On the Distortion of Shape Recovery from Motion

Tao Xiang $^a$,* and Loong-Fah Cheong $^b$

$^a$Department of Computer Science
Queen Mary, University of London, London E1 4NS, UK
txiang@dcs.qmul.ac.uk

$^b$Department of Electrical and Computer Engineering
National University of Singapore, Singapore 119260
eleclf@nus.edu.sg

Abstract

Given that most current Structure From Motion (SFM) algorithms cannot recover true motion estimates, it is important to understand the impact such motion errors have on the shape reconstruction. In this paper, various robustness issues surrounding different types of second order shape estimates recovered from motion cue are addressed. We present a theoretical model to understand the impact that errors in the motion estimates have on shape recovery. Using this model, we focus on the recovery of second order shape under different generic motions, each of which presenting different degrees of error sensitivity. Understanding such different distortion behavior is important if we want to design better fusion strategy with other shape cues.

Keywords: Shape recovery; Structure from motion; Error Analysis; Iso-distortion framework

1 Introduction

In spite of the best efforts of a generation of computer vision researchers, we still do not have a practical and robust system for reconstructing shape from a sequence of moving imagery. Questions have been asked
about the goals of shape reconstruction as well as the nature of shape representation appropriate for achieving these goals. Much of the intervening research emphasized either the importance of a closer coupling with tasks to be performed by the systems [2] and/or, often correlatively, a more qualitative representation of depth [9, 12]. The view is that the physical world imposes complexity such that adequate interpretation is impossible outside the context of tasks. Accordingly, less metric forms of depth representation such as projective depth and ordinal depth might be adequate for accomplishing these tasks.

It is also obvious that robustness of shape recovery depends on the cues (such as shading, contour, range data, as well as motion cue) used for depth reconstruction. Shape perception typically occurs in a context where a rich nexus of cues are available. Algorithms that estimated shape from certain cue may perform well only in restricted domains [19, 5, 27]. For instance, in the context of shading, [14] argued that reduction of stimulus has gone too far; subjects feel uneasy in judging relief of shaded images without contours. Thus, while it is well known that the recovery of second order shape is sensitive to errors, it does not mean that the extracted information is irrelevant. If the confidence level of the shape recovery from various cues can be ascertained against various motion-scene configurations, it then becomes possible to fuse the results of algorithms using different cues, and might even be the best strategy. In particular, if one type of cue is used as the main input, we would like to know under what configuration this type of cue would not be able to provide the useful information for the particular task and would seek help from other cues. Few works have been devoted towards this direction. This paper is a step towards this direction, especially with regards to the recovery of second order curved surfaces from the motion cue.

It is well known that current Structure From Motion (SFM) algorithms that recover structure parameters, especially those of second order quantities such as curvatures, are sensitive to various types of errors such as noise [1, 26]. However, not many researchers have addressed explicitly the problem of reconstruction accuracy given some errors in the 3-D motion estimates. Given that most current structure from motion algorithms cannot recover true motion estimates, our investigation seeks to clarify the impact such motion errors have on the shape estimates; in particular, we attempt to understand the severity of the resultant shape distortion under different motion-scene configurations. Are there generic motion types that can render depth recovery more robust and reliable? What are the likely problem conditions of motion cue? If such understanding could be achieved, we could use other shape cues to supplement motion cue when the latter
is inadequate so as to achieve robust shape perception.

In this paper, iso-distortion framework [4] is employed to analyze the errors in second order shape recovery, with regards to local representations such as normal curvatures and shape index [13]. As was shown in [5, 27], there exists certain dichotomy between forward (perpendicular to the image plane) and lateral (in the image plane) motion in terms of both 3-D motion and depth recovery. For instance, useful information like depth order is preserved under a lateral motion in a small field of view, even though it is difficult to disambiguate translation and rotation in such case; while under forward motion, a robust recovery of 3-D structure is much more difficult. In this paper, we make explicit the local shape distortion properties under lateral motion and forward motion.

The main contribution of this paper lies in the elucidation of the impact of errors in 3-D motion estimates on the second order shape recovery in a systematic manner. In particular, our findings show that the second order shape is recovered with varying degrees of uncertainty depending on the types of 3-D motion executed. Furthermore, we also make clear that different shapes exhibit different sensitivities in their recovery to errors in 3-D motion estimates. Evidently, these results remind us the importance of understanding the behavior of various SFM algorithms under different motion-scene configurations, which in turn might permit a better fusion strategy with other shape cues. Experiments are conducted to verify various theoretical results obtained in this paper. Our findings are also employed to explain some well-known psychophysical experiments results concerning human perception of second order shapes.

