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 quadratic 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 quadratic surfaces 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 accurately reconstructing shape from a sequence of moving imagery. Shape reconstruction

* Corresponding author. Tel: (+44)-(0)20-7882-5230; Fax: (+44)-(0)20-8980-6533

typically occurs in a context where different cues such as shading, contour, range data and motion are available. It has been realised that the robustness of shape recovery depends on the cues used for depth reconstruction. In particular, algorithms that estimated shape from certain cue may perform well only in restricted domains [14, 4, 17]. Thus, while it is well known that the recovery of second order shape (i.e. 3-D quadratic surfaces) 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 to produce better results. More specifically, 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 3-D quadratic 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, 16]. 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 SFM 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 shape 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 [3] is employed to analyze the errors in second order shape recovery, with regards to local representations such as normal curvatures and shape index [10]. As was shown in [4, 17], 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.

2 Background

2.1 Local shape representation

To begin, we look at a smooth surface patch:

$$\begin{aligned} Z &= \frac{1}{2}\left(\cos^2\theta(k_{min}X^2 + k_{max}Y^2) + \sin^2\theta(k_{max}X^2 + k_{min}Y^2)\right) \\ &\quad + \cos\theta\sin\theta(k_{min} - k_{max})XY + d \end{aligned} \tag{1}$$

where k_{min}, k_{max} are the two principal curvatures with $k_{min} < k_{max}$. 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 two principal directions (directions of principal curvatures) can be obtained by rotating the $X-$axis and $Y-$axis around the $Z-$axis by θ in the anti-clockwise direction.

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 [10] 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:

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

$$C = \sqrt{\frac{k_{min}^2 + k_{max}^2}{2}} \tag{3}$$

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.

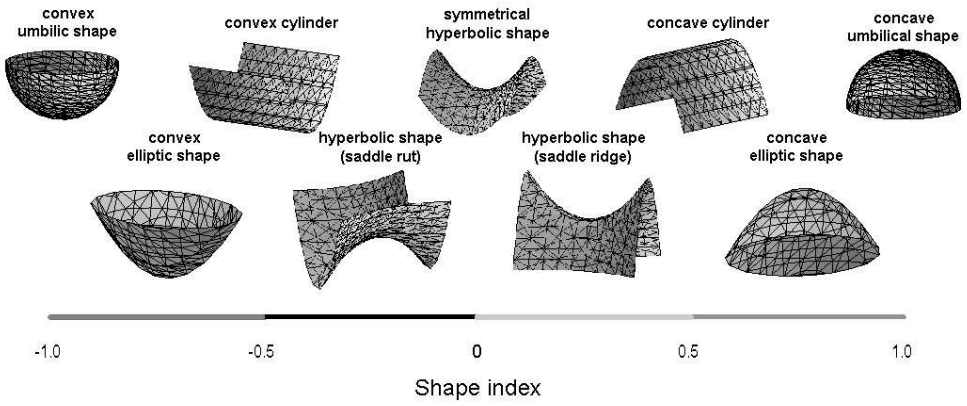


Figure 1: Examples of quadratic surfaces with correspondent shape index values. $S = \pm 1$: umbilic shape (spherical cap and cup); $S = \pm 0.5$: cylindrical shapes; $S \in [-1.0, -0.5]$ or $S \in [0.5, 1.0]$: elliptic shapes; $S \in [-0.5, 0.5]$: hyperbolic shapes.

When shape is reconstructed based on the 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 factor. 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 (α, β, γ), the resulting optical flow (u, v) at an image location (x, y) can be expressed as the well-known equation [11]:

$$\begin{aligned} u &= u_{trans} + u_{rot} \\ &= (x - x_0)\frac{W}{Z} + \frac{\alpha x y}{f} - \beta\left(\frac{x^2}{f} + f\right) + \gamma y \\ v &= v_{trans} + v_{rot} \\ &= (y - y_0)\frac{W}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \frac{\beta x y}{f} - \gamma x \end{aligned} \tag{4}$$

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 $W = 1$ for the case of general motion; the case of pure lateral motion ($W = 0$) will be discussed separately where required. Then the scaled depth of a scene point recovered can be written as:

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

where $\mathbf{n}$ 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. In this paper, we denote the estimated parameters with the hat symbol (ˆ) 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:

$$\hat{Z} = Z\left(\frac{(x - \hat{x}_0, y - \hat{y}_0) \cdot \mathbf{n}}{(x - x_0, y - y_0) \cdot \mathbf{n} + (u_{rot_e}, v_{rot_e}) \cdot \mathbf{n}Z}\right) \tag{5}$$

From (5) 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 $\mathbf{n}$ whose value depends on the scheme we use to recover depth. In the forthcoming analysis, we will take the "epipolar reconstruction" scheme [3], i.e. $\mathbf{n}$ is along the estimated epipolar direction with $\mathbf{n} = \frac{(x-\hat{x_0}, y-\hat{y_0})}{\sqrt{(x-\hat{x_0})^2+(y-\hat{y_0})^2}}$. The statistical and geometrical reasons for choosing this scheme of reconstructing shape were explained in [4].

