THE STRONG ROLE OF CONTOUR GEOMETRY IN STRUCTURE-FROM-MOTION AND ITS IMPLICATIONS FOR PROJECTIVE CONSISTENCY

By

XIAOLI HE

A dissertation submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Psychology

Written under the direction of

Manish Singh

And approved by

———————————————

———————————————

———————————————

———————————————

New Brunswick, New Jersey

May, 2020

ABSTRACT OF THE DISSERTATION

The strong role of contour geometry in Structure-from-motion and its implications for projective consistency

by XIAOLI HE

Dissertation Director:

Manish Singh

Structure from motion (SFM) studies have shown that observers can perceive 3D structure in dynamic dot displays that are projected from rotating 3D objects. The SFM literature has focused on explaining how human observers 'reconstruct' the 3D structure from dynamic dot displays. Some analyses are based on the positions of dots, others are based on the velocity field. For a long time, there have been two implicit assumptions in SFM research: first, the perceived 3D structure should be projectively consistent with the 2D display; second, the computation of 3D percept does not depend on contour geometry, only on the motion profile of the dots. However, there have been some suggestions in the literature that neither of the two assumptions is correct (Ramachandran, Cobb, & Rogers-Ramachandran, 1988; Thompson, Kersten, & Knecht, 1992; Froyen, Feldman, & Singh, 2013; Tanrıkulu, Froyen, Feldman, & Singh, 2016). Under certain conditions, the perceived 3D structure is not projectively consistent with the 2D motion, and in those cases, the 3D percept is generally dominated by the contour geometry (despite the inconsistency with the motion). Nevertheless, the role of contour geometry in SFM remains largely ignored, and none of the existing SFM models (in both Psychology and Computer Science) incorporate its role. The current dissertation aims to provide more direct evidence against these two assumptions and investigate possible connections

between them.

In chapter 1, we overview the literature in the field of human structure-from-motion research. The studies on SFM have been either based on the position of individual dots, or on the analysis of motion profile. Both types of studies assume that the perceived 3D structure should be projectively consistent with the 2D display. Also, they seem to ignore the role of contour geometry in the determined 3D structures. We show that some studies have provided evidence that questions the above two assumptions. The aim of current dissertation is to explore both topics, and its potential connections. Our hypothesis is that both motion and contour geometry play important roles in SFM, and when the two sources of information conflict with each other, the influence of contour geometry can be strong enough to override the inconsisency in motion profile. Therefore, this may provide an explanation for why projective consistency is not a necessary requirement in SFM.

In chapter 2, we manipulate the dots' 2D speed profile in symmetric asymmetric and parellel shapes using a method introduced in He (2016), which allows us to systematically control the degree of projective consistency in the displays. The latter two types of shape contain projective inconsistency both in the speed profile and contour geometry. However, our results show that the volumetric percept with asymmetric displays is still vivid, providing strong evidence that projective consistency is not a necessary requirement in SFM. The similar strength of the 3D percept between symmetric and asymmetric displays also suggests that the global contour symmetry may not be as important as we thought in SFM. In these displays, the speed profile is locally consistent with contour geometry along each horizontal cross-section, so the zero-speed at the boundary suggests a slant of $90°$, which may serve as a cue to be an occluding contour of a 3D object. It suggests that instead of global geometry, we should probably focus on the local presence of 2D contour, which may itself provide a cue to the occluding contour of a 3D object, even if the speed profile does not go down to 0 there.

In chapter 3, we investigate the role of contour geometry by manipulating it separately from the speed profile of the dots. Specifically, we manipulate the shape of the aperture through which the same image motion is shown. Because of the absence of any occlusion cues, the shape of the aperture essentially defines the bounding contour of the SFM display. We start with a 'motion region' (such as rectangle, trapezoid, hexagon, barrel, diamond, ellipse) containing

dot motion consistent with 3D rotation, and then transform the shape of this motion region to define a smaller aperture. In many conditions, rather than looking like dots moving behind an aperture, the contour captures and defines the perceived 3D shape. The question is how much can the aperture shape be transformed and still determine the 3D percept (i.e. contour-defined percept which is now projectively inconsistent with the image motion). We used Method of Constant Stimuli to find the thresholds for the contour-defined percept. In Experiment 1, we generated smaller apertures by compressing the horizontal width of the motion region. We found that when the horizontal width ratio ω was as low as 0.2 (ellipse) or 0.6 (rectangle), the SFM percept was still dominated by the contour shape. This shows the importance of contour geometry in SFM. In Experiments 2 and 3, we broke coaxiality (the aperture shape and motion region no longer shared the same axis) by either translating or rotating the aperture within the motion region. We found that observers were quite sensitive to the deviation from co-axiality. In Experiment 4, we introduced variation of horizontal width ratio $\Delta\omega$ within the same display (such as a rotating sphere display shown through a triangle). We found no systematical effect of $\Delta\omega$. However, the results did show systematic effects of convexity and contour smoothness.

Overall: 1) These studies provide strong evidence that projective consistency is not a necessary requirement in SFM. Observers can readily perceive a 3D shape even when it is not projectively consistent with the SFM display; 2) These studies document the prominent role played by contour geometry in SFM, as well as some of the ways in which contour geometry and image motion interact in SFM. The prominent influence of contour geometry also in part explains why SFM percepts can fail to be projectively consistent with image motion. Rather than ignore contour geometry (as the SFM literature has largely done), the current study highlights the need for further systematic investigations into how contour geometry and image motion combine in SFM to generate a 3D percept. This in turn would make it possible to develop mathematical models that incorporate both of these cues in SFM.

Acknowledgements

Dedication

This dissertation is dedicated to my supportive family and my dear friends in the U.S. and in China without whom I cannot accomplish so much.

Table of Contents

List of Figures

Chapter 1

Introduction

1.1   Literature review of SFM studies

Structure from motion (SFM) refers to the phenomenon that human observers can perceive the three-dimensional structure from two-dimensional image sequences that contain only 2D motion information. It is closely related to the phenomenon of kinetic-depth effect (KDE) studied in the early twentieth century (Wallach & Oconnell, 1953). Like in Figure 1.1, a typical SFM display is a sequence of 2D images where dots move with different velocities in 2D, corresponding to points on the surface of an object undergoing some motion in 3D, usually a rigid rotation about an axis. The object transformation is usually rigid, or partially nonrigid in some more recent studies, and the projection can be perspective or orthographic. There are two characteristics in common in the displays: first, no 3D motion information is conveyed by each static frame; and second, such displays eliminate other cues such like shading, and observers continue to be able to perceive 3D structure from image motion when variations in texture density are also eliminated (Sperling, Landy, Dosher, & Perkins, 1989). Therefore, the only available information in the display is the image motion. Since the image projection loses one spatial dimension from the 3D object, there are infinite combinations of 3D structures and 3D motion that are consistent with the 2D image sequences—though some interpretations are simpler, and some can be very complicated. One interesting topic of SFM studies is to explore how people perceive 3D structures with limited 2D information.

Over the decades, researchers have been interested in how people perceive 3D structure from 2D SFM displays. One approach has been to investigate the SFM problem theoretically, i.e. to examine the conditions under which the SFM problem is mathematically solvable, and to find the minimum information required to get a unique 3D structure (Ullman, 1979, 1983; Bennett & Hoffman, 1985). These studies have also inspired work in computer vision, e.g.,
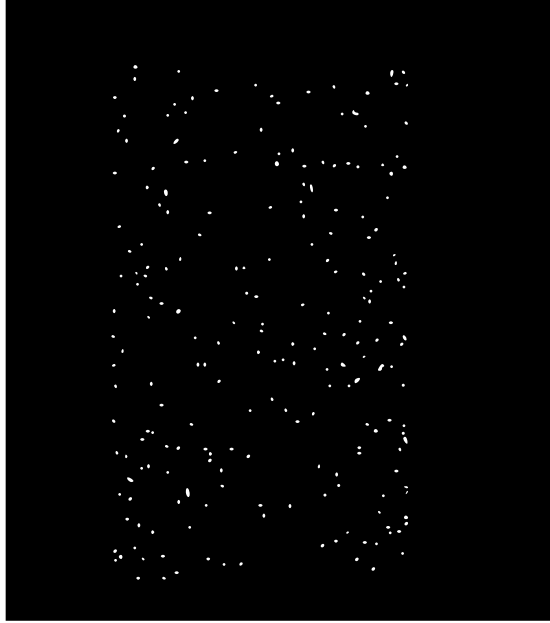
Figure 1.1: A typical SFM display of a rotating cylinder

(Tomasi & Kanade, 1992; Morita & Kanade, 1997; Anandan & Irani, 2002). On the other hand, researchers are interested in understanding human perception of 3D structure from image motion—what constraints and what information human observers use in SFM. They are concerned not only with whether certain mathematical models can in principle yield the 'correct' 3D solution, but also with how well humans can perceive 3D structure from image motion under different conditions. Based on the above two types of questions, many models have been developed over the last 60 years. There are mainly two kinds of models—position-based models and velocity-based models. Position-based models take as input the position of each dot in multiple static frames, whereas velocity-based models take as input the velocity field (i.e. the velocity at each image location), also known as optic flow.

1.1.1 Position-based Models

The most widely used SFM model is the rigidity-based model proposed by Ullman (1979). According to his rigidity assumption, in order to perceive 3D structure, the 2D display should satisfy the rigidity constraint that the distance between any two points on the surface of the 3D object should remain the same over time. Specifically, Ullman's theorem shows that as

long as rigidity is maintained, given as few as four non-coplanar points in three distinct orthographic views, the 3D structure of the four points can be uniquely determined (up to a depth reversal). This therefore provides the minimum information needed to solve SFM. For more complicated displays with more elements, the rigidity assumption can be applied by dividing the elements into many small groups of such four non-planar elements, and testing for a possible rigid solution within each group, and then combining these local structures into a global 3D representation. In summary, the model assumes that as long as the 2D display is projectively consistent with a rigidly moving object and certain minimal conditions obtain, the 3D structure and motion can be uniquely determined.

Ullman's rigidity-based model is a computational level of analysis of SFM. Since its publication, it has greatly inspired other theoretical research in human SFM as well as machine vision algorithms (Tomasi & Kanade, 1992; Morita & Kanade, 1997; Anandan & Irani, 2002). For human SFM studies, there have been many empirical studies exploring the minimum information needed by human observers to perceive a unique solution (number of views and number of points). Ullman proved that three distinct views of four non-coplanar points was the minimum requirement to solve SFM. However, later on, other minimum requirements were proposed using slightly different constraints. For example, Tsai and Huang (1984)'s analysis based on singular value decomposition (SVD) showed that the 3D structure and motion can be uniquely determined by seven points correspondences with only two perspective views. Hoffman and Bennett (1986), instead, showed that if two rigidly linked points are rotating along a fixed axis at a constant angular velocity, four orthographic projections of those two points can lead to a unique interpretation. These studies add other constraints related to rotation, which helps to reduce the required number of frames and/or number of elements.