2 Background

2.1 Local shape representation

To begin, we look at a smooth local surface patch represented by its Taylor series up to the second order terms:

$$Z = \frac{1}{2} (k_{\text{min}} X^2 + k_{\text{max}} Y^2) + d + O^3(X, Y)$$  \hspace{1cm} (1)

where $k_{\text{min}}, k_{\text{max}}$ are the two principal curvatures with $k_{\text{min}} < k_{\text{max}}$. This represents a canonical case whereby the observer is fixating straight ahead at the surface patch located at a distance $d$ unit away, and the
The $X$-axis and the $Y$-axis are aligned with principal directions (directions of principal curvatures). Rotating the aforementioned surface patch around the $Z$-axis by $\theta$ in the anti-clockwise direction, we obtain a more general surface patch whose principal directions do not coincide with the $X-Y$ frame:

$$Z = \frac{1}{2} \left( \cos^2 \theta (k_{\min} X^2 + k_{\max} Y^2) + \sin^2 \theta (k_{\max} X^2 + k_{\min} Y^2) \right) + \cos \theta \sin \theta (k_{\min} - k_{\max}) XY + d + O^3(X, Y)$$

In this paper, we focus on the local shape measurements at the fixation point ($X=Y=0$), and we believe that these measurements influence significantly the perception of global shape. In the forthcoming analysis, since we are mainly interested in the second order shape quantities, we neglect the $O^3(X, Y)$ term in Equations (1) and (2).

Local representations of curved surfaces include the classical differential invariants of Gaussian and mean curvatures which are computed based on the principal curvatures. A good shape descriptor should correspond to our intuitive idea of shape: shape is invariant under translation and rotation, and more importantly, independent of the scaling operation. Principal curvatures, as well as the Gaussian and mean curvatures satisfy the former but do not satisfy the latter condition because they still contain the information of the amount of curvature. Koenderink [13] proposed two measures of local shape: shape index ($S$) and curvedness ($C$) as alternatives to the classical differential shape invariants. $S$ and $C$ are defined respectively as follows:

$$S = \frac{2}{\pi} \arctan \frac{k_{\min} + k_{\max}}{k_{\min} - k_{\max}}$$

$$C = \sqrt{\frac{k_{\min}^2 + k_{\max}^2}{2}}$$

$S$ is a number in the range of $[-1, +1]$ and obviously scale invariant and $C$ is a positive number with the unit $m^{-1}$. Shape index and curvedness provide us with a description of 3-D quadratic surfaces in terms of their types of shape and amount of curvature. Figure 1 shows the examples of quadratic surfaces with correspondent shape index values.

When shape is reconstructed based on motion cue, it is more appropriate to use shape index as the local measurement of shape type rather than other differential invariants. Due to the well-known ambiguity of SFM, the scale of the recovered objects and the speed of translation can only be determined up to a common
Figure 1: Examples of quadratic surfaces with correspondent shape index values. $S = \pm 1$: umbilic shape (spherical cap and cup); $S = \pm 0.5$: cylindrical shapes; $S \in [-1.0, -0.5]$ or $S \in [0.5, 1.0]$: elliptic shapes; $S \in [-0.5, 0.5]$: hyperbolic shapes.

The recovered scale information is thus meaningless unless additional information is available. Using shape index, we at least know that it can be estimated correctly when the SFM algorithm gives correct result.

2.2 General nature of distortion

This section derives the distortion in the recovered shape given that the motion parameters are imprecise. In the most general case, if the observer is moving rigidly with respect to its 3-D world with a translation $(U, V, W)$ and a rotation $(\alpha, \beta, \gamma)$, the resulting optical flow $(u, v)$ at an image location $(x, y)$ can be expressed as the well-known equation [15]:

\[
\begin{align*}
  u &= u_{\text{trans}} + u_{\text{rot}} \\
  &= (x - x_0) \frac{W}{Z} + \frac{\alpha xy}{f} - \beta \left( \frac{x^2}{f} + f \right) + \gamma y \\
  v &= v_{\text{trans}} + v_{\text{rot}} \\
  &= (y - y_0) \frac{W}{Z} + \alpha \left( \frac{y^2}{f} + f \right) - \frac{\beta xy}{f} - \gamma x
\end{align*}
\]

(5)

where $(x_0, y_0) = (f \frac{U}{W}, f \frac{V}{W})$ is the focus of expansion (FOE), $Z$ is the depth of a scene point, $u_{\text{trans}}, v_{\text{trans}}$ are the horizontal and vertical components of the flow due to translation, and $u_{\text{rot}}, v_{\text{rot}}$ the horizontal and vertical components of the flow due to rotation, respectively.
Since the depth can only be derived up to a scale factor, we set $W = 1$. Then the scaled depth of a scene point recovered can be written as

$$Z = \frac{(x - x_0, y - y_0) \cdot n}{(u - u_{rot}, v - v_{rot}) \cdot n}$$

where $n$ is a unit vector which specifies a direction.

If there are some errors in the estimation of the intrinsic or extrinsic parameters, this will in turn cause errors in the estimation of the scaled depth, and thus a distorted version of the space will be computed. Denoting the estimated parameters with the hat symbol (\(\hat{\cdot}\)) and errors in the estimated parameters with the subscript $e$ (where the error of any estimate $p$ is defined as $p_e = p - \hat{p}$), the estimated depth $\hat{Z}$ can be readily shown to be related to the actual depth $Z$ as follows:

$$\hat{Z} = Z \left(\frac{(x - \hat{x}_0, y - \hat{y}_0) \cdot n}{(x - x_0, y - y_0) \cdot n + (u_{rot}, v_{rot}) \cdot nZ + (N_x, N_y) \cdot nZ}\right)$$

(6)

where $(N_x, N_y)$ is a noise term representing the error in the estimate of optical flow. Past work in understanding the behavior of SFM algorithms has focused on the statistical aspect of noise, without an adequate accounting of the geometric aspects of the problem. In the ensuing analysis we want to emphasize the geometric aspect of the problem and we will therefore ignore the noise term$^1$.