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 $\left(\hat{X}, \hat{Y}, \hat{Z}, \hat{W}\right)$, 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 $\frac{x\hat{Z}}{f} = D\frac{xZ}{f} = DX$; similarly, $\hat{Y} = DY$. In general, the transformation from physical to perceptual space belongs to the family of Cremona transformations [3]. The resulting transformation ϕ is more complicated than a projective transformation, and very little statements can be made about the nature of the distorted shape. Indeed, some readers 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, depth order information is preserved under lateral motion, even though it is difficult to disambiguate the translation and the rotation in such case [4, 17].

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 [15] 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 motions. For an active vision system, these two types of motions are often purposefully executed for various tasks such as depth recovery, which means that the agent executing such motions is at least aware of the generic type of motion being executed. For a general vision system, the 3-D Motion needs to be estimated from captured image sequences. In [17], we have shown that current 3-D motion algorithms are able to accurately detect the generic type of motion being executed under lateral and forward motions. We thus have:

$$\begin{cases} \hat{W} = W = 0 & \text{when lateral motion is executed} \\ \\ \hat{U} = U = \hat{V} = V = 0 & \text{when forward motion is executed} \end{cases} \tag{6}$$

3 Distortion of shape recovery under generic motions

3.1 Lateral motion

We consider an agent performing a lateral translational motion $(U, V, 0)$, coupled with a rotation (α, β, γ). Under the aforementioned assumption of $\hat{W} = W = 0$, the estimated epipolar direction would be a fixed direction for all the feature points with $\mathbf{n} = \frac{(\hat{U}, \hat{V})}{\sqrt{\hat{U}^2 + \hat{V}^2}}$. In typical visual systems, rotation about the optical axis is usually not executed ($\gamma = 0$). It is also shown in [17] that rotation about the optical axis can be estimated accurately by current SFM algorithms. It is thus reasonable to assume $\hat{\gamma} = \gamma = 0$. The distortion factor can therefore be expressed as:

$$D = \frac{\left(\hat{U}^2 + \hat{V}^2\right) Z}{\left(\begin{array}{l} (U\hat{U} + V\hat{V})Z + \hat{U}\beta_e \left(X^2 + Z^2\right) - \\ \hat{V}\alpha_e \left(Y^2 + Z^2\right) + \left(\hat{V}\beta_e - \hat{U}\alpha_e\right) XY \end{array} \right)} \tag{7}$$

from which the mapping from a point (X, Y, Z) in the physical space to its reconstruction $\left(\hat{X}, \hat{Y}, \hat{Z}\right)$ can be established.

Consider that a surface $\mathbf{s}(X, Y) = (X, Y, Z(X, Y))$, parameterized[1] by X and Y, is transformed under the mapping to the reconstructed surface $\hat{\mathbf{s}}(\hat{X}, \hat{Y}) = (\hat{X}, \hat{Y}, \hat{Z}(\hat{X}, \hat{Y}))$, which is parameterized by $\hat{X}$ and $\hat{Y}$. Using Equation (7), the reconstructed surface can also be parameterized by X and Y as follows:

$$\hat{\mathbf{s}}(X, Y) = (DX, DY, DZ) \tag{8}$$

where D is defined in (7) and for brevity, $Z(X, Y)$ has been written as Z. Given this parameterization, *Mathematica*[2] is used to manipulate the symbols as well as to investigate and visualize the distortion of the local shape in the following sections.

[1] Such parametric representation may not be possible for more complex global shapes.

[2] *Mathematica* is a registered trademark of Wolfram Research, Inc.

3.1.1 Curvatures

Using *Mathematica*, we can obtain the principal curvatures and principal directions of any point on the distorted surface patch [5]. 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 only give the results under the special case of $\mathbf{n} = (1, 0)$ and $(X = Y = 0)$, unless otherwise indicated. It is found that many of the results obtained under this special case are still hold under a general case as can be observed in the simulations and experiments presented later. For the local surface patch given by Equation (1), we obtain the perceived principal curvatures $\hat{k_{min}}$ and $\hat{k_{max}}$ as follows:

$$\begin{aligned} \hat{k_{min}} &= \frac{1}{2}\left(\frac{(k_{min}+k_{max})U - 2\beta_e}{\hat{U}} - \sqrt{\frac{Q}{\hat{U}^2}}\right) \\ \hat{k_{max}} &= \frac{1}{2}\left(\frac{(k_{min}+k_{max})U - 2\beta_e}{\hat{U}} + \sqrt{\frac{Q}{\hat{U}^2}}\right) \end{aligned} \tag{9}$$

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 distorted principal curvature (and other shape measures based on principal curvatures).