In summary, the position-based models keep track of positions of specific points across different frames, and reach a rigid 3D interpretation. The main contribution of position-based models is that they show that SFM problem is mathematically solvable under certain well-defined conditions—it's a theory of competence. However, they tell us little about a theory of performance, i.e., do human observers actually use rigidity assumption and rigidity-based algorithms to perceive 3D structure? One challenge that position-based models face is the correspondence problem. Note that most stimuli used in the above studies contain only small

numbers of points in a few distinct views, because in order to test for rigidity, the algorithm needs to find the dots' correspondence between adjacent views. In other words, the position-based models need to know which dots correspond to which in the other frames. This leads to some limitations when trying to apply these models to empirical studies. For example, for situations involving more dots and more frames, the computational work of the above models will increase dramatically. More importantly, such models will fail when it's impossible to track individual dots, such as when the dots have limited lifetime (Sperling, Dosher, & Landy, 1990). In one extreme, if the dots in the display have lifetimes of only two frames, human observers still have high shape identification accuracy, but the position-based models will fail because it's impossible to track dots in this case (Dosher, Landy, & Sperling, 1989). These limitations imply that human visual system may be using other computations that do not simply rely on the positions of individual dots in multiple frames.

1.1.2 Velocity-based Models

Whereas the first class of models is based on analyzing dot positions as a function of discrete time (or frame number), the second is based on the analysis of velocity field (or optic flow). One advantage of using velocity field is that it does not require tracking correspondence of individual dots over time, which makes it more robust for the visual system. This way of encoding motion information may also be more natural from the point of view of biological vision. Indeed, there are some studies showing that human perception of SFM is based on velocity computations. For example, Treue, Husain, and Andersen (1991) found that the minimum dot lifetime required to perceive the 3D structure is around 50-85 ms, which is similar to the threshold for estimating 2D velocity. This suggests that the visual system uses velocity information to process SFM. Landy, Dosher, Sperling, and Perkins (1991) also showed that human observers use velocity to compute the 3D shape. They used different random dots stimuli depicting hills wiggling around a vertical axis. Some stimuli contained strong 1st-order energy, some only contained 2nd-order energy. Their results showed that weakening 1st-order energy correspondingly impaired the 3D shape discrimination, which implies that 1st-order motion is very important for SFM. There is also biological evidence that motion signal is important in SFM. Siegel and Andersen (1990)'s monkey studies showed that areas MT and MST (which are specific for processing motion

signals) were involved in SFM. The fMRI study by Paradis et al. (2000) also showed that the superior occipital gyrus (SOG; presumptive V3/V3A) and the left occipito-temporal junction had specific sensitivity to SFM dispalys depicting curved surfaces, compared to 2D coherent motion and random motion. Since these areas are specific to motion perception, they also imply that SFM involves analysis of the motion field in the image.

When researchers try to explain human SFM, most models are based on motion signal, and they have delved more deeply into the details of what and how motion signal is used in SFM.

Braunstein and Andersen (1984) ran a series of experiments and proposed the constant angular velocity heuristic. Instead of testing dots' consistency directly with a rigid interpretation, they assumed that the visual system analyzed the projected image motion (i.e. 2D speed profile) of the dots to test for consistency with 3D rotation. They found that if the dots' 2D speed profile was a cosine function along each horizontal cross-section, observers had high ratings for both depth and volumetric percepts. Therefore, they proposed that the visual system is biased to assume a constant angular speed for a rotating object.

Besides velocity, there was also another potential cue available to the 3D structure in their study —dot density. Braunstein and Andersen (1984) generated the displays by randomly and uniformly sampling dots on the surface of a 3D sphere, and projecting the dots onto a 2D image. Therefore, the density of dots in the image is not uniform and observers could have potentially used the density cue rather than motion to perceive the 3D shape. In order to eliminate the influence of density cue, Sperling et al. (1989) made the density uniform in the 2D display by introducing limited dot lifetime. They divided the whole display into a square grid. At each time step, they either added or deleted some dots from each square, in order to ensure that each square had the same number of dots. Some reserchers argued that Sperling's approach for density control might introduce another aritificial cue —a region-specific variation in scintillation (the central region will have more scintillation because it requires more dot insertions/deletions to make the density uniform).

In another paper (Sperling et al., 1990), they discussed avoiding the influence of these artificial cues by either measuring the strength of artificial cues alone (no motion) or by directly

removing these cues from the display. They found that each artificial cue alone (density or scintillation) was not sufficient for perceiving 3D shape from SFM displays. Also, they designed a 3D shape identification task, where they removed the dot density cue by introducing limited dot lifetime and controlled the strength of scintillation to be low (3%). After removing these artificial cues, their observers were still able to perceive the correct 3D structure. Their method eliminated the artificial cues, leaving alone the motion information, thereby showing that 2D velocity field was necessary and sufficient for SFM.

The above studies show that the global velocity field, is important in SFM. Moreover, a series of studies have shown that people mainly use first-order properties of optic flow in SFM (Koenderink & Van Doorn, 1986; James T Todd & Bressan, 1990; Landy et al., 1991). Indeed, when the second-order optic flow (acceleration) information was provided in the display, observers did not use it in SFM task, and they were only sensitive to the sign of accelerations (James T Todd & Bressan, 1990; Norman & Todd, 1993).

Mathematically speaking, the Euclidean structure can only be calculated veridically from first order-optic flow (velocity) together with second-order optic flow (acceleration), whereas first-order properties alone can only provide the affine structure that contains an infinite number of interpretations that fall in an affine family (James T Todd & Bressan, 1990). If the visual system is only sensitive to first-order property, i.e., velocity, how does it estimate 3D structure from a velocity field? Most of the studies mentioned above have shown that people are only sensitive to first-order properties of optic flow, and that is only sufficient to get an affine structure, but little is known about how observers get the unique solution from the affine family. Domini et al.'s work provides a probabilistic account showing how the visual system can compute a consistent and unique percept based on a simple heuristic.

The optic flow can be decomposed into three differential invariants—*curl*, *shear* (deformation, or *def*) and *div* (divergence) (Koenderink & Van Doorn, 1986) (see Figure 1.2, from Domini and Caudek (2003a)). *curl* refers to rigid image rotation, *div* is the isotropic expansion or contraction, and *def* is the amount of shape change over time. *curl* and *div* are also influenced by other factors (Todd, 1995), but *def* is only influenced by the object's 3D structure. In other words, *def* is very informative about the local 3D structure such as relative slant. Domini and Caudek (1999) showed that the component *def* is related to the surface slant and rotation

speed.
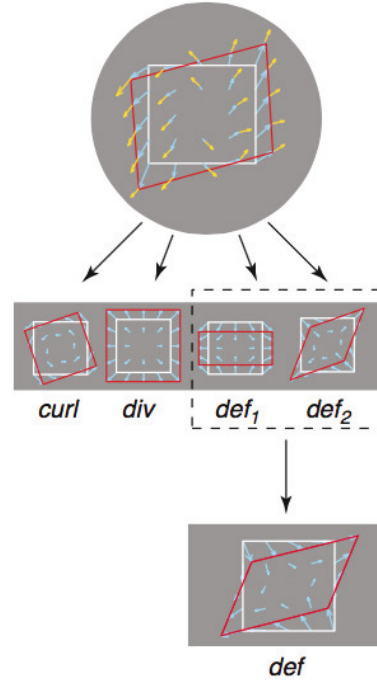
Figure 1.2: Decomposition of first-order optic flow into *curl*, *div* and *def* (from Domini and Caudek (2003a), Figure 1c)

Domini, Caudek, and Proffitt (1997), Domini and Caudek (1998, 1999), Domini, Vuong, and Caudek (2002), Domini and Caudek (2003a), Di Luca, Domini, and Caudek (2004) considered a planar patch rotating rigidly in depth along a vertical axis. In their model, they showed that for a linear velocity field that has deformation (non-zero *def*) and no translation, the deformation $def = \sigma\omega$, where σ is the local surface slant, and ω is the angular speed of rotation. The equation indicates that there are infinite combinations of σ and ω that can generate the same value of *def* (those pairs lie on the hyperbola defined by the above equation). In other words, based on first-order optic flow (i.e. the instantaneous velocity field), one cannot get a unique solution for surface slant and angular speed. This is consistent with the affine structure analysis mentioned above. However, in their experiments, Domini et al. manipulated *def* and ω independently and found that the perceived slant is an increasing function of *def*, despite the fact the 'correct' slant (from the actual display) remained the same across different *def* values.
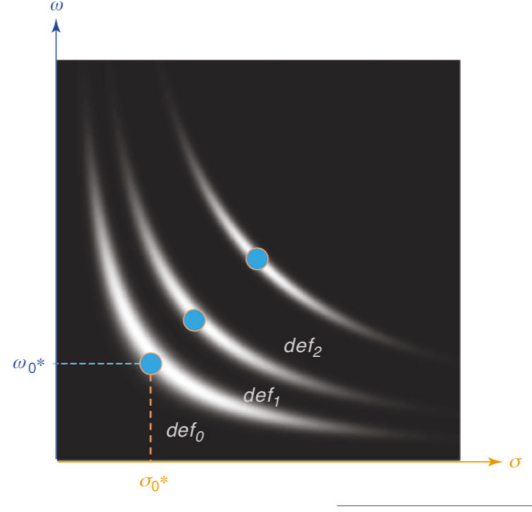
Figure 1.3: The probability functions of $p(\sigma, \omega | def)$ are coded by gray levels in which white represents the highest value and corresponds to the most likely σ, ω pair given def (red-blue dots). (from Domini and Caudek (2003a), Figure 3b)

Based on these results, they proposed a heuristic model based on the equation $def = \sigma\omega$: they assume that there is Gaussian noise for the measurement of def, so the distribution of def, as a function of local slant σ and angular speed ω, is not uniform and has the shape like Figure 1.3 (from Domini and Caudek (2003a). They mathematically prove that, under these conditions, the likelihood function has a unique maximum at $\sigma = \omega = \sqrt{def}$. Their experimental results show that human observers choose the pair of (σ, ω) that maximizes the likelihood of def, and the perceived slant σ is an increasing function of def (though the veridical local slant was constant).