From (6) we can see that $\hat{Z}$ is obtained from $Z$ through multiplication by a factor given by the terms inside the bracket, which we denote by $D$ and call the iso-distortion factor. The expression for $D$ contains the term $n$ whose value depends on the scheme we use to recover depth. In the forthcoming analysis, we will take the “epipolar reconstruction” approach [4], i.e. $n$ is along the estimated epipolar direction with

$$n = \frac{(x - x_0, y - y_0)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}}$$

If we denote, with slight abuse of notation, the homogeneous co-ordinates of a point $P^3$ by $(X, Y, Z, W)$, and the estimated position $\hat{P^3}$ by $(\hat{X}, \hat{Y}, \hat{Z}, \hat{W})$, the distortion transformation $\phi: P^3 \rightarrow \hat{P^3}$ can be written down. Note that to obtain the estimated $\hat{X}$, we use the back-projection given by $\frac{\hat{X}}{\hat{Z}} = D\frac{X}{Z} = DX$; similarly, $\hat{Y} = DY$. In general, the transformation from physical to perceptual space belongs to the family of Cremona transformations [4]. The resulting transformation $\phi$ is more complicated than a projective transformation, and very little statements can be made about the nature of the distorted shape. Indeed, some readers

$^1$The experiments presented in both [5] and this paper shows that our theoretical analyses are valid even with the presence of noise.
may observe that at certain points, the distortion $D$ is not defined. These points are known as the fundamental elements of the Cremona transformation and indeed characterize properties of the transformation. Despite the complex nature of the distortion, the recovered depths may nevertheless possess good properties under certain types of generic motions. For instance, ordinal depth information is preserved under lateral motion, even though it is difficult to disambiguate the translation and the rotation in such case [5, 27].

Many biological organisms often judge structure by making lateral motion, although the lateral translations are usually coupled with certain amount of rotation. On the other hand, psychophysical evidence [25] showed that under pure forward translation, human subjects were unable to recover structure unless favorable conditions such as large field of view exist. Thus it seems that not all motions are equal in terms of robust depth recovery and that there exists certain dichotomy between forward and lateral motion. In the following, we investigate the distortion of second order shape under lateral and forward motion. Since these two types of motions are often purposefully executed for various tasks such as depth recovery, we assume that the agent executing such motions is at least aware of the generic type of motion being executed. That is,

- When lateral motion is executed, $\hat{W} = W = 0$.
- When forward motion is executed, $\hat{U} = U = \hat{V} = V = 0$

### 3 Distortion under lateral motion

We consider an agent performing a lateral translational motion $(U, V, 0)$, coupled with a rotation $(\alpha, \beta, \gamma)$. Under the aforementioned assumption of $\hat{W} = W = 0$, the estimated epipolar direction would be a fixed direction for all the feature points with $\mathbf{n} = \frac{(\hat{U}, \hat{V})}{\sqrt{\hat{U}^2 + \hat{V}^2}}$. If the rotation around the optical axis is always not executed (which is usually true for a general visual system) so that the agent sets $\hat{\gamma} = 0$, we can further assume $\hat{\gamma} = \gamma = 0$. From (6) the distortion factor can be expressed as:

$$D = \frac{\left(\hat{U}^2 + \hat{V}^2\right) Z}{\left(\hat{U}\hat{\hat{U}} + V\hat{V})Z + \hat{U}\beta_c (X^2 + Z^2) - \hat{V}\alpha_c (Y^2 + Z^2) + (\hat{V}\beta_c - \hat{U}\alpha_c) XY\right)}$$

(7)
from which the mapping from the point \((X, Y, Z)\) to \((\hat{X}, \hat{Y}, \hat{Z})\) can be established.

Consider a surface \(s(X, Y, Z(X, Y))\), parameterized\(^2\) by \(X\) and \(Y\), transformed under the mapping to the perceived surface \(\hat{s}(\hat{X}, \hat{Y}, \hat{Z}(X, Y))\). By using Equation (7), the perceived surface can be parameterized by \(X\) and \(Y\) as follows:

\[
\hat{s}(X, Y) = (DX, DY, DZ)
\]

where \(D\) is defined in (7) and for brevity, \(Z(X, Y)\) has been written as \(Z\). Given this parameterization, *Mathematica*\(^3\) is used to manipulate the symbols as well as to investigate and visualize the distortion of the local shape in the following section.