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 [5] 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:

$$\begin{aligned} Z_{XX} &= k_{min}\cos^2\theta + k_{max}\sin^2\theta \\ Z_{YY} &= k_{min}\sin^2\theta + k_{max}\cos^2\theta \end{aligned} \tag{10}$$

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

$$\hat{Z}_{\hat{X}\hat{X}} = \frac{U}{\hat{U}}Z_{XX} - \frac{2\beta_e}{\hat{U}}$$

$$\hat{Z}_{\hat{Y}\hat{Y}} = \frac{U}{\hat{U}} Z_{YY} \tag{11}$$

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

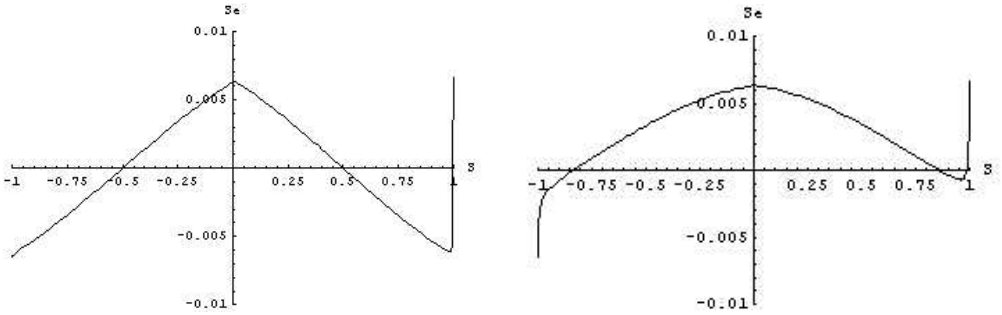
3.1.2 Shape index

From the expression of the distorted principal curvatures, the distorted shape index at the fixation point can readily be obtained by substituting (9) into (2):

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

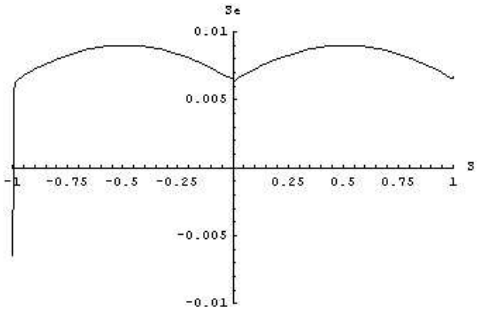
As the expression for $\hat{S}$ is very complex, we have assorted some diagrams to graphically illustrate the error in the shape index estimate with respect to the true shape index. The curvedness is fixed for each diagram so as to factor out the influence of curvedness. Figure 2 shows the distortion obtained by varying 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).

Other factors that may have an impact on the sensitivity of shape perception are studied in Figure 3. β_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 β_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, α_e seems to have little influence on the recovered shape index, as shown

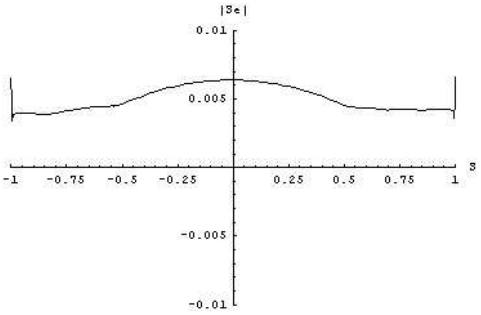


(a)

(b)



(c)