In summary, there are two types of models—position-based and velocity-based models. The main difference between them is how the motion information is represented/used. The position-based models use the dots positions in multiple frames, whereas the velocity-based models analyze the optic flow, such as velocity field. Moreover, the position-based models emphasize more on finding the unique mathematical solution based on some constraints such as rigidity, and they also focus on finding the minimum conditions (number of dots and numbers of frames) required to get a unique interpretation. In other words, they try to solve SFM as a math problem, but they are less concerned with whether the analysis is actually used by human perception or not (at least the initial studies along these lines were less concerned).

Velocity-based models, on the other hand, focus more on how the human visual system deals with SFM. They assume that the visual system uses velocity field in SFM. Some studies, for example, have shown that humans are not sensitive to higher-order optic flow and only uses first-order optic flow properties (James T Todd & Bressan, 1990; Koendreink, 1991). From a mathematical point of view, if observers only use first-order properties, they can only perceive affine structures from 2D displays. Beyond the stage of affine analysis, Domini et al. propose that observers choose the most likely solution based on a simple heuristic that relates the local surface property and angular rotation.

1.2 Two prevalent assumptions in the SFM literature

Although the SFM literature has used a number of approaches, two themes stand out as being almost universal. First, SFM models assume that the perceived 3D structure must be projectively consistent with the image motion. Second, the literature ignores any possible contribution of contour geometry, essentially assuming that it does not contribute to perceived 3D structure in SFM.

1.2.1 Projective consistency

First, the SFM literature has assumed that the perceived 3D structure must bt projectively consistent with the 2D motion. We call this projective consistency. The position-based studies seek a unique solution by assuming some constraints such as rigidity. For example, according to rigidity-based models, the reconstructed 3D structure should be consistent with a rigid configuration while being projectively consistent with the given image sequence. The velocity-based studies, on the other hand, assume that the perceived 3D structure should be consistent with the 2D velocity field in the display. For example, the constant angular velocity heuristic assumes that the human visual system is biased to use constant angular velocity to get the 3D structure. Regardless of the concrete form and input of each model, the literature has assumed that the perceived 3D structure should be projective consistent with the image motion.

In fact, projective consistency is an implict assumption underlying almost all of vision research, far beyond SFM. Vision is commonly viewed as an 'inverse' problem: there is one

spatial dimension missing in the 2D retinal image projected from a 3D scene, and the aim of vision is to infer the 'correct' 3D scene from the 2D projection. Therefore, it is natural to take projective consistency as a necessary assumption for all models: whenever you are seeking a solution, it should be projectively consistent with the 3D 'ground truth'. For example, when you see a picture of a railway track, although the two lines are not parallel in the picture, you will automatically interpret it as a parallel railway, because you know that the picture is projectively consistent with a parallel track under perspective projection.

However, there is some evidence that calls into question the assumption of projective consistency in SFM.

Some studies have shown that linear motion can induce a percept of 3D rotation, despite the fact that the constant speed in the image is not projectively consistent with 3D rotation. In Froyen et al. (2013) 's figure-ground studies, the 2D random dot texture within several vertically oriented columns was moving with constant speed in opposite directions. They found that the linear motion could induce a vivid percept of 3D rotating columns, suggesting that rotation percept does not require projectively consistent 2D motion. Thompson et al. (1992) showed a display where a central rectangular region contained dots moving at constant speed, and the surrounding region contained dots moving at constant speed in opposite direction. They showed that when the rectangle was narrow enough, it was percieved as a rotating cylinder despite the constant speed of the dot motion.

In He (2016), we directly manipulated the 2D speed profile along multiple cross-sections within an ellipse, so that the speed was a linear combination of cosine speed component (which is projectively consistent with 3D rotation) and a constant speed component (which is not projectively consistent). We applied such speed profile to dots within an ellipse, and also manipulated the motion direction from $0°$ to $60°$. We found that when the speed contained as high as 40% of inconsistent motion, the 3D percept was still vivid. Also, we found that observers had large tolerance to the deviation of motion direction: when the motion direction was around $30°$ away from horizontal, the 3D percept was still strong.

In summary, the above studies have questioned the role of projective consistency in SFM. This shows that the 3D percept may not require projectively consistent image motion, and suggests that it is not necessary in SFM.

1.2.2   Role of contour geometry

Second, traditional SFM studies only focus on the analysis of motion profile, leaving little to no room for the role of contour geometry. The position-based studies try to get the 3D structure by applying some constraints on the position of individual elements. For example, the pairwise distances between individual elements should remain constant throughout rotation (rigidity assumption). They build the global structure from local estimates using those constraints, and treat all elements equally in the display, mentioning no role of contour geometry. The velocity-based studies try to estimate the 3D object by analyzing the whole velocity field. For example, Braunstein's constant angular velocity heuristic is based on the fact that the projected speed of a rotating 3D object (with constant speed) should have a cosine speed profile along each slice in the direction of motion, and does not include any contribution of the specific shape of contour. The studies based on decomposition of first-order optic flow (Koenderink & Van Doorn, 1986; Norman & Todd, 1995; Domini et al., 2002; Domini & Caudek, 2003a; Di Luca et al., 2004), build up the global structure from local analysis of optic flow deformation. They recover the local slant and angular speed from the deformation of optic flow, and gradually integrate those local measurements into a single global solution. Again, in both local measurement and global integration, no role is played by the geometry of the occluding contour. Even in the surface interpolation literature in SFM, little has been mentioned about the role of 2D contour. Usually a surface interpolation algorithm (such as Treue, Andersen, Ando, and Hildreth (1995)) interpolates a smooth surface from the estimated 3D coordinates of dots. It uses the relative depth between elements and the smoothness constraint to interpolate a surface between 3D dots. The interpolation is not dependent on the 2D contour geometry (e.g., they do not require the interpolated surface to pass through the 2D contour).

However, there is some evidence that the role of contour geometry cannot be ignored. In fact, the importance of contour geometry may be closely related to the issue of projective consistency. Whenever projective consistency is violated in SFM, the contour geometry seems to determine the 3D percept. Ramachandran et al. (1988) showed an interesting demo in which a rotating cylinder was occluded by a triangular aperture that has the same base edge as the cylinder, and there was no relative depth or other cues about the presence of the aperture. They

noted that this display tends to be perceived as a rotating cone instead of a rotating cylinder. This was surprising given that the percept was not consistent with the 3D object that generated the image motion, and more importantly it shows that the perceived 3D structure was dominated by the shape of contour, when there was inconsistency between motion and contour shape. Moreover, in the figure-ground study (Froyen et al., 2013), they found that the regions with convex contour were more likely to be perceived as rotating in 3D. It suggests that some shape properties tend to make the 3D percept stronger.

1.3 Motivation and current approach

Thoughout the literature, projective consistency has been a prevalent assumption in SFM, and the role of contour geometry has been largely ignored, essentially assuming that it plays no role in the perception of 3D structure in SFM. There is some evidence to suggest that neither of the two assumptions is correct in SFM. However, these issues have not been systematically investigated and there are many unanswered questions.

First, some evidence comes from different areas, which involves displays where there may be other cues aside from SFM. For example, in the figure-ground studies, one could argue that the accretion/deletion cue along boundary pushes one region to be figural and the other one to be background, and this competition promotes the 3D percept.

Second, there is no systematic analysis of these two issues. Ramachandran et al. (1988) only presented an interesting demonstration, but did not investigate the phenomenon systematically (such as by manipulating the properties of the speed profile or the aperture shape in an experiment). And interestingly this issue was not followed up anyone in subsequent years.

Third, note that the two notions often seem to go together—when projective consistency is violated, contour geometry seems to dominate the percept (Ramachandran et al., 1988; Froyen et al., 2013). This suggests a possible connection between them. However, none of the studies has tried to connect the two notions together.

The current study attempts to fill in the gap in the study of SFM with respect to the role of projective consistency and the role of contour geometry in SFM. Our hypothesis is that both motion and contour geometry play important roles in SFM, and when the two sources of

information conflict with each other, the influence of contour geometry can be strong enough to override the fact that the image motion is not projectively consisent with the 3D shape. This dominant effect of contour geometry can thus lead to 3D percepts that are projectively inconsistent with the image motion. Therefore, this may provide an explanation for why projective consistency is not a necessary requirement in SFM.

For a long time, the role of projective consistency in SFM has remained unquestioned, and the role of contour geometry has also been largely ignored. This may in part be due to the way in which SFM displays are typically generated. In traditional SFM literature, the displays are generated from the projection of a real or simulated 3D object that is rotating. Therefore, the 2D display is by construction consistent with the 3D object —both in terms of speed profile and 2D contour shape. Thus, the information contained in contour is always consistent with the motion. In the other word, the contribution of contour is covered by 2D motion. This makes it impossible to study the respective contributions of both factors. We therefore use a different approach where motion and contour information can be manipulated independently, which enables us to study the two factors directly and systematically.

In the first study (chapter 2), we start by defining the 2D speed profile directly, so the display is not necessarily projectively consistent with 3D rotation. By manipulating the proportion of cosine speed profile (i.e., α, which is projectively consistent with an ellipsoid rotating with constant speed under orthographic projection, see Figure 2.2 for details.) in a linear mixture with constant speed profile, we can control the degree of projective consistency in the display. Moreover, we can define SFM displays with asymmetric shapes. In the first study (Chapter 2), we investigated the topic of projective consistency by introducing stronger projective inconsistency in terms of contour shape (in addition to the projective inconsistency with respect to the speed profile). In particular, we used symmetric and asymmetric 2D contours, and the asymmetric displays have an especially high degree of projective inconsistency because the contour does not change throughout rotation, which is not physically possible with an actual asymmetric 3D object that is rotating. We also introduced different proportions of projectively consistent motion in each type of contour. In a yes/no task, we asked observers if they saw a volumetric rotating object, and used method of Constant Stimuli to find the α threshold for each type of contour. Asymmetric contours contain a greater degree of projective

inconsistency, so we predicted that they should have larger α thresholds, meaning that they would need higher proportions of projectively consistent motion to be perceived as 3D.