### 3.1 Curvatures

Using *Mathematica*, we can obtain the principal curvatures and principal directions of any point on the distorted surface patch [8]. The expressions of the distorted curvatures under the general case (i.e. any point on the surface patch and any epipolar direction) are very complex. Due to space constraint, we focus on the special case of \(n = (1, 0)\) and \((X = Y = 0)\), unless otherwise indicated. This case is most relevant to a situated agent, which often executes horizontal lateral motion to judge distance. In such case, it is reasonable to assume that \(\hat{V} = 0\) although the true \(V\) is not necessarily zero due to perturbances. In other words, the estimated epipolar direction can be regarded as along the \(X\)-axis direction. Furthermore, it is reasonable to believe that the local shape distortion at the fixation point (i.e. \((X = Y = 0)\)) would influence strongly the viewer’s perception of the surface in the surrounding region; this phenomenon can be observed in the simulations and experiments presented later for the case of lateral motion. For the local surface patch given by Equation (2), we obtain the perceived principal curvatures \(k_{\min}^\wedge\) and \(k_{\max}^\wedge\) as follows:

\[
\begin{align*}
k_{\min}^\wedge &= \frac{1}{2} \left( \frac{(k_{\min} + k_{\max})U - 2\beta_e}{U} - \sqrt{\frac{Q}{U^2}} \right) \\
k_{\max}^\wedge &= \frac{1}{2} \left( \frac{(k_{\min} + k_{\max})U - 2\beta_e}{U} + \sqrt{\frac{Q}{U^2}} \right)
\end{align*}
\]

where \(Q = U^2(k_{\min} - k_{\max})^2 + 4(\alpha_e^2 + \beta_e^2) - 4U(k_{\min} - k_{\max})(\beta_e \cos 2\theta + \alpha_e \sin 2\theta)\). Notice that the parameter \(d\) which indicates the distance between the surface patch and the observer does not affect the

\(^2\)Such parametric representation may not be possible for more complex global shapes.

\(^3\)Mathematica is a registered trademark of Wolfram Research, Inc.
From (9), it is obvious that Gaussian and mean curvatures are not preserved under the distortion; even the signs of principal curvatures are not necessarily preserved. The principal directions of the perceived surface patch can be obtained by computing the eigenvectors of the Weingarten matrix [8] of the parametrical shape representation. It was found that generally the principal directions are not preserved. Only when the principal directions are aligned with the $X -$ and the $Y -$axis and $\alpha_e = 0$, they are preserved, though the directions of the maximum curvature and that of the minimum curvature may swap.

We are also interested in the distortion in normal curvatures, especially the normal curvatures along the horizontal and vertical directions which are denoted as $Z_{XX}$ and $Z_{YY}$ respectively and given by:

$$Z_{XX} = k_{\min} \cos^2 \theta + k_{\max} \sin^2 \theta$$
$$Z_{YY} = k_{\min} \sin^2 \theta + k_{\max} \cos^2 \theta$$ (10)

The distorted normal curvatures along the horizontal and vertical directions can be obtained as:

$$\hat{Z}_{\hat{X}\hat{X}} = \frac{U}{\hat{U}} Z_{XX} - \frac{2\beta_e}{\hat{U}}$$
$$\hat{Z}_{\hat{Y}\hat{Y}} = \frac{U}{\hat{U}} Z_{YY}$$ (11)

Interestingly, $\hat{Z}_{\hat{X}\hat{X}}$ and $\hat{Z}_{\hat{Y}\hat{Y}}$ are not affected by $\alpha_e$. Equation (11) shows that normal curvatures along different directions have different distortion properties.

### 3.2 Shape index

From the expression of the distorted principal curvatures, the distorted shape index at the fixation point can readily be obtained by substituting (9) into (3):

$$\hat{S} = \frac{2}{\pi} \arctan \left( \frac{(k_{\min} + k_{\max})U - 2\beta_e}{\hat{U} \sqrt{\frac{2}{\hat{U}}}} \right)$$ (12)

As the expression for $\hat{S}$ is very complex, we have assorted some diagrams to graphically illustrate the error in the shape index estimate with respect to the true shape index. The curvedness is fixed for each diagram so as to factor out the influence of curvedness. Figure 2 shows the distortion obtained by varying
Figure 2: Distortion of shape index with different principal directions, plotted in terms of $S_e$ against $S$. $	heta = 0$ for (a), $\theta = \frac{\pi}{4}$ for (b) and $\theta = \frac{\pi}{2}$ for (c). (d) was obtained by averaging curves with $\theta = 0 + n \times \frac{\pi}{8}$ where $n = 0, 1, \cdots 8$. All the other parameters are identical: $\alpha_e = 0.005$, $\beta_e = -0.02$, $U = 0.2$ and $\hat{U} = 0.18$.

