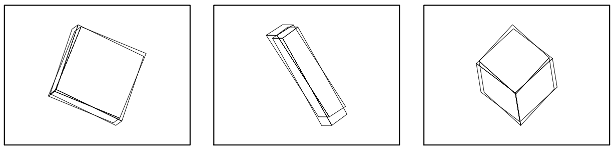

Figure 7: When the images in the top row of Fig. 1 are matched to the most likely object (illustrated by thin black arrows in Fig. 1), the matching is not precise. The difference between the closest image of the recognized object to the given image are shown.

Step 2: maximum likelihood recognition: $l(\vartheta, \varphi)$ is computed for each pair [model, image] obtained as a possible interpretation in step 1. Table 1 contains the *view likelihood* values, assuming a uniform prior over viewpoints.

	left image	middle image	right image
object (a)	0.17, 0.17	0.2,—	0.24, 0.23
object (b)	0.44, 0.4	0.008, 0.035	0.008,—
object (c)	0.21,—	0.3, 0.33	0.005, 0.002

Table 1: The *view likelihood* of each [model,image] pair, defined respectively in (7),(6) before normalization. Only the feasible matches, from the list produced by an affine-based and similarity-based recognition algorithms, are given.

Clearly the most-likely and most-stable interpretation for each image, regardless of whether the recognition system uses similarity or affine matching, is:

left image: object (b) (the flat box, twice as likely as the other objects)

middle image: object (c) (the elongated box, 10 times as likely as the other object suggested by the alignment method)

right image: object (a) (the cube, 100 times more likely than any other object)

The most likely interpretation is shown with thin black arrows in Fig. 1.

5.2 Maximum likelihood reconstruction

When there exist cues for $3D$ structure, but many interpretations are feasible, the likelihood functions are used to select the best among the interpretations under consideration. We call this process maximum likelihood reconstruction.

5.2.1 Fiducial points

Consider the battery charger picture shown in Fig. 8. From the text on the object we know that the object is a battery charger, and we presumably know that battery chargers are box-like in shape. Our reconstruction task is to compute the dimensions (up to scaling) of the charger from the image coordinates of the 7 visible vertices of the charger's enveloping box. We denote the dimensions of the charger by $u \times v \times 28$, where u is the depth of the charger (the width of the left face), v its height, and 28 the length of the front face (which scales the remaining measurements).

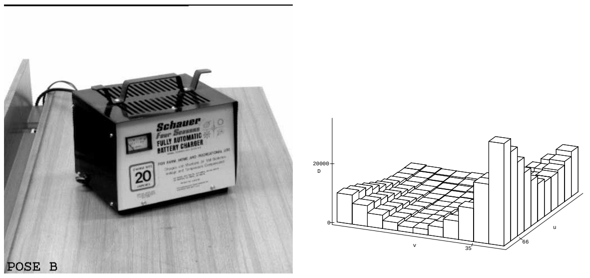

Figure 8: left: A battery charger. Right: the model-to-image distances, as a function of the parameters u, v described in the text. The coordinates are drawn in log scale in the ranges $u \in [14, 86]$, $v \in [14.5 - 41]$.

To find the best interpretation, we search the parameter space (u, v) in 2 stages:

1. We first compute the model-to-image distance[1] between each model and the picture. This gives us the function shown in Fig. 8-right. A picture of the correct model, which obtains a small image error, is shown in Fig. 9-left.

Figure 9: Left: an image of the correct object (which obtains a small model-to-image distance), super-imposed on the image of the original vertices of the charger. Right: the image of the most likely object, super-imposed on the original vertices. Clearly, both interpretations match the data reasonably well.

[1] A model-to-image metric measures the distance between the closest view of a $3D$ model to the given $2D$ image, up to some $2D$ image transformation [9].

2. We estimate an upper bound on the noise in the image to be 4 times the $3D$-affine distance between the model and the image[2]. Among all the models, whose model-to-image distance is smaller than this noise threshold, we compute the likelihood measure $l(\vartheta, \varphi)$ and choose the most likely interpretation.

Fig. 10 shows the likelihood of all the interpretations for which the model-to-image distance was smaller than the noise threshold; this function was computed as described in the object recognition example above. It follows that the most likely interpretation of the picture is the box $19.5 \times 21 \times 28$ (see Fig. 9-right), whereas the actual dimensions of the box are $22.6 \times 19.1 \times 28$ cm. Thus the picture is interpreted as a bit flatter and shorter than it really is. (We see the same kind of bias in the assessment a human would give to the dimensions of the battery charger, based on the single given image.)