In the second study (chapter 3), we focus more directly on the role of contour geometry. Inspired by Ramachandran's demo, we separate the region that defines the speed profile and the smaller aperture through which the motion display is seen. We call it an aperture for easier understanding, but there are no depth or occlusion cues between the aperture and dot motion, and so this 'aperture shape' essentially defines the contour geometry of the SFM display. The separation of two regions provides larger freedom for us to play with projective consistency. More importantly, the definition of motion is independent from the contour shape, so we can study the contribution of contour geometry more cleanly. We try to explore under what conditions would contour geometry dominate the 3D percept. We manipulated the horizontal width ratio ω between the two regions in experiment 1, and broke the coaxiality between the rotation aixs of motion region and symmetric axis of aperture (2D bounding contour) by translation or rotation in Experiment 2 and Experiment 3 (but still keeping ω constant within each display, i.e, the motion region and aperture shape belongs to the same shape class). In the last experiment, we introduce variation of horizontal width ratio as well as convexity/concavity. The variation of horizontal width ratio ω means that the aperture and the motion region do not belong to the same shape class. We predict that the contour geometry dominates the 3D percept at first, but that its dominance would decrease as we introduce greater inconsistency (such as smaller ω, larger deviation from coaxiality, as well as larger variation of ω). Also we used several shape classes: rectangle, trapezoid, hexagon, diamond, barrel, ellipse. The exact shapes used vary across different experiments because of the constraints of the particular experiment. The shapes vary in the following two properties: variation in horizontal width and smoothness.

Chapter 2

The role of projective consistency in SFM

2.1 Introduction

2.1.1 Phenomenon of Structure from motion

Structure from motion (SFM) refers to the phenomenon that human observers can perceive three-dimensional structure from a two-dimensional image sequence that is consistent with the 3D structure, and the 2D displays only contain motion information (no shading or other depth cues). It is closely related to a phenomenon called kinetic depth effect (KDE) studied in the early twentieth century. A typical SFM display is a sequence of 2D image projections where dots move with different velocities in 2D, corresponding to points on the surface of a 3D object undergoing some motion in 3D, usually rotating about an axis. The objects can be rigid, or partially nonrigid (such as bending, stretching, deforming), and the projection can be perspective or orthographic. There are two characteristics in common in the displays: first, no 3D information is conveyed by each static frame; and second, such elements eliminate other cues like shading, texture or color. Therefore, there are many possible combinations of 3D structures and motions that are consistent with any 2D image sequence—though some interpretations are simpler, and some can be very complicated. One interesting topic of SFM is to explore how people perceive 3D structure from limited 2D information.

2.1.2 Projective consistency—the implicit assumption behind SFM research

Over the decades, researchers have been interested in what constraints and what information human observers use in SFM. There are mainly two types of theories—position-based models and velocity-based models. As the names show, position-based models take as input the position of each dot in multiple frames, whereas velocity-based models take as input the whole

velocity field (i.e. the velocity at each image location), also known as optic flow. No matter which input they are based on, there is one hidden assumption—projective consistency. This means that these models are seeking a solution of a combination of 3D structure and 3D motion that is projectively consistent with the 2D display. Here, we will briefly introduce several well-known SFM models, and their underlying assumption of projective consistency.

The most famous SFM model might be Ullman's rigidity-based model (Ullman, 1979). The model is based on the assumption that the interpretation of the visual system has a bias towards the interpretation of rigid objects, and it shows that as long as rigidity is maintained for as few as four non-coplanar points in three distinct orthographic views, the 3D structure of the four points can be uniquely determined. Ullman also showed that the algorithm can deal with more complicated displays by applying rigidity assumption on small subsets of such four non-planar elements and then combining the local structures into a global 3D representation.

Later on, the rigidity assumption has been relaxed to apply to certain non-rigid SFM displays that deviate a little bit from the strict rigid SFM, such as adding a small proportion of non-rigid distortion aside from the rigid motion (Ullman, 1983). The modified model starts with a flat internal model, and as new frames come in, it updates the model by allowing minimum nonrigid transformation, so that it will fit the display as much as possible and also maintain the maximum rigidity at the same time. Ullman showed that the scheme can converge to a unique solution with both strictly rigid and nonrigid SFM displays that have moderate deviations from rigidity.

Since its publication, the rigidity assumption has influenced a large variety of SFM algorithms, both in computer vision (Tomasi & Kanade, 1992; Morita & Kanade, 1997; Anandan & Irani, 2002), and in human studies (Hoffman & Bennett, 1986; Dosher et al., 1989; Sperling et al., 1990). In summary, Ullman's rigidity assumption and other rigidity-based models aim to find the interpretation which is projectively consistent with a rigid rotating object, and if such a solution is possible, according to the theory, it should be chosen by the visual system.

Unlike position-based models, the velocity-based models take the whole velocity field as the input, and therefore, they do not have to deal with the correspondence problem explicitly—the correspondence of dots between adjacent frames. Different velocity-based models emphasize different ways that human observers can use the motion information.

Braunstein and Andersen (1984)'s constant angular velocity heuristic assumed that the visual system analyzes the projected image motion (i.e. 2D speed profile) of the dots to test for consistency with 3D rotation. They found that if the dots' 2D speed profile was a cosine function along each horizontal cross-section (assuming that the object is centered at the origin), observers have high ratings for both depth and volumetric percepts.

Another velocity-based model is based on the motion perspective cue that velocity is inversely proportional to the depth between object and observer under perspective projection (Sperling et al., 1989). Based on this relationship, observers can assign relative depth according to the minima and maxima of velocities. Studies have shown that the motion perspective model prediction is consistent with human percepts with rigid SFM displays (Sperling et al., 1989; Liter, Braunstein, & Hoffman, 1993), as well as certain types of nonrigid SFM (Jain & Zaidi, 2011). However, note that it only works under perspective projection; with orthographic projection, there's no such cue.

In summary, the above two types of velocity-based models try to find a solution that is consistent with the projected velocity from a 3D object, although the concrete constraints they are based on are different—either constant angular speed which is consistent with a rotating object, or the relative velocity based on motion perspective cues.

Another velocity-based approach considers the analysis of optic flow. We know that the relative motion between observer and a 3D object will generate some optic flow in the image, and the optic flow can be decomposed into three differential invariants–*curl*, *shear* (*def*) and *div* (divergence)(Koenderink & Van Doorn, 1986). *curl* corresponds to rotation in the image, *shear* is the amount of shape change over time, and *div* is the isotropic expansion or contraction. Norman and Todd (1995) found that *curl* and *div* are also influenced by other factors, but *def* is only influenced by the object's 3D structure. In other words, *def* is very informative about the local structure such as relative surface slant or curvature, and it can be used to recover the 3D structure of an object. Koenderink and Van Doorn (1986) showed that *def* can also provide a partial solution to an object undergong bending deformation. As with the other approaches, this decomposition approach assumes that the solution given by the visual system should be projectively consistent with the optic flow.

To sum up, a common assumption underlying SFM models is projective consistency.

The rigidity-based models assume that the solution should preserve rigidity of the 3D object, and if the display contains deviations from rigidity, it should at least approximately preserve local rigidity. The velocity-based models, on the other hand, look for a solution that is consistent with the projected velocity field—focusing either on the speed profile (such as constant angular velocity, motion perspective cue), or the optic flow generated by the relative motion between an observer and a 3D object.

In fact, projective consistency is an implict assumption underlying almost all of vision research, far beyond SFM. Vision is commonly viewed as an 'inverse' problem: there is one spatial dimension missing in the 2D retinal image projected from a 3D scene, and the aim of vision is to infer the 'correct' 3D scene from the 2D projection. Therefore, it is natural to take projective consistency as a necessary assumption for all models: whenever you are seeking a solution, it should be projectively consistent with the 3D 'ground truth'. For example, when showning a 2D image of a rectangle, you are seeking solution from the affine family of polygons in 3D which can generate the 2D rectangle.

2.1.3 Projective consistency is not a necessary requirement

However, there is some evidence suggesting that projective consistency may not be a necessary requirement in SFM.

Some evidence comes from figure-ground studies. In a series of studies by Froyen et al. (2013), Tanrıkulu et al. (2016), they used displays containing alternating light and dark vertical regions, where random-dot textures moved horizontally at constant speed but in opposite directions in alternating regions. Surprisingly, the regions that were percieved as figural were also perceived as 3D columns rotating in depth. As noted above, the speed in those figural regions was constant, which is not consistent with rotating columns, but the observers still perceived those constant speed regions as 3D rotating columns. Hence, in the context of figure-ground studies, their work suggests that sometimes human percepts are not projectively consistent with the image input.

One may argue that in figure-ground stimuli, the figural regions are 'pushed' to be in front and look like rotating objects because of the figure-ground competition along the shared boundary. However, some evidence also comes from SFM studies where the displays contain

only a single region and there is no figure-ground competition. Ramachandran et al. (1988) mentioned an interesting demo containing a random-dot SFM display generated from a rotating cylinder. When the whole display was visible, their observers perceived it as a rotating cylinder. However, when the display was occluded by a triangular aperture (with its base equal to the width of the cylinder, so only a portion of the display was visable), surprisingly, the observers reported it as a rotating cone, despite the fact that the speed profile of the display was projected from a rotating cylinder. In this demo, due to the influence of contour geometry, the perceived structure was not consistent with the actual speed profile in the image.

It may sound unbelievable at first glance that projective consistency is not necessary, since it seems like a natural basis for visual inference. Domini et al. (1997), Domini and Caudek (1998, 1999), Domini et al. (2002), Domini and Caudek (2003b), Di Luca et al. (2004) showed that people seem to use other rules. They used a display where a planar patch was rotating rigidly in depth along a vertical axis. They expressed the deformation *def* as a product of σ (local surface slant) and ω (rotating angular speed): *def* = $\sigma\omega$. Thus there are infinitely many combinations of σ and ω that can generate the same value of *def* (those pairs lie on the hyperbola defined by the above equation). In other words, based on first-order optic flow (i.e. the instantaneous velocity field), one cannot get a unique solution for surface slant and angular speed. Their experiment manipulated *def* and ω independently and kept slant σ the same across different conditions, and they found that the perceived slant σ is an increasing function of *def*, despite the fact that the 'actual' σ (based on the actual 3D stimuli) remained the same. They proposed a heuristic model based on the relation between local surface slant and angular speed: they assume Gaussian noise for the measurement of *def*. They mathematically proved that if one assumes a noisy estimate of *def*, the likelihood surface has a unique peak, rather than being equally high at all points along the hyperbola (see Figure 1.3). Their empirical studies showed that human observers choose the pair of (σ, ω) that maximizes the likelihood of *def* (the peak of hyperbola), regardless of the ground truth. Domini and Caudek (2003a)'s work gives us some insights that instead of looking for a solution that is consistent with the ground truth, our visual system chooses the most likely solution given the 2D displays.