principal directions while fixing all the other parameters. Figures 2(a), (b) and (c) show the sensitivity of shape recovery to the motion estimation errors for three principal directions: $\theta = 0$, $\theta = \frac{\pi}{4}$ and $\theta = \frac{\pi}{2}$. Figure 2(d) show the average sensitivity for different principal directions. Note that the sudden jump of $S_e$ in some of the curves corresponds to the case where the direction of the maximum curvature and that of the minimum curvature have swapped. It can also be seen that the principal direction does affect the robustness of the shape perception. We can also infer from Figure 2 that different shapes have different distortion properties. The average curve shown in Figure 2(d) gives us a rough idea on the overall robustness of the perception of different shapes: the saddle-like shapes (also known as “hyperbolic shapes” with $k_{\min}$ and $k_{\max}$ having different signs) are more sensitive to the errors in 3-D motion estimation than the concave and convex shapes (“elliptic shapes” or “umbilic shapes” with $k_{\min}$ and $k_{\max}$ having the same sign).
Figure 3: Other factors that affect the distortion of shape index (see text). $\beta_e = -0.03$ for (a), $\beta_e = 0.02$ for (b) and $\beta_e = -0.02$ for (c) and (d); $\alpha_e = -0.05$ for (c) and $\alpha_e = 0.005$ for (a), (b) and (d); $C = 20$ for (d) and $C = 10$ for (a), (b) and (c). $\theta = 0$, $U = 0.20$ and $\tilde{U} = 0.18$ for all the diagrams.

Other factors that may have an impact on the sensitivity of shape perception are studied in Figure 3. $\beta_e$ has a significant influence on the shape perception. Comparing Figure 3(a) with Figure 2(a), we can see that the error in the estimated shape index increases with increasing value of $|\beta_e|$. The sign of $\beta_e$ will determine whether the shape index will be under-estimated or over-estimated for each shape type (comparing Figures 3(a) and 3(b)). On the contrary, $\alpha_e$ seems to have little influence on the recovered shape index, as shown in Figure 3(c). This anisotropy between $\alpha_e$ and $\beta_e$ is related to the directional anisotropy that exists in the distortion of the normal curvatures, derived earlier in Equation (11). This anisotropy is of course a result of the estimated translation being in the horizontal direction. Finally, our analysis seems to support the view that the more curved surfaces will be perceived with high accuracy (see Figure 3(d)).
4 Distortion under forward motion

We assume pure forward translation is performed, and the agent executing such motion is aware that such generic type of motion is being executed. We thus have \( \dot{U} = U = \dot{V} = V = 0 \). We also assume \( \dot{\gamma} = \gamma = 0 \) as we did before. Adopting “epipolar reconstruction” approach, the distortion factor can be expressed as:

\[
D = \frac{X^2 + Y^2}{X^2 + Y^2 + \alpha_c Z (X^2 Y + Y Z^2 + Y^3) - \beta_c Z (X Y^2 + X Z^2 + X^3)}
\]  

(13)

Following the same procedure of deriving the distorted curvatures as in the lateral motion case, we can obtain the distorted principal and normal curvature expressions on any point of the distorted surface. The distorted shape index can also be computed thereafter. These expressions are very complex and are thus not presented here. In particular, at the fixation point, the distorted local shape measurements are undefined due to the undefined distortion factor (see Equation (13)). Our analysis in [5] shows that the distortion factor varies wildly around the fixation point, which implies that the distorted surface patches are also not smooth. Therefore, the distorted local shape measurements at any particular point do not make much sense.

To have an idea on how curved surface patches will be distorted under different generic motions, we use Mathematica to obtain the shapes and graphically display them. Figure 4 compares the distortion in the original and recovered surfaces under lateral and forward motions. We consider a vertical cylinder (upper row) and a horizontal cylinder (lower row), given by the equations \( Z = \frac{1}{2}X^2 + 4 \) and \( Z = \frac{1}{2}Y^2 + 4 \) respectively. For the lateral motion case and given the parameters stated in the caption of Figure 4, it is easy to show that from Equations (9) and (11) that at the fixation point a vertical cylinder will be perceived as a less curved vertical cylinder and a horizontal cylinder will be perceived as a saddle-like shape. The principal directions remain unchanged. Figures 4(b) and (e) show the full reconstructed surfaces over the entire cylinders; clearly the distorted surface patches are smooth and the distortion behavior at the fixation point seems to be qualitatively representative of the global shape distortion even with a large view angle. The recovered surface patches under forward motion, by contrary, are not smooth and show large distortion. Even when the errors in the rotation estimates are much smaller than those in the case of lateral motion, the perceived surface patches are obviously not quadrics with major distortion occurring around the central region of attention.

Not surprisingly, the psychophysical experiments conducted under the forward relative motion [20, 25]
Figure 4: Distortion of cylinders under lateral and forward motion. (a) and (d) are the original vertical and horizontal cylinders respectively. (b) and (e) are the distorted vertical and horizontal cylinders under lateral motion respectively, with $U = 0.9$, $\hat{U} = 1.0$, $V = 0.5$ and $\hat{V} = 0$. (c) and (f) are the distorted vertical and horizontal cylinders under forward motion respectively. The errors in the estimated rotation are: $\alpha_e = 0$ and $\beta_e = 0.05$ for (b) and (e), and $\alpha_e = 0$ and $\beta_e = 0.001$ for (c) and (f). The field of view is $127^\circ$ for all the diagrams.
showed that the perceived shape experienced significant distortion when motion was the only cue. On the other hand, human seems to recover shape relatively successfully when lateral motion is executed, although distortion was also reported [3, 6, 7].