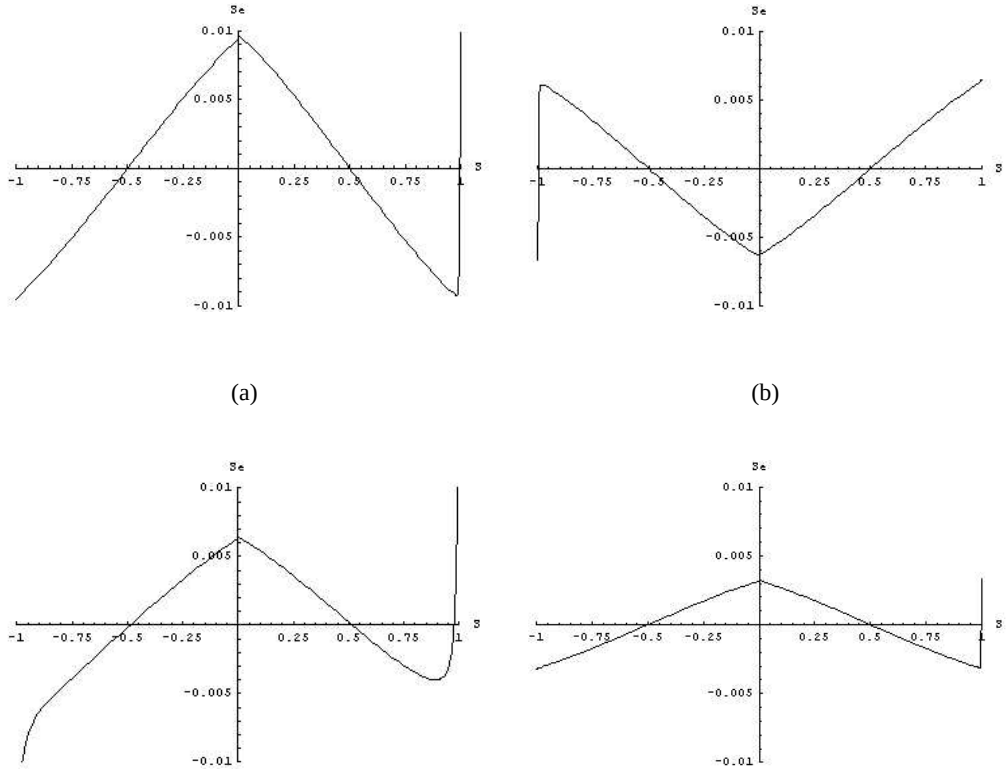
(d)

Figure 2: Distortion of shape index with different principal directions, plotted in terms of S_e against S. $\theta = 0$ 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 $\hat{U} = 0.18$.

in Figure 3(c). This anisotropy between α_e and β_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)).

3.2 Distortion under forward motion

In this section, we assume pure forward translation is performed. According to Equation (6), we have $\hat{U} = U = \hat{V} = V = 0$. We also assume $\hat{\gamma} = \gamma = 0$ as we did before. Adopting "epipolar reconstruction"



(a) (b)

(c) (d)

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 $\hat{U} = 0.18$ for all the diagrams.

approach, the distortion factor can be expressed as:

$$D = \frac{X^2 + Y^2}{X^2 + Y^2 + \alpha_e Z(X^2 Y + Y Z^2 + Y^3) - \beta_e Z(X Y^2 + X Z^2 + X^3)} \tag{13}$$

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. It is noted that at the fixation point, the distorted local shape measurements are undefined due to the undefined distortion factor (see Equation (13)). Our analysis in [4] 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.

3.3 Simulations

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.