Our previous studies (He, 2016) questioned the necessity of projective consistency more directly. Unlike the previously reviewed studies, our studies manipulated the 2D speed

profile directly. To determine each dot's actual 2D speed, at a given horizontal cross-section, we took a linear combination of the cosine speed profile (which is projectively consistent with an ellipsoid rotating with constant speed under orthographic projection) and a constant speed profile. We varied the linear weight of the cosine speed so that the displays contained different degrees of projective consistency (see details in Experiment 1 **Stimuli and design**). We applied such speed profile to randomly sampled dots within an ellipse region. In addition, we also varied the angle between dots' motion direction and the ellipse's primary axis. We found that when the 2D speed profile contained only 60% of the projectively correct motion, our observers already perceived the 2D display as a vivid rotating ellipsoid. And we also found that our observers were not very sensitive to the deviation of correct motion direction (the angle could be up to 30° before it resulted in a visible performance difference). All in all, our studies showed that the observers could tolerate large deviations of motion from the projectively correct stimuli in terms of the speed and motion direction. Thus it suggests that projective consistency is not necessary to perceive a vivid 3D structure.

2.1.4   Current approach

In our previous studies, we only explored projective inconsistency in terms of the 2D speed profile. Also, most SFM studies only focus on the motion information and ignore the role of 2D contour. However, note that the inconsistency can also come from the 2D contour. Ramachandran et al. (1988)'s demo is a piece of clear evidence suggesting the importance of 2D contour, but they only reported an interesting phenomenon, with no systematic experiments. This points to the need for a detailed exploration on projective consistency in terms of the 2D contour. Inspired by our previous approach, here we continue to manipulate the amount of projective inconsistency in the 2D display by the manipulation of 2D speed profile, and also in terms of the 2D contour.

Symmetry is quite commonly seen in daily life, in both living things (human body, trees, planets, etc.) and manufactured objects (chairs, windows, etc.), and it contains useful information about the object's skeletons, and therefore contains useful information about their 3D structure (Blum, 1973; Feldman & Singh, 2006). In SFM displays, we know that a rotating object will change its projected silhouette over time, unless it is rotationally symmetric and

rotating about its axis of symmetry. In other words, if the SFM display is asymmetric in 2D, and stays fixed over time, it will introduce strong inconsistency with an actual rotating object, and we can manipulate the degree of inconsistency by manipulating the degree of symmetry (it is actually asymmetry, but for simplicity, let us call it symmetry). In particular, in the current studies, we use different splines to create the left and right portions of the 2D contour, and close the shape with straight lines at the top and bottom. We manipulate different levels of symmetry by varing the parameters of the splines. In the first two experiments, we explore the role of global symmetry, while in the third experiment, we explore further by introducing parallelism (local symmetry). Within these shapes we added moving dots with similar speed profiles as in our previous studies. The task for the observers is a Yes/No task that whether they perceive the display as a solid volumetric rotating object. We calculate the threshold of cosine-speed weight for each level of symmetry, which means the minimum amount of projectively-consistent motion needed to override the inconsistency from 2D contour and have a vivid 3D percept. In other words, these 'thresholds' indicate the amount of tolerance obsevers can have about deviations from projective consistency. Smaller threshold means greater tolerance. Since symmetric shapes have less inconsistency, we predict that they may have smaller thresholds than asymmetric shapes.

2.2 Experiment 1: symmetry to asymmetry

In the first experiment, we briefly explored the 3D percepts with symmetric and asymmetric SFM displays.

2.2.1 Method

2.2.1.1 Participants

Six Rutgers students with normal/corrected-to-normal vision participated in the experiments (three females and three males). They were naive to the experimental purpose, and were paid after the experiment for their participation.

2.2.1.2 Stimuli and design

The stimuli were 2D dynamic random dot displays shown within different shapes. We manipulated both the shape of contours and the speed profile of random dots.

2.2.1.2.1 Symmetry of the contour

We used two symmetrical contours and one asymmetrical contour. The contours were generated in the following way. First, we generated two pairs of vertical splines, one with three bumps, and the other with four bumps (See Figure 2.1). The two splines have the same height ($13.24°$, in dva, all the stimuli unit are in dva) (i.e., the distance connecting the top and bottom endpoints is the same), but the bumps have different amplitudes and periods. Then, we reflected and translated one of the two splines to get the other corresponding spline (Figure 2.1, middle and right); meanwhile, we use the two splines to get the asymmetric left and right spline (Figure 2.1, left). The average dva between the left and right spline is ($5.57°$). In the end, we connected the two endpoints at the top and at the bottom of each spline to form a closed contour.
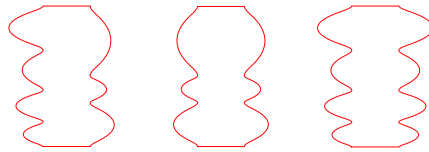
Figure 2.1: Contour shapes in Experiment 1

2.2.1.2.2 Speed profile of the dots

There were approximately 220 white dots of radius $0.03°$ (in dva) on a black screen. The dots were moving within the above shape. Therefore, there was no actual contour (or contrast border), only the illusory contour formed by the moving dots. The initial locations of the dots were randomly sampled. The motion was induced by changing the dots' location frame by

frame. The dots moved along the same axial direction. The speed of each dot **S** is defined according to the following 2D speed profile *S(r)*:

$$S(r) = \alpha C(r) + (1 - \alpha)L \qquad (2.1)$$

$$C(r) = 2\pi\omega R \times cos(0.5\pi\frac{r}{R}) \qquad (2.2)$$

In the second equation, let us call *C(r)* the cosine speed profile. **r** is the distance of a dot to the shape's central axis (see the red dotted line in Figure 2.2(b)); **R** is the radius of the horizontal cross-section. ω is the angular speed, we set it to $2.66°/s$. Therefore, when the dot is on the primary axis, its projected speed reaches the maximum, and when it reaches the boundary, the projected speed reduces to zero. In the experiment, the actual 2D speed of dot *S(r)* is a weighted sum of *C(r)* and a constant speed *L*, and α is the weight for *C(r)*. Note that based on the above

(a) A rotating ellipsoid
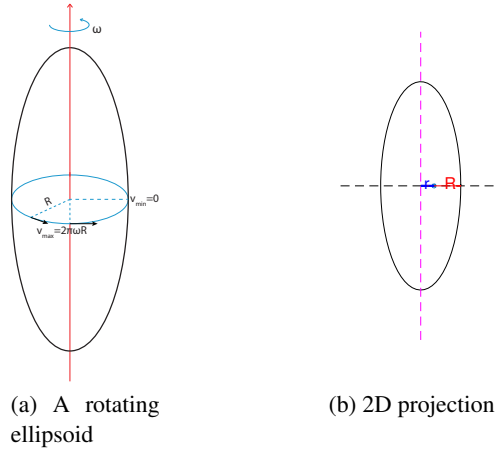
(b) 2D projection

Figure 2.2: 2D Speed profile. (a) An ellipsoid rotating with a constant angular velocity ω. R is the radius of the ellipsoid, *V*s are the projected linear speed under orthographic projection, *V* gets its maximum V_{max} when the dot moves to the center of the projective ellipse(see b), and it gets its minimum $V_{min} = 0$ when the dot moves to the boundary. (b) The corresponding ellipse of the ellipsoid in (a) under orthographic projection.

equation, when $\alpha = 1$, the dots at the boundary have a velocity of zero, and they will form a salient boundary which makes the display look unnatural. Therefore, in the actual displays, we set a small speed threshold so that when the speed calculated by equation (1) drops below that threshold, it will move with that minimum speed.

In our previous experiments (He, 2016), we have shown that observers' 3D percepts

systematically increase as α increases. Therefore, the threshold of α shows the point where the 3D percept begins to appear. A good analogy is light intensity and perceptual brightness.

2.2.1.2.3 Other controls

Uniform dot density in 3D projects to non-uniform dot density in 2D, which might provide extra cues to the 3D structure. To avoid this possible confounding factor, we kept the dots 2D density uniform using the method proposed by (Sperling et al., 1989). We divided the area within the shape into 10×10 (unit: pixel) small squares, and kept the dot density constant and uniform in each square by adding new dots on each frame, or deleting some dots as needed. Also, to eliminate possible motion direction biases, in half of the trials, the dots were moving towards left, and in the other half were moving towards right. We also counterbalanced the shape presentations by flipping the contour left/right and top/down.

In summary, there were mainly two manipulations: we had six different weights of cosine speed profile α, equally spaced from 0 to 1, and we had two types of contour: asymmetric and symmetric contours. We used the Method of Constant Stimuli to estimate the threshold of α in each type of contour.

2.2.1.3 Procedure

The experiment was run in a quiet dark room. Observers sat at a distance of 105 cm from an iMac screen (refresh rate: 85 Hz, resolution 1280×1024 pixels). The experiment was implemented and presented using Psychtoolbox-3 on a Mac version of MATLAB(R2014b) (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

In the beginning, the instructions were shown on the screen, observers were asked to judge whether the given stimuli looks like a rotating 3D object or not. After reading the instructions, the observer was shown two classical SFM demos showing clearly a flat surface and a vivid 3D rotating cyinder. Then observers completed eight practice trials to get acquainted with the procedure and the task. In the experimental trials, there was a 500 ms fixation at the center of the screen, followed by a 2s dynamic SFM display. Then the question appeared at the center of the screen:

Do you see a rotating 3D object? Press Left for Yes, and Right for No.

It would go to the next trial after recording observer's response. There was no response time limit.

There were 384 trials in total, half of them with symmetric shapes, the other half with asymmetric shapes. The whole experiment took about 30-35 minutes, and observers took a short break after every 50 trials. The order of trials was randomized for each level of symmetry and for each observer.

2.2.2 Results and discussion

Figure 2.3 shows each observer's proportion of 'yes' responses as a function of α, as well as psychometric fits (using logistic regression). The blue curve shows the psychometric functions for symmetric shapes, and the red ones are for asymmetric shapes.
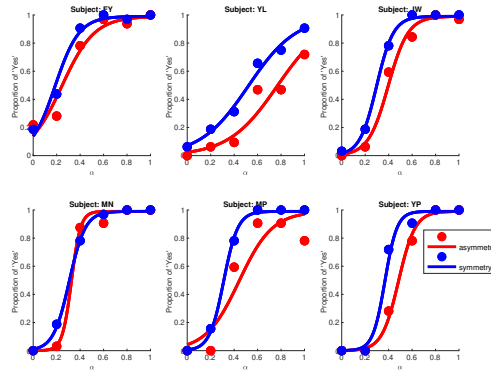
Figure 2.3: Psychometric function for each observer in Experiment 1

First, both symmetric and asymmetric shapes have nice logistic curves: the proportion of 'Yes' response increased as α increased, and plateaued at certain level of α. This is consistent with the intuition that, when α increases, the 2D display is more projectively consistent with a solid rotating object, and therefore the 3D percept should be stronger. However, the α thresholds were about 0.25 to 0.6 for different observers. The small thresholds also replicate our previous finding that the volumetric percept does not require projectively consistent speed profile.