5 Human perception of shape under lateral motion

With the preceding as a theoretical backdrop, we will now discuss two important results of psychophysical experiments performed on curved surface perception, namely that of [3] and [6, 7]. By explaining these results with our preceding theoretical model, we demonstrate the usefulness of our theoretical predictions and hope to further illustrate the properties of these distortions.

5.1 Directional anisotropy in curvature perception

The directional anisotropy in the detection and discrimination of the surface curvature from motion cue has been reported in many psychophysics papers. Rogers and Graham [21] showed that depth percept occurred for surface defined by either motion parallax or stereopsis. They found that the percept was obtained only when the depth varied in the horizontal direction, and did not occur when the depth variation was along the vertical direction. Cornilleau-Pérez [3] found a similar anisotropy for cylindrical surfaces defined by motion. In the psychophysical experiments conducted, the motion executed was that of the cylinders rotating about a vertical axis passing through the object center (equivalent to a horizontal translation $U$ and a rotation $\beta$ in the observer’s reference frame). The subjects were asked to judge on the normal curvatures in the horizontal and the vertical directions in a small field of view. It was found that the curvatures were much less accurately detected when the cylinders were curved in the direction of rotation than when they were curved in a perpendicular direction. Norman and Lappin [17] confirmed Cornilleau-Pérez’s findings and extended to the discrimination between different quadric surfaces. Basically it was found that the direction of frontal-parallel translation and the curvature orientation will influence the perception of the surface curvature.

From the standpoint of the iso-distortion framework, the shapes perceived by human subjects are likely to be a distorted version of the true shape since it is in general very difficult to recover the true motion. Given the simple experimental configuration in [3], we expect that the human subjects will be aware of the
type of generic motion being executed, i.e. $\hat{V} = V = 0$. Furthermore it is reasonable to expect that human subject employed a one-dimensional model, i.e. $\alpha_e = 0$ and the only ambiguity is between $U$ and $\beta$. In this case, at the fixation point $X = Y = 0$, we have the principal directions unchanged and thus the principal curvatures would be just given by the normal curvatures along the $X$ and $Y$ direction ($\hat{Z}_{X,X}$ and $\hat{Z}_{Y,Y}$ in equation (11)).

It is clear that given such horizontal movements and the attendant distortion, the perception of the vertical curvature can be performed with a greater degree of confidence since $\hat{Z}_{Y,Y}$ is directly proportional to $Z_{Y,Y}$. In particular, if $Z_{Y,Y} = 0$, there would be no bending in the vertical direction (i.e. $\hat{Z}_{Y,Y} = 0$). On the other hand, and this is where directional anisotropy arises, the equation for $\hat{Z}_{X,X}$ contains a proportional term and a constant term (clearly such anisotropy arises as a consequence of the estimated translational direction, which in turn is influenced by the true translational direction). It is reasonable to expect that $U$ and $\hat{U}$ have the same sign and thus the first term alone will not change the sign of the curvature. The second term represents an offset term; it adds a fixed amount of convexity or concavity to the actual curvature, depending on the sign of $\beta_e$ relative to $\hat{U}$. The effects of this second term are two-fold: 1) a plane will be perceived as a vertical cylinder ($\hat{Z}_{X,X} \neq 0$), similar to the phenomenon of Apparent Fronto Parallel Plane (AFPP) reported in both stereo and motion cues experiments [10, 18, 24]; 2) it is harder to judge on concavity/convexity especially for the cases where the actual curvatures are small. The degree of difficulty experienced in concavity/convexity judgments depends on the relative magnitude of the term $\beta_e$ and the estimated translation $\hat{U}$, and thus indirectly on the actual translation $U$.

## 5.2 Perception of shape index

Based on the work of [13], Damme et al. [6, 7] conducted experiments on the perception of shape index and curvedness. Shape perception under active vision (observer are moving during perception) was studied using dynamic random dot patterns. It is believed that 3-D structure from active vision is different from 3-D structure from moving object, in the sense that under active vision more cues are available, thus the perception should be more accurate. The random dots simulated continuous quadratic surfaces which are defined by equation (2). The principal directions were random during each trial. Just Noticeable Differences (JND) between shape index and between curvedness were utilized as an indicator of the sensitivity or dis-
criminability of shape perception. Damme et al. [7] reported that given the same curvedness, some types of shapes can be perceived quite accurately, whereas others are more difficult to be distinguished. In particular, the recognition of parabolic shapes \((S = \pm 0.5)\) would be relatively robust, whereas that for umbilic and hyperbolic shapes \((S = \pm 1.0\) and \(S = 0\) respectively) would be difficult. The overall result exhibits a W-shaped sensitivity curve as shown in Figure 5.

Figure 5: From [7]: Just noticeable shape index differences for the three subjects FV, WD and IH for the three conditions \(C = 10m^{-1}\), \(C = 20m^{-1}\) and \(C\) = random.