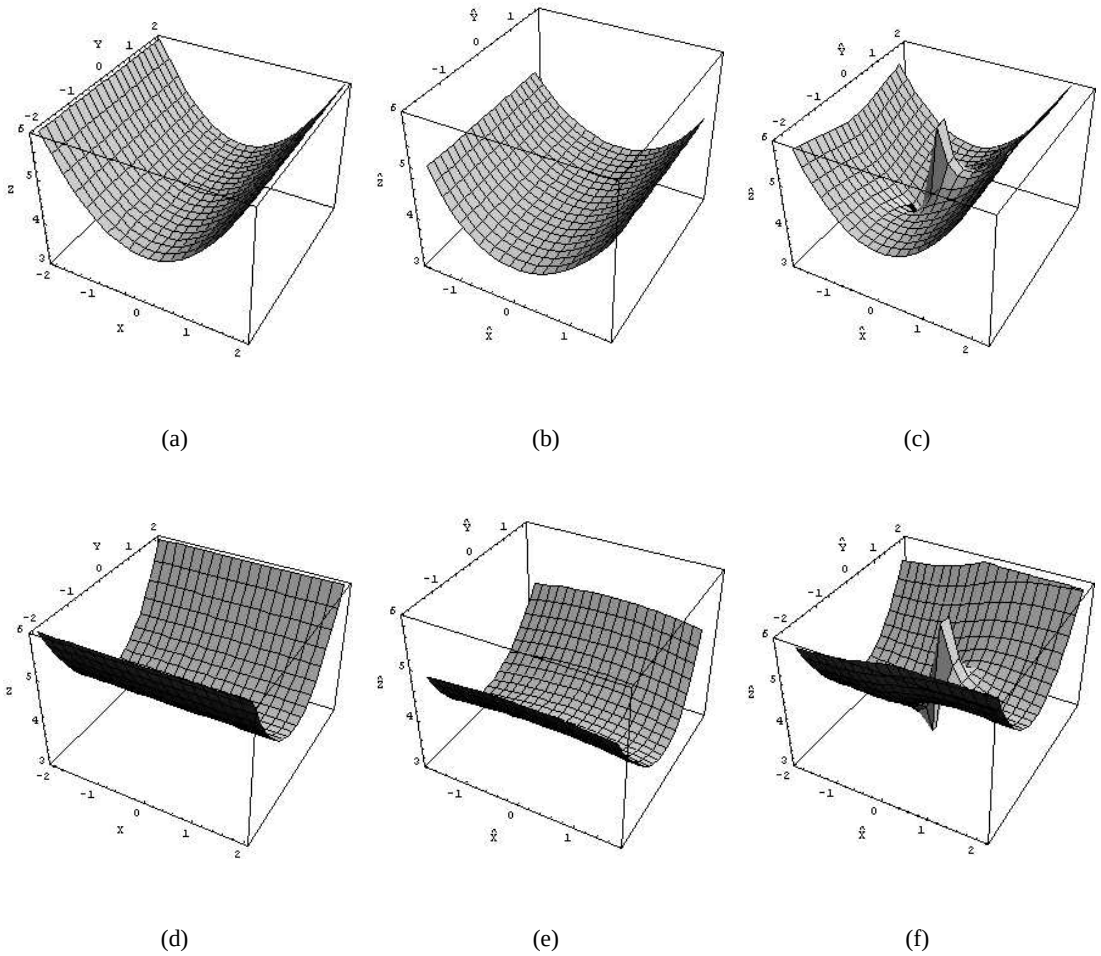
4 Experiments

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

4.1 Computer generated image sequences

SOFA [3] 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

[3] courtesy of the Computer Vision Group, Heriot-Watt University (http://www.cee.hw.ac.uk/∼mtc/sofa).



(a) (b) (c)

(d) (e) (f)

Figure 4: Distortion of cylinders under lateral and forward motions. (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^o for all the diagrams.

of 309 pixels and a field of view of approximately 45^o. 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 [12], 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 5(a)). Clearly the entire feature points on the top face of the cylinder in SOFA1 (which are delineated in Figure 5(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 $(\alpha, \beta, \gamma) = (-0.0203, 0.0284, 0)$. For the reasons we discussed in Section 2, we assume that current SFM algorithms can estimate W accurately (i.e. $\hat{W} = 0$). Given the 3-D motion estimation, depth is estimated as:

$$\hat{Z} = \frac{-\hat{f}(\hat{U}, \hat{V}) \cdot \mathbf{n}}{(u, v) \cdot \mathbf{n} - (\hat{u_{rot}}, \hat{v_{rot}}) \cdot \mathbf{n}} \tag{14}$$

Different SFM algorithms may give different errors in the estimated 3-D motion parameters. Instead of being constraint by one specific SFM algorithm, let us consider all the possible configurations in the 3-D motion estimation errors [4]. The resultant recovered depths are illustrated in Figure 5 using 3-D plot viewed from the side. In particular, Figure 5(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 5(c) depicts the shape recovered when $\beta_e > 0$ and $\hat{U} < 0$. According to Equation (11), we have $\hat{Z}_{\hat{X}\hat{X}} > 0$. It can be seen from Figure 5(c) that the plane was indeed reconstructed as a convex surface. Conversely, when $\beta_e < 0$ and $\hat{U} < 0$, concave surfaces were perceived (Figures 5(d), (e) and (f)). Comparing Figure 5(d) with Figure 5(e), we find that larger $\hat{U}$ in general resulted in smaller curvature distortion, whereas the results of Figure 5(d) with Figure 5(f) show that large β_e resulted in larger curvature distortion. Figures 5(c)–(f) show that the reconstructed surfaces were not curved in the Y direction. All these results confirm to our prediction made in Section 2.

[4] For our experiments, if the subspace algorithm proposed by Heeger and Jepson [7] was used, we obtained the reconstruction shown in Figure 5(c); if the algorithm proposed by Ma et. al. [13] was adopted, we obtained the reconstruction shown in Figure 5(d).