Second, note that in our displays, there are two possible sources of inconsistency—the

speed profile of the dots and the shape of contour. When α is smaller than 1.0, or when the shape is asymmetric, the SFM display is projectively inconsistent with a rotating 3D object. This means that an asymmetric display contains more inconsistency than a symmetric display with the same α. Thus asymmetric shapes should have higher α threshold, and their curve should be to the right of the symmetric displays. Figure 2.3 confirms this trend in that the curve of asymmetric shape was to the right, which is to say, asymmetric shape had higher α thresholds (it was slightly harder to perceive 3D structure because it contained larger projective inconsistency).

However, though the trend was consistent across observers, the difference between symmetric and asymmetric shapes was surprisingly small. The average α threshold for symmetric shape was 0.39 (std = 0.14), while it was 0.48 (std = 0.22) for asymmetric shape. This is surprising because the violation of projective consistency with asymmetric shapes is so much stronger than symmetric shapes. If there is an asymmetric object rotating in 3D, its silhouette would grossly change over time. However, the silhouette in our display remained constant throughout the motion, so it is little likely to have an actual 3D object that could project to our asymmetric SFM displays.

Overall, the results of Experiment 1 indicate that human SFM 3D percepts are surprisingly resilient to deviations from projective consistency, both in terms of the 2D dot motion and the shape of occluding contour. Although the difference is in the predicted direction, the small magnitude of the difference between asymmetric and symmetric shapes is surprising, given that asymmetric shape had a much stronger level of projective inconsistency.

However, there might be several possibilities that need to be further discussed. For example, it is possible that the absolute difference between symmetric and asymmetric shapes was not large enough to be noticed. In our displays, the asymmetric shape has three bumps on one side and four bumps on the other side, however, it is possible that three vs four bumps is not a very noticable difference in SFM displays. Based on oral reports after the experiment, almost half observers did not even notice the difference between the two contours. Therefore, in the second experiment, we used one bump vs four bumps on two sides, to make asymmetry highly salient. Also, in the current experiment, we used only two types of shape (symmetric and asymmetric shapes), which might lead to another concern that the observers were likely to

remember the exact shapes. To avoid this possiblity, we added another three intermediate levels of symmetry between the two extreme cases in Experiment 2.

2.3 Experiment 2: levels of asymmetry

To further explore the role of symmetry in SFM, as well as to rule out the possibility that the insensitivity to asymmetry is due to the above two reasons, in the second experiment, we introduced larger difference between symmetry and asymmetry by having one bump vs four bumps on the left/right boundary (and again close the boundary with straight lines at the top/bottom). Also, we introduced more shapes along the spectrum from asymmetric shape to symmetric shape, to avoid the possibility that the observers would remember the exact shapes.

2.3.1 Method

2.3.1.1 Participants

Seven Rutgers graduate students (24 - 30 years old) participated in the experiment. Three were male, and four were female. All of them have normal or corrected-to-normal vision. The observers were paid after the experiment for their participation.

2.3.1.2 Stimuli and Design

The stimuli were still 2-second 2D dynamic random dot displays within a fixed shape of contour. The shape of contours would change from trial to trial.

The contours still consisted of left and right splines (height: $13.24°$, which were connected by two straight lines at the top and bottom. Instead of three bumps and four bumps, here we used one bump on one side, and four bumps on the other side (see the most left shape in Figure 2.4). The average distance between them was the same ($5.27°$). The most extreme asymmetric shape was composed from the one-bump and four-bump splines. The left and right contours of the symmetric shape, on the other hand, was composed from the four-bump spline(see the rightmost shape in Figure 2.4). Then we morphed the one-bump spline into the four-bump spline to create three more intermediate symmetry levels (see the other three shapes in Figure 2.4). In total, we have five levels of symmetry. We call the symmetry level (or the

morph weights) as β. β = 1 is symmetric shapes, and β= 0 corresponds to the extreme case of one bump vs. four bumps.
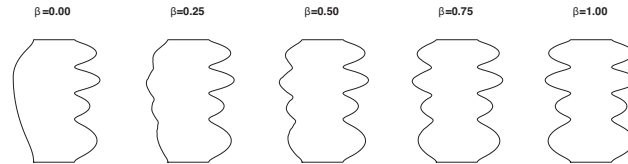
Figure 2.4: Contours of experiment 2. The β ratio is the proportion of one-bump curve in the morphed spline. 1 means symmetric contour, and 0 means the most asymmetric contour.

Then, we applied the same 2D speed profile to define the dot motion within each shape, and used α to manipulate the amount of projectively consistent motion (cosine speed profile $C(r)$) (see Stimuli and design in Experiment 1). But here, instead of Method of Constant Stimuli, we used Staircase (Psi Method) to find the α threshold for each degree of symmetry. The α threshold indicates the point where people start to have 3D percepts in an SFM display with a given shape. Smaller α threshold means that the display is easier to perceive as 3D. We used the thresholds from Experiment 1 to define the prior for the Staricase to determine the initial α parameters.

The other parmeters for the display, such as density control, minimum speed threshold, stimulus duration, remained the same as the first experiment.

2.3.1.3 Procedure

The experiment was conducted in a quiet dark room. The observers were sittting in front of an iMac screen (refresh rate: 85 Hz, resolution 1280×1024 pixels), at a distance of 105 cm.

The experiment was implemented and presented using Psychtoolbox-3 on a mac version of MATLAB(R2014b) (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

After signing the informed consent, the instructions were shown on the screen. Observers were instructed to respond based on their spontaneous percepts, and look at the display as a whole. After the observers finished reading the instructions, they pressed the space bar to enter the example phase. Two examples were shown with the same image contour (the right-most symmetric shape in Figure 2.4), but in one example with $\alpha = 1$, which is consistent with a rotating 3D object; in the other example with $\alpha = 0$, which corresponds to a flat translating surface. The experimenter would ask the observers to describe their percepts of those two examples, to make sure they could see 3D structure in the display with $\alpha = 1$. After the examples, there were eight practice trials to help observers better understand the task. The practice trials included the whole range of α (0, 0.2, 0.8, 1) and symmetry level β (0, 0.25, 0.5, 0.75, 1) with the same image contours used in the actual experiment.

In each trial, a fixation cross was shown in the center of the screen for 500 ms, followed by an SFM display for 2 s. Then the question would appear on the screen:

Do you see a rotating 3D object? Press Left Arrow for Yes, Right Arrow for No.

Observers needed to answer with left or right keys. If they pressed other keys, the following warning would be shown, and wait for observer's new input, until they pressed the Left or Right Key:

Invalid key. Please press Left Arrow for Yes, RightArrow for No.

No time limit was imposed. It would go to the next trial after observers' valid response (left/right). In the next trial, the α value would be chosen based on previous response and previous α (Prins et al., 2016).

We used the PsiMethod to calculate α threshold for each shape (i.e., each level of symmetry β). The process of PsiMethod for each shape would end when α converged to a stable value, or the maximum number of trials (50 trials for each shape) was reached. Most observers needed to finish the whole 50 trials for their trials to converge. Also, the order of different shapes was randomized for each new calculation of α. For counterbalancing purpose,

we randomized the motion direction (left/right) and shape handedness (original or left-right mirrored shape in Figure 2.4).

2.3.2 Results and discussion

We used Palamedes toolbox (Prins & Kingdom, 2018) to calculate the α threshold of each shape for each observer. The results are shown in Figure 2.5, and the last subplot shows the α threshold averaged across all observers.
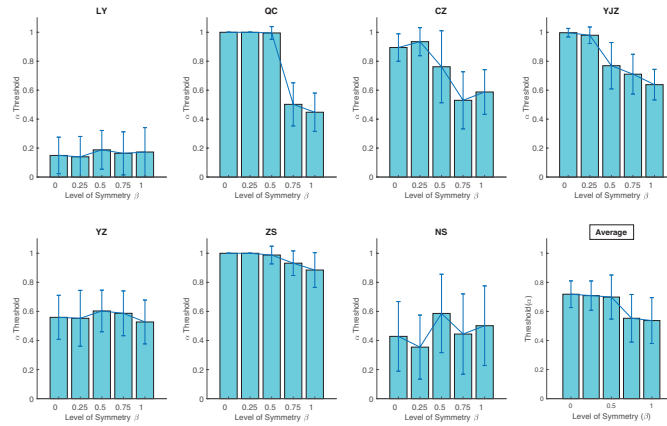
Figure 2.5: The α threshold of each level of symmetry β for each observer. Error bar shows the standard error of each corresponding α threshold. The last plot shows the average results from all observers.

For symmetric shapes, since a symmetric contour is consistent with a rotating object, the source of inconsistency in the dynamic display comes from the dot motion. As α gets larger, the proportion of projectively consistent motion also gets larger. The threshold of α shows the transition where people start to have vivid 3D rotating percept. Figure 2.5 shows that the average α threshold was 0.54 ($\sigma = 0.16$) for the symmetric shape ($\beta = 1$), which means that observers had vivid 3D percept when the display contained at least 54% projectively consistent motion. The finding is consistent with Experiment 1, as well as our previous studies. This supports our claim about projective consistency that human observers can have high tolerance of deviations from projective consistency in dot motion, and can have vivid 3D percepts when the proportion of inconsistent motion is around 40% - 50%.

For asymmetric shapes, there are two sources of inconsistency, one is from the dot

motion, the other one is from the asymmetry itself. Therefore, if the shape is less symmetric (smaller level of symmetry β), it contains a greater level of projective inconsistency from both sources, and therefore should require more projectively consistent dot motion to have vivid 3D percept. In other words, the displays with smaller β should have higher α threshold. The average plot in Figure 2.5 shows a clear but rather small decreasing trend from $\beta = 0$ (maximally asymmetric) to $\beta = 1$(symmetric). We also note that there are nonnegligible individual differences in the absolute value of the α threshold. One observer (ZS) had overall large α thresholds across all β, while another observer (LY) had overall small α thresholds. But if we concentrate on each observer's own trend from $\beta = 0$ to $\beta = 1$, four observers (QC, CZ, YJZ, ZS) had a clear decreasing trend, and the other three (LY, YZ, NS) showed little difference between different levels of β.