From the standpoint of the iso-distortion framework, given some errors in the motion estimation, the shape index of the surface patch would be perceived with distortion. In Section 3.2, we have derived the expression for \(\hat{S}\). The error behavior given various parameters has also been investigated. That there is different error sensitivities for different shapes is not unexpected based on our illustration of the relationship between \(S_e\) and \(S\) for fixed curvedness (Figure 2). For \(\theta = 0\), Figure 2(a) showed that the absolute error \(|S_e|\) is indeed a W-shaped function (note that \(|S_e|\) and the just noticeable shape index difference are not exactly the same concept ). However, with different principal directions, our derivation yielded somewhat different error sensitivity functions, as illustrated by Figures 2(b) and 2(c). Depending on the principal directions, the
sensitivity curve is not necessarily W-shaped. Figure 2(d) showed that on average, the perception of cylinder is now not as robust as that of the sphere and the symmetrical saddle, and that the average sensitivity curve does have a somewhat W-shaped appearance.

5.3 Discussions

The results of the preceding psychophysical experiments corroborate the theoretical predictions regarding perceived shape distortion when only motion cue is present. That is, the optical flow can be a misleading depth cue, leading to a curved percept when the surface is flat, or to a wrong perception of the sign of the curvature and the types of shape. Similar effects were found for stereopsis, as the apparent frontal-parallel plane can correspond to a real surface which is concave or convex [10]. Therefore, in situations where the perception of the apparent shape of large surfaces is critical (e.g. during aircraft landing, performing navigation in narrow environments), it may be appropriate to consider the use of static perspective cues to complement the motion cues. Indeed the works by Stevens and Brooks [23], and Sparrow and Stine [22] have shown that for human, these static cues can indeed dominate stereopsis or motion cues during the perception of plane orientation.

6 Experiments

In this section, we conducted computational experiments on computer generated and real images to further verify the theoretical predictions.

6.1 Computer generated image sequences

SOFA 4 is a package of 9 computer generated image sequences designed for testing research works in motion analysis. It includes full ground truth on the motion and camera parameters. Sequence 1 and 5 (henceforth abbreviated as SOFA1 and SOFA5) were chosen for our experiments, the former depicting a lateral motion and the latter a forward motion. Both of them have an image dimension of 256 × 256 pixels, a focal length

4courtesy of the Computer Vision Group, Heriot-Watt University (http://www.cee.hw.ac.uk/~mtc/sofa).
of 309 pixels and a field of view of approximately 45°. The focal length and principal point of the camera were fixed for the whole sequence. Optical flow was obtained using Lucas and Kanade’s method [16], with a temporal window of 15 frames. Depth was recovered for frame 9. We assumed all the intrinsic parameters of the camera were estimated accurately throughout the experiments and concentrated on the impact of errors in the extrinsic parameters.

The 3-D scene for SOFA1 consisted of a cube resting on a cylinder (Figure 6(a)). Clearly the entire feature points on the top face of the cylinder in SOFA1 (which are delineated in Figure 6(a)) lie on a plane. The reconstructed shape of this plane was used to testify our theoretical predictions in the lateral motion case. The camera trajectory for SOFA1 was a circular route on a plane perpendicular to the world Y-axis, with constant translational parameters \((U, V, W) = (0.8137, 0.5812, 0)\) and constant rotational parameters \((\alpha, \beta, \gamma) = (-0.0203, 0.0284, 0)\). If the observer or the system was aware that a lateral motion is being executed, then \(\dot{W} = 0\), and the following equation can be used to recover depth:

\[
\hat{Z} = \frac{-\hat{f}(\hat{U}, \hat{V}) \cdot \mathbf{n}}{(u, v) \cdot \mathbf{n} - (u_{\text{ref}}, v_{\text{ref}}) \cdot \mathbf{n}}
\]  

(14)

The translation estimates were fixed along the horizontal direction, which means that \(\mathbf{n}\) would be fixed as \((1, 0)\). The resultant recovered depths, given different errors in the 3-D motion estimates, were illustrated in Figure 6 using 3-D plot viewed from the side. In particular, Figure 6(b) shows that, with motion parameters accurately estimated, the plane (top of the cylinder) remained as a plane, although the noise in the optical flow estimates made some points ‘run away’ from the plane. Figure 6(c) depicts the shape recovered when \(\beta_e > 0\) and \(\dot{U} < 0\). According to Equation (11), we have \(\hat{Z}_{\hat{X},\hat{X}} > 0\). It can be seen from Figure 6(c) that the plane was indeed reconstructed as a convex surface. Conversely, when \(\beta_e < 0\) and \(\dot{U} < 0\), concave surfaces were perceived (Figures 6(d), (e) and (f)). Comparing Figure 6(d) with Figure 6(e), we find that larger \(\dot{U}\) in general resulted in smaller curvature distortion, whereas the results of Figure 6(d) with Figure 6(f) show that large \(\beta_e\) resulted in larger curvature distortion. Figures 6(c)–(f) show that the reconstructed surfaces were not curved in the \(Y\) direction. All these results can be explained using Equation (11).