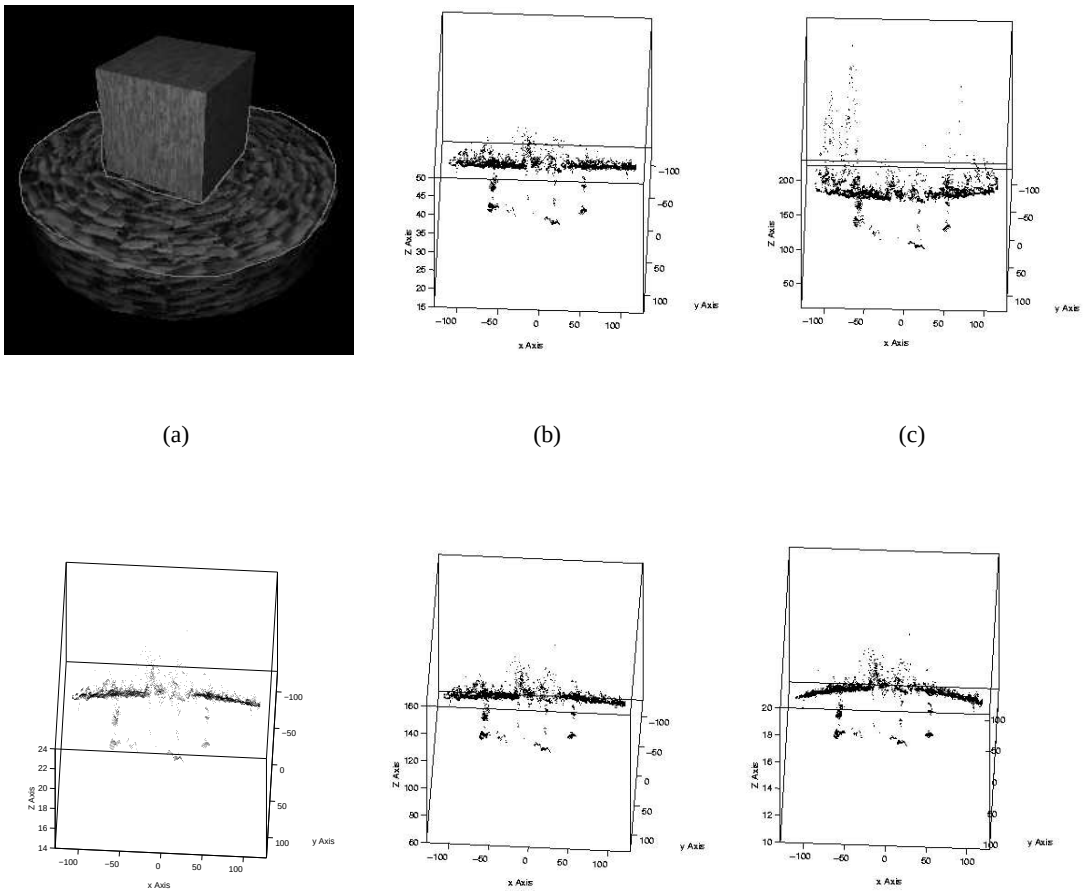


Figure 5: Computer generated lateral motion sequence and shapes recovered when different 3-D motion estimates were used for reconstruction. (a) SOFA1 frame 9 with the top face of the cylinder delineated; (b) Reconstruction with true motion parameters. (c) Reconstruction with $(\hat{U}, \hat{V}) = (-1, 0)$ 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 $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.01, 0.06, 0)$.

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 6(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 $(\alpha, \beta, \gamma) = (0, 0, 0)$ respectively. Current SFM algorithms have no difficulty in estimating the translation accurately for the SOFA5 sequence [17]. Therefore, the equation for reconstructing depth for each feature point would be:

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

The resultant recovered depths, given different typical errors in the 3-D motion estimates, are shown in Figure 6. Figure 6(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 6(c)). It is in accordance with our prediction that large depth distortion is expected when forward motion is performed.

4.2 Real image sequence

Shape 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 7(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 $(U, V, W) = (\beta_0 Z_0, 0, 0)$ and $(\alpha, \beta, \gamma) = (0, -\beta_0, 0)$ where β_0 was the rotational velocity of the basket ($\beta_0 > 0$ in this case)and Z_0 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 [4] 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 7. Figure 7(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

[5]courtesy of Dr Andrew Calway of Department of Computer Science, University of Bristol.