Overall, our results show that asymmetric shapes have slightly higher α thresholds than symmetric shapes. But as in Experiment 1, the difference between them are surprisingly small. This again shows that human 3D percepts are surprisingly resilient to deviations from projective consistency, both in terms of 2D speed profile and the occluding contours. The most suprising fact is that the projective inconsistency stemming from asymmetry was so obvious, and yet it seemed to have little influence on obsevers' 3D percept.

This result suggests that breaking global symmetry does not influence the vividness of observers' 3D percept. It is quite surprising, given that shape is an important factor in biological motion, and also given that symmetry is an important factor in shape skeletal representation (Blum, 1973; Siddiqi, Tresness, & Kimia, 1996). Therefore, it is worthwhile to delve deeper into different ways of breaking symmetry and its influence on SFM percepts. We know that symmetry can be defined in both global sense and local sense. In experiment 1 and 2, we broke bilateral symmetry in a global sense. It provides evidence that breaking global symmetry is not vital in SFM though it introduced stronger deviations from projective consistency. In the next experiment, we are going to explore one specific way of breaking global bilateral symmetry—parallelism, where the local symmetry is still preserved. Specifically, if symmetry is only preserved in a local sense, would it still generate vivid 3D percepts?

2.4 Experiment 3: parallelism to symmetry

In this experiment, we aim to further investigate the role of symmetry in SFM by using a specific form of asymmetry—parallelism. Global symmetry plays a more important role in SFM since a self-rotating symmetric object should have a symmetric 2D contour. But local symmetry is important in skeletal shape representation (Blum, 1973; Siddiqi et al., 1996). If the contour contains only local symmetry, would the 3D percept still be strong? Is there any difference in performance between globally symmetric and locally symmetric contours?

2.4.1 Method

2.4.1.1 Participants

Eight Rutgers graduate students (24 - 32 years old) participated in the experiment. Two were male, and six were female. All of them have normal or corrected-to-normal vision. The observers were paid after the experiment for their participation.

2.4.1.2 Stimuli and design

In this experiment, we continued to use similar 2-second display presentations as before. We applied the same α speed profile to random dots within a fixed shape. The shape of contour remained constant in one display, but varied across different conditions. The density of dots were uniform in 2D (on average 0.005 dots $/pixel^2$).

In this experiment, the shape of displays are morphed from parallel shapes to symmetric shapes (Figure 2.6). For the parallel shape, we first generated a sine curve (with a whole 2π cycle) and rotated it to vertical direction, which would be the axis for the final shape (the red dotted line in Figure 2.6). Then, we shifted the sine curve to the left and to the right through the same distance, and the two curves we got defined the left and right boundaries of the final shape. Finally, we connected the top and bottom endpoints of the two curves with two straight lines. We generated three sine curves with different initial phases (see the three shapes in the first row of Figure 2.6). The symmetric shapes (the last row of Figure 2.6) were obtained by shifting a phase shift of 180 ° to one of the two curves. The other intermediate shapes were
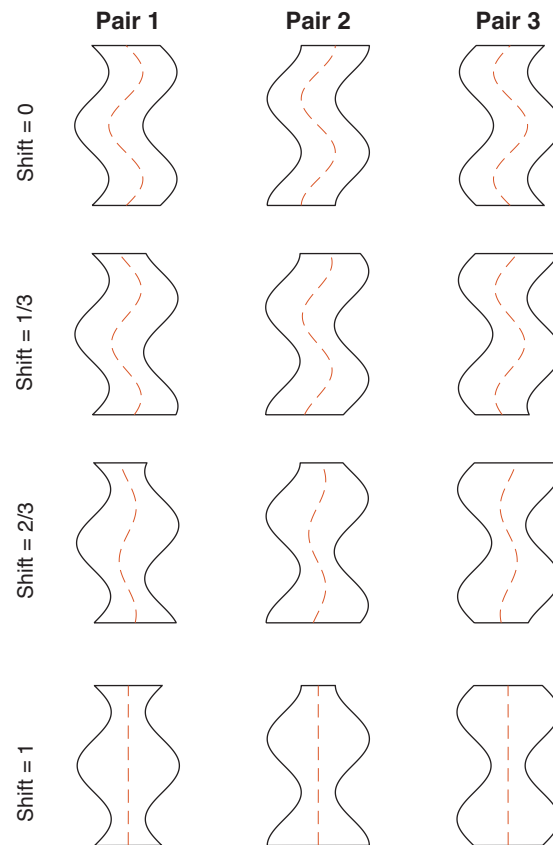
Figure 2.6: Contours (black lines) and axis (red dotted lines) of experiment 3. Shift = 0 means parallel shapes. Shift = 1 means symmetric shapes. For each level of shift, there were three different pairs of shapes.

generated by shifting the sine curve by either 60° or 120°. This resulted in a total of twelve different shapes: there were four shift levels, ranging from 0 (parallel) to 1 (symmetric). And for each level of shift, we had three different pairs of curves (with a different common phase). Each display was 2.80° (horizontal width), 7.5° (vertical height). Then we applied our α speed profile to each display. In particular, six levels of α were used: 0, 0.2, 0.4, 0.6, 0.8, 1. For each display, we asked observers the question whether they perceived the display as a solid rotating 3D object. We used Method of Constant Stimuli to find α threshold, which represent how easily the display can be perceived as 3D (given the inconsistency in the display). We had 720 trials in total, with 30 repetitions for each condition (each combination of α and shift ratio), including the three different sine curves and motion direction counterbalance(left/right).

In traditional SFM displays, the contours are relatively simple (less curvature, and curvature changes), so even if the contour shape is formed by the dot motion (i.e., the contour is 'illusory'), it is relatively easy to see the (motion-defined) contour shape. However, in our displays, the contour shapes are more complicated, and it is easier to mis-perceive a smooth maximum of curvature as a sharp corner (or vice versa). Therefore, it is possible that this may make the contour geometry more difficult to see. Since we are trying to explore the influence of contour shape on SFM, it is important to make sure both motion and contour cues are available in the display clearly and explicitly. Therefore, in the third experiment, we introduced a mild contrast difference between the in-shape region (RGB: [18, 18, 18]) and the black background (RGB: [0,0,0]), to reduce the ambiguity on the perceived contour shape.

2.4.1.3   Procedure

The experiment was conducted in the same quiet dark room. The observers were sittting in front of an iMac screen (refresh rate: 85 Hz, resolution 1280 × 1024 pixels), at a distance of 105 cm. The experiment was implemented and presented using Psychtoolbox-3 on a Mac version of MATLAB(R2014b) (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

The procedure was similar to Experiment 1 where we used Method of constant stimuli. After signing the informed consent, observers entered the screen of actual experiment. In the instructions, the experimenter would explain the display and task in general. The task is for each display, *do you see a single solid 3D object that is rotating?* The experimenter explained

the criteria in more detail, to make sure the observers understand that *a solid 3D object* means that the perceived 3D object should have a 3D volume, not just an open curved surface, and also that the object should not change its shape over time. After that, a rotating 3D cylinder and a translating rectangular 2D surface were shown on the screen, and the observers were asked to describe their percepts, to make sure they were able to see 3D structure in SFM displays. Then we selected eight displays with different α (0, 0.2, 0.8, 1) and all four shift ratios as practice trials. The results of practice trials were recorded but not analyzed.

The experimenter asked the observers if they had any questions, if not, they would enter the actual experiment. Each trial consisted of 0.5s fixation + 2s display + question. There was no time limit for each trial, and it would go to the next trial after a valid key is pressed (Left key for Yes, and Right key for No, and 'Q' key for escaping the whole experiment). There was no feedback in both practice and experimental trials.

2.4.2 Results and discussion
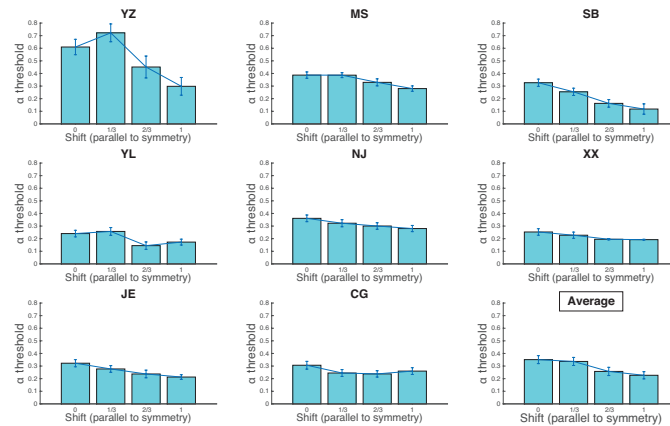
Figure 2.7: The α threshold of each level of shift from parallel to symmetry for each observer. Error bar shows the standard error of each corresponding α threshold. The last plot shows the average α threshold for all observers.

We recorded the Yes/No responses of the eight observers, and calculated their percent of 'Yes' responses. We used Palamedes toolbox to estimate the α threshold for each level of shift (parallellism to symmetry). The individual thresholds and the average threshold are shown in Figure 2.7. Each bar corresponds to the α threshold of one single shift condition.

Zero represents parallel shapes, and one represents symmetric shapes.

First, if we focus on the symmetric shapes, the α thresholds were within the range of 0.14 to 0.33 (mean: 0.22, std: 0.06). This is a little smaller than Experiment 1 where the α threshold of symmetric shape was around 0.21 to 0.5 (mean = 0.39, std = 0.14). Perhaps the smaller stimuli size in Experiment 3 resulted in smaller α thresholds, since for a smaller display, it is harder to distinguish between cosine speed profile and constant speed profile, and the 3D percept is stronger. This is consistent with Thompson et al. (1992)'s finding that narrower rectangle has stronger 3D percept (see discussion of the display in Introduction).

Second, Figure 2.7 shows some consistent patterns as our previous experiments. When we compare the α thresholds across different levels of shift, it was clear to see that most observers (MS, SB, NJ, XX, JE) had a clear decreasing pattern from parallel (locally symmetric) shapes to symmetric (globally symmetric) shapes. If we compare only the parallel shapes and symmetric shapes, YZ also had a decreasing trend. This means that symmetric shapes had overall smaller α threshold than parallel shapes. On the one hand, the decreasing trend is consistent with the fact that global symmetric shapes are more projectively consistent with rotating 3D objects compared to parallel shapes, and therefore they should have lower α thresholds. On the other hand, as in the previous two experiments, the influence of symmetry is surprisingly small—on average, the α thresholds were 0.35 (std: 0.13) for parallel shapes and 0.22 (std: 0.07) for symmetric shapes. This suggests that breaking global symmetry (but preserving local symmetry) has little influence on the 3D percept.