The SOFA5 sequence was used in the next set of experiments to verify predictions in the case of forward motion. The 3-D scene for SOFA5 comprised of a pile of 4 cylinders stacking upon each other and in front of a frontal-parallel background (Figure 7(a)). The camera trajectory for SOFA5 was parallel to the world \(Z\)-axis and the corresponding translational and rotational parameters were \((U, V, W) = (0, 0, 1)\) and
Figure 6: Computer generated lateral motion sequence and depth recovery. (a) SOFA1 frame 9 with the top face of the cylinder delineated; (b) Reconstruction with true motion parameters. (c) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.03, 0.01, 0)$. (d) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.01, 0.04, 0)$. (e) Reconstruction with $(\hat{U}, \hat{V}) = (-2, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.01, 0.04, 0)$. (f) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.01, 0.06, 0)$. 
(\alpha, \beta, \gamma) = (0, 0, 0) \) respectively. We assumed that the observer or the system was aware that a forward motion is being executed, i.e. \( \hat{U} = \hat{V} = 0 \). Performing epipolar reconstruction, the equation for calculating depth for each feature point would be:

\[
\hat{Z} = \frac{x^2 + y^2}{(u, v) \cdot (x, y) - (u_{rot}, v_{rot}) \cdot (x, y)}
\]

Figure 7(b) depicts the case of no errors in the motion parameters. It can be seen that the background plane was preserved roughly. However, when there was small amount of errors in the rotation estimates, a complicated curved surface with significant distortion was reconstructed (shown in Figure 7(c)). It is in accordance with our prediction that large depth distortion is expected when forward motion is performed.

![Image](image1.png)

(a) (b) (c)

Figure 7: Computer Generated forward motion sequence and depth recovery. (a) SOFA5 frame 9. (b) Reconstruction with true motion parameters. (c) Reconstruction with \((\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.001, -0.001, 0)\).

6.2 Real image sequence

Depth recovery was also performed on a real image sequence BASKET\(^5\). It was taken by a stationary video camera on an upturned basket being rotated on an office chair (Figure 8(a)). The true intrinsic and extrinsic parameters are not available. However, we know that the basket was rotating about a vertical axis which was approximately parallel to the \(Y\)-axis of the camera co-ordinate system and located along the \(Z\)-axis of the camera co-ordinate system. The equivalent egomotion can thus be expressed as \((U, V, W) = (\beta_0Z_0, 0, 0)\) and \((\alpha, \beta, \gamma) = (0, -\beta_0, 0)\) where \(\beta_0\) was the rotational velocity of the basket \((\beta_0 > 0\) in this case) and \(Z_0\)

\(^5\) courtesy of Dr Andrew Calway of Department of Computer Science, University of Bristol.
the distance between the optical center of the camera and the centroid of the basket. Depth was recovered for frame 9 using Equation (14). Notice that since the camera was not calibrated, we have arbitrarily set the value of the focal length as $\hat{f} = 300$. Fortunately, we have proven in [5] that under lateral motion the errors in the intrinsic parameters would not affect the depth recovery results qualitatively. These recovered shapes, given different errors in the 3-D motion estimates, were shown in Figure 8. Figure 8(b) depicts the shape recovered when $\hat{U} < 0$ and $\beta_e < 0$ ($\beta_e < 0$ since $\beta < 0$ and $\hat{\beta} > 0$). Substituting the signs of these terms into Equation (11), we obtained $\hat{Z}_{XX} < Z_{XX}$, which means that the recovered surface should be less convex than the true shape. Comparing the results of Figure 8(b) with the true shape measured by us manually, this is indeed the case. Furthermore, according to Equation (11), a smaller $\beta_e$ or a larger $\hat{U}$ (with the signs of these terms unchanged) would result in an even less convex shape being recovered. Figures 8(c) and (d) clearly support our prediction.

7 Conclusions and future work

This paper presented a geometric investigation on the reliability of shape recovery from motion cue given some errors in the estimates for motion parameters. Distortion in the local shape recovered was considered under different generic motions. Specifically, our distortion model has shown that the types of motions executed are critical for the accuracy of the shape recovered. In the case of lateral motion, the accuracy of curvature estimates exhibits anisotropy with respect to the estimated translation direction, in that the surface curvature in the estimated translational direction is better perceived than that in the orthogonal direction. Correlatively, the recognition of different shapes (classified in terms of shape index) exhibits different degrees of sensitivity to noise. In the case of forward motion, the reconstructed shape experiences larger distortion compared with the case of lateral motion and is not visually smooth any more.

These findings suggest that shape estimates are recovered with varying degrees of uncertainty depending on the motion-scene configuration. If the confidence level of the shape estimates under various motion-scene configurations can be ascertained, the robustness of shape perception can be enhanced by including additional static cues to restrict the space of possible interpretation. This study has taken a step towards this direction. More works need to be done to understand the robustness of shape recovery under more general
Figure 8: Real lateral motion sequence and shapes recovered. (a) BASKET frame 9. (b) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.05, 0.05, 0)$. (c) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.05, 0.1, 0)$. (d) Reconstruction with $(\hat{U}, \hat{V}) = (-0.5, 0)$ and $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.05, 0.05, 0)$. 

22
configurations. The investigation should also be extended to the other cues. Practical and robust fusion algorithms would not be possible until a full understanding of the robustness of the shape recovered based on various cues can be achieved.

References