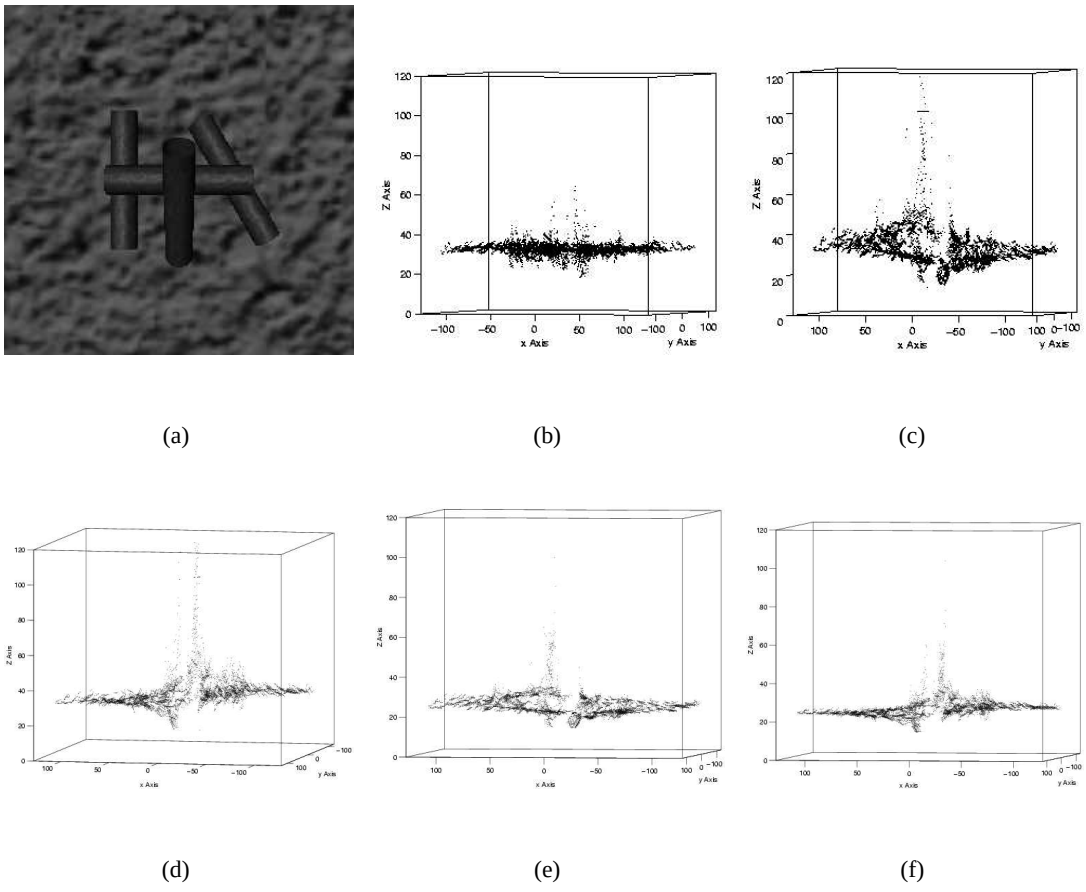
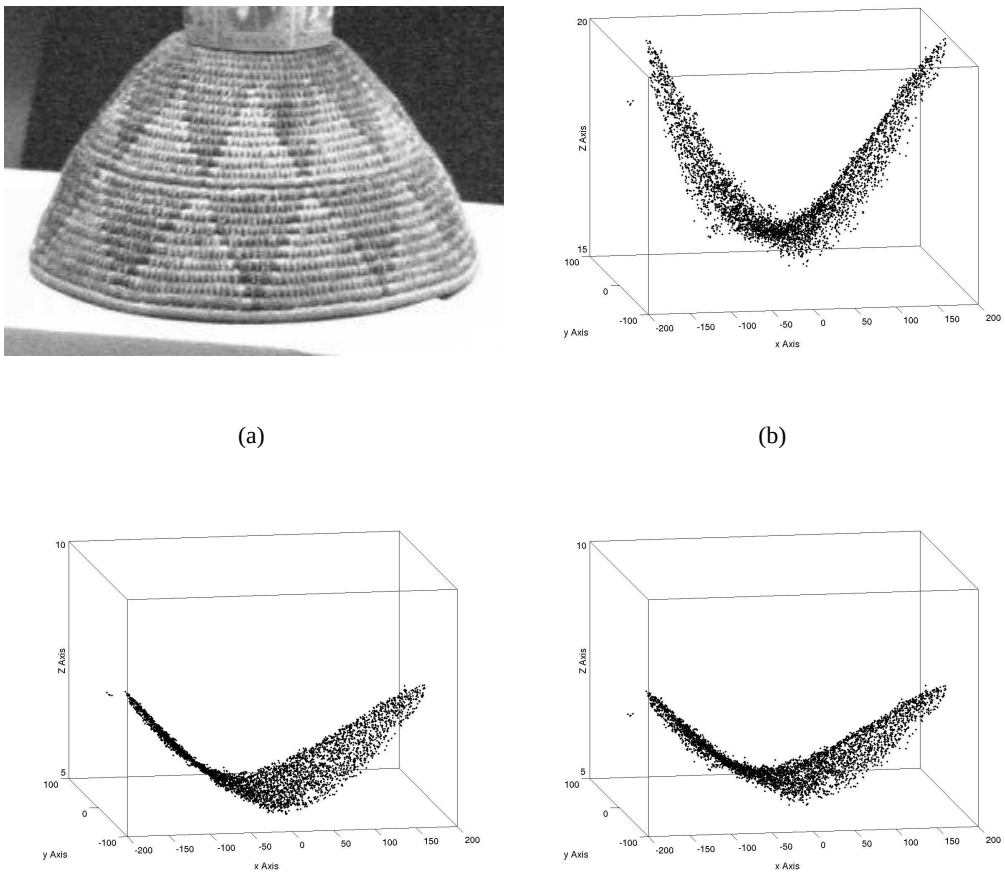