Overall, our results show that both parallel displays and symmetric displays had small α thresholds, and the parallel displays had a slightly higher α thresholds compared to symmetric displays. It again provides evidence that observers had surprisingly high tolerance to large deviations from projective consistency in terms of contour symmetry. The volumetric percepts are still vivid regardless of symmetry or asymmetry, and regardless of how symmetry is broken. It suggests that symmetry, which is a global property of contour shape, plays little role in SFM percepts.

2.5   General discussion

Most of the human SFM literature assumes projective consistency. One possible reason is that normally SFM displays are generated by projecting a rotating 3D object, so the 2D display is projectively consistent with the 3D object. However, there is some empirical evidence suggesting that the 3D percept from SFM displays does not require projectively consistent image motion (Ramachandran et al., 1988; Thompson et al., 1992; Froyen et al., 2013; He, 2016). Moreover, these studies suggest that, no matter where the inconsistency comes from, either from the projectively inconsistent speed profile (He, 2016), or an SFM display being shown through a shaped aperture (Ramachandran et al., 1988), or like in some studies, the inconsistency was introduced on purpose (Froyen et al., 2013; Tanrıkulu et al., 2016), when projective consistency is violated, there are conflicts between the image speed profile and the contour geometry, and the percept is dominated by the contour geometry. It suggests the potential importance of contour geometry in SFM. Then the questions on the table is: what shape properties would influence its dominance when it conflicts with the 2D motion?

When we analyze the displays where 2D contour dominates, such as Ramachandran's illusory rotating cone display (Ramachandran et al., 1988), and the rotating column displays from Froyen and Tankululu's figure-ground studies (Froyen et al., 2013; Tanrıkulu et al., 2016), we notice that the displays are always symmetric. Symmetry is quite common in daily life, and it is also very important in shape skeletal representations (Blum, 1973; Asada & Brady, 1986; Siddiqi et al., 1996). Therefore, in the current project, we tried to explore the influence of breaking symmetry in SFM displays. We were wondering if the 3D percept would be influenced by the loss of symmetry. In a series of experiments, we applied our α speed profile to different types of shapes. We controlled the degree of inconsistency from 2D motion by using different levels of α—higher α means greater projective consistency. We also controlled the degree of inconsistency from 2D contour shape by using different symmetric and asymmetric shapes. The 2D shape stayed fixed during the motion sequence.

In Experiment 1 and Experiment 2, we focused on the role of global symmetry. We defined the left and right parts of asymmetric contour using splines with different numbers of bumps, and then closed the contour by straight lines at the top and bottom. In Experiment 1,

we used three parts bumps vs. four parts bumps on the two sides. In Experiment 2, we used one vs four bumps to make the asymmetry more visually salient, and we also introduced three intermediate levels of asymmetric shapes. Using different psychophysical methods (Method of Constant Stimuli and Staircase), we obtained similar results that asymmetric shapes had slightly higher α thresholds compared with symmetric shapes. The difference is consistent with the fact that asymmetric displays contain more deviations from projective consistency. However, the small magnitude of this difference also suggests that observers had surprisingly high tolerance to various forms of asymmetry, despite that none of these asymmetric displays could possibly be projected from an actual 3D rotating object.

Given the results of Experiment 1 and 2, we wonder whether we can conclude that bilateral symmetry does not play a vital role in SFM. However, we realized that in these two experiments, symmetry/asymmetry was only defined in a global way. We know that bilateral symmetry can be defined locally or globally. A shape can be locally symmetric, but globally asymmetric. One such instance is bilateral parallel shape (i.e., the left and right contours are parallel to each other, see Figure 2.6). Local symmetry (parallelism) is an important concept in shape skeletal representations (Blum, 1973; Siddiqi et al., 1996). In Experiment 3 we investigated a different way of breaking global symmetry—one in which local symmetry is nevertheless preserved. We wondered whether the 3D percepts in displays containing only local symmetry would differ from general symmetric/asymmetric displays.

The results from Experiment 3 again showed that the 3D percepts from parallel displays does not differ much from globally asymmetric displays. Firstly, there was a clear increasing trend of α threshold from global symmetry to parallelism, but the difference was still relatively small. This pattern is similar with the pattern between globally symmetric and asymmetric shapes in Experiment 1 and 2. Secondly, the α thresholds of parallel displays were consistently but only slightly lower than other asymmetric shapes (in Experiment 1 and 2). It suggests that parallelism is not that different from general asymmetry in terms of SFM percepts.

All together, the three experiments indicates that breaking bilateral symmetry does have an influence in the predicted direction on the vividness of 3D percepts, however, the influence is small—the α thresholds between parallel, asymmetric and symmetric displays are quite similar. They provide strong evidence that projective consistency is not a necessary requirement

in SFM.

In the current study, we defined the speed profile of dots directly from 2D contour shape, and this is quite different from the traditional way that starts from a 3D object. One may argue that our displays were generated from manipulating the 2D speed profile directly, and the asymmetric displays do not have a simple 3D configuration of motion and structure, so they should always have 'No' responses. Our observers may just have been pushed to give a 'Yes' response by the relative comparison with other asymmetric displays, but overall the asymmetric displays (regardless of α) did not look like 3D rotating objects at all. We have many reasons to rule out this possibility. Firstly, for asymmetric displays, there is a clear increasing trend of 'Yes' responses as α increases, and the proportion of 'Yes' responses ranges from 0 to 1. These two aspects suggest that observers can distingush asymmetric displays with different α values, and that observers did have very strong 3D percept with large values of α. Secondly, the comparison of symmetric and asymmetric displays also provides more confidence to deal with this concern. The symmetric displays is physically consistent with some objects rotating in 3D when α is 1. Also its α threshold are comparable with our previous findings with a rotating ellipsoid (He, 2016). Therefore, the α threshold of symmetric displays can serve as a baseline for the asymmetric displays. Our data have shown a clear and reasonable trend that asymmetric displays always had a slightly higher α threshold than symmetric displays, and that parallel (locally symmetric, but globally asymmetric) displays had a slightly lower α threshold than other globally asymmetric displays. This is consistent with the amount of projective inconsistency each type of displays contains. Thirdly, based on the oral report after the experiments, some observers did not even notice that the asymmetric displays cannot be physically possible until the experimenter pointed out that the fixed asymmetric contour cannot be the sillouhette for a rotating object. Therefore, we believe that our observers' responses indeed reflected their actual 3D percepts.

Overall, our experiments have provided strong evidence for the surprising fact that projective consistency is not necessary in SFM. It is suprising not only in the sense of inconsistency in dot motion, but more importantly, in the sense of introducing asymmetry in the contour shape. Regardless of the concrete forms of asymmetry—in a gobal sense or a local sense, and regardless of the specific psychophysical methods we used—Method of Constant Stimuli or

Staircase, the 3D percepts are still vivid in spite of the fact that the asymmetric displays can not be projectively consistent with simple physicial rotation. One way to understand the results of the current study may be that the visual system does not impose a requirement of projective consistency with respect to global shape. Note that something like symmetry requires global (or at least non-local) comparisons. Hence whatever role contour shape is playing in SFM is not captured by global geometry. In the next study, will we examine more local aspects of contour geometry that are likely to influence 3D SFM percepts.

2.6   Conclusion

The current study used a novel way to explore the role of contour shape in SFM. It used displays in which not only the motion profile but also the occluding contour (in the asymmetric shapes) was projectively inconsistent with a 3D shape. Surprisingly, the effect of asymmetry was quite small, so the visual system appears to be quite tolerant to deviations from projective consistency, whether in terms of the image motion, or in terms of global geometry. Although the visual system does not seem to require projective consistency with global aspects of shape (such as bilateral symmetry), we do believe that contour geometry plays an important role in SFM, and we explore this further in the next study.

Chapter 3

The role of contour geometry in SFM

3.1 Introduction

3.1.1 The phenomenon of structure from motion

The term structure-from-motion (SFM) refers to the phenomenon where people can perceive 3D structure from a 2D dynamic display that contains only 2D motion information. Usually an SFM display is generated from a simulated rotating 3D object. The elements (dots) are usually projected from corresponding elements on the surface of a 3D object under orthographic or perspective projection. For example, in Figure 3.1, for a cylinder rotating with a constant angular speed, if you sample dots randomly on the surface of the cylinder, the dots will rotate rigidly with the cylinder, and then you project the dots onto a 2D image, it will create a typical SFM display. Note that there is no other information such as shading or color, and the only available infomation is the moving pattern of the elements.
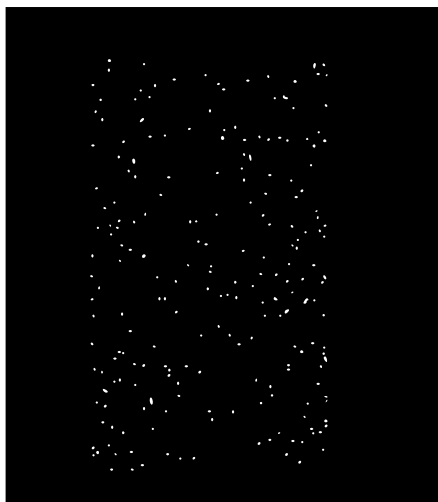
Figure 3.1: A typical SFM display of a rotating cylinder

SFM is a typical inverse problem in vision. The 2D display lacks one spatial dimension

information from the 3D scene and therefore there are many combinations of 3D structure and motion that are consistent with the SFM display. The task for our visual system is to infer the underlying 3D structure and 3D motion from an SFM display (what is usually called 'interpretation' or 'solution' in SFM literature). For example, in the above rotating cylinder example, one possible solution is that the dots are translating within the 2D rectangular plane with a cosine speed profile (that matches the projected speed of a rotating cylinder with constant angular speed). It is amazing that in most cases our visual system chooses a single solution for this challenging inference task, given that there are many possible combinations that are consistent with the display.

3.1.2 Empirical and theoretical studies on SFM

Researchers from both the field of human vision and the field of computer vision have long been interested in getting a unique solution for the inverse problem, and there are many empirical and theoretical studies on SFM in both fields.

The most widely known SFM models are the rigidity-based models which are based on Ullman's rigidity assumption (Ullman, 1979). The rigidity assumption states that if the display is consistent with an interpretation of a rigid body moving in place, this interpretation should be chosen by the visual system. Based on this assumption, the rigidity-based models check for rigidity among elements, and Ullman showed that four non-coplanar points in three distinct orthographic views are sufficient to mathematically recover the unique rigid solution. More-over, if global rigidity is not possible, the algorithms allow certain non-rigid transformation (such as articulation