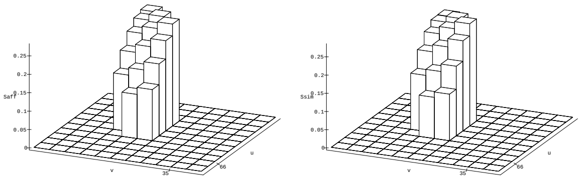

Figure 10: The likelihood of the battery charger shown in Fig. 8, as a function of the parameters u, v. The coordinates are drawn in log scale, in the ranges $u \in [14, 86]$, $v \in [14.5 - 41]$. The likelihood is set to 0 for parameter values for which the model-to-image distance is larger than the noise threshold. The left image corresponds to (7), and the right corresponds to (6).

5.2.2 Gray levels of smooth objects

Consider the ellipsoids in Fig. 11: the lengths of the first two axes of the ellipsoid on the left are immediately measurable from the picture, but the height (depth) can only be derived from the gray levels. If we want to select the best interpretation for the height of the ellipsoid given that the picture might be noisy, we pick the most likely height such that the distance between rendered pictures is small (less than 5 gray level values per pixel on the average). This results in choosing a flatter ellipsoid with parameters $1, 2, 3.2$ instead of $1, 2, 5$. This flattening affect is consistent with psychological evidence in humans.

6 Discussion

Two novel contributions are described in this work:

1. A workable definition of *view likelihood* and *view stability*, which can be readily used with any object recognition and reconstruction method. These two measures characterize the quality of a $3D$ interpretation to a given $2D$ image, in terms of robustness to accidental matches between the image and the "wrong" $3D$ interpretation.

2. Simple expressions of *view likelihood* and *view stability*, which were worked out for objects composed of localized $3D$ features, smooth ellipsoids, and angles.

We demonstrated an enhancement of existing object recognition techniques with maximum-likelihood estimation using real and simulated objects. We also showed how to compute the most likely (or stable) $3D$ interpretation of an ambiguous image. These applications are made possible with the use of the concepts of *view likelihood* and *view stability* developed here.

[2] $3D$-affine distance between a $3D$ model and a $2D$ image is the difference between the closest view of the model, transformed by any linear transformation, to the image. This distance should be 0 (ignoring noise) in our example for every model under consideration, since the picture is an image of a box, which is related by a linear transformation to any other box.

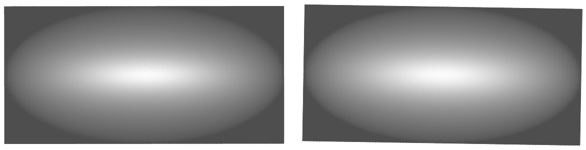

Figure 11: Left: the rendering of an ellipsoid with axis of length $1, 2, 5$. Right: the rendering of the most likely reconstruction of the left picture, an ellipsoid with axis of length $1, 2, 3.2$.

References

[1] J. Ben-Arie. *T-PAMI*, 12(8):760–774, 1990.

[2] J.B. Burns, R. Weiss, and E. Riseman. *T-PAMI*, 15(1):51–68, 1993.

[3] S. J. Dickinson, A. P. Pentland, and A. Rosenfeld. *T-PAMI*, 14(2):174–198, 1992.

[4] W. T. Freeman. In *Proc. of the 4th ICCV*, pages 347–356, Berlin, Germany, 1993. IEEE, Washington, DC.

[5] D. P. Huttenlocher and S. Ullman. *IJCV*, 5:195–212, 1990.

[6] K. Kanatani. *Group Theoretical Methods in Image Understanding*. Springer, Berlin, 1990.

[7] Y. Lamdan and H. Wolfson. In *Proc. of the 2nd ICCV*, pages 238–251, Tarpon Springs, FL, 1988. IEEE, Washington, DC.

[8] S. Ullman and R. Basri. *T-PAMI*, 13(10):992–1006, 1991.

[9] D. Weinshall and R. Basri. In *Proc. IEEE Conf. on CVPR*, pages 220-225, New-York City, NY, 1993. IEEE, Washington, DC.

[10] D. Weinshall and M. Werman. In *Proceedings of the IEEE International Symposium on Computer Vision*, pages 425–430, Coral Gables, USA, 1995.

[11] M. Werman and D. Weinshall. *T-PAMI*, 17(8):810–814, 1995.