Figure 6: Computer generated forward motion sequence and shape recovered when different 3-D motion estimates were used for reconstruction. (a) SOFA5 frame 9. (b) Reconstruction with true motion parameters. (c) Reconstruction with $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.001, -0.001, 0)$. (d) Reconstruction with $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.001, 0.001, 0)$. (e) Reconstruction with $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.001, 0.001, 0)$. (f) Reconstruction with $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.001, -0.001, 0)$.

these terms into Equation (11), we obtained $\hat{Z}_{\hat{X}\hat{X}} < Z_{XX}$, which means that the recovered surface should be less convex than the true shape. Comparing the results of Figure 7(b) with the true shape measured by us manually, this is indeed the case. Furthermore, according to Equation (11), a smaller β_e or a larger $\hat{U}$ (with the signs of these terms unchanged) would result in an even less convex shape being recovered. Figures 7(c) and (d) clearly support our prediction.



(a) (b)

(c) (d)

Figure 7: Real lateral motion sequence and shapes recovered when different 3-D motion estimates were used for reconstruction. (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 $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (0.05, 0.05, 0)$.

5 Conclusions

This paper presented a geometric investigation on the reliability of shape recovery from motion cue given 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. When designing a visual system to perform shape recovery, these findings can help us choose the best motion strategy. In the meantime, 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. The work presented in this paper has taken a step towards this direction. More works need to be done to understand the robustness of shape recovery under more general 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

[1] G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, IEEE Trans. PAMI, 11(5) (1989) 477–489.

[2] J. Aloimonos, I. Weiss and A. Bandopadhay, Active vision, International Journal of Computer Vision, 2 (1988) 333–356.

[3] L-F. Cheong and K. Ng, Geometry of distorted visual space and Cremona Transformation, International Journal of Computer Vision, 32(2) (1999) 195–212.

[4] L-F. Cheong and T. Xiang, Characterizing depth distortion under different generic motions, International Journal of Computer Vision, 44(3) (2001) 199–217.

[5] A. Davies and P. Samuels, An introduction to computational Geometry for curves and surfaces, Clarendon Press, 1996.

[6] O. D. Faugeras, Stratification of three-dimensional vision: projective, affine, and metric representations, J. Opt. Soc. Am. A, 12 (1995) 465–484.

[7] D.J. Heeger and A.D. Jepson, Subspace methods for recovering rigid motion I: Algorithm and implementation, International Journal of Computer Vision, 7 (1992)95–117.

[8] J. J. Koenderink and A. J. van Doorn, Relief: pictorial and otherwise, Image and Vision Computing, 13(5) (1995) 321–334.

[9] J. J. Koenderink and A. J. van Doorn, Affine structure from motion, J. Optic. Soc. Am., 8(2) (1991) 377–385.

[10] J. J. Koenderink and A.J. van Doorn, Surface shape and curvature scales, Image and Vision Computing, 10(8) (1992) 557–564.

[11] H. C. Longuet-Higgins, A computer algorithm for reconstruction of a scene from two projections, Nature, 293 (1981) 133-135.

[12] B. D. Lucas, Generalized image matching by the method of differences, PhD Dissertation, Carnegie-Mellon University, 1984.

[13] Ma, Y., Košecká, J. and Sastry, S, Linear Differential Algorithm for Motion Recovery: A Geometric Approach, International Journal of Computer Vision, 36(1) (2000) 71–89.

[14] J. Oliensis, A critique of structure-from-motion algorithms, Computer Vision and Image Understanding, 80 (2000) 172–214.

[15] S. Ullman, The Interpretation of Visual Motion, MIT Press, Cambridge and London, 1979.

[16] J. Weng, T. S. Huang, and N. Ahuja, Motion and Structure from Image Sequences, Springer-Verlag, 1991.

[17] T. Xiang and L-F. Cheong, Understanding the behavior of structure from motion algorithms: a geometric approach, International Journal of Computer Vision, 51(2) (2003) 111-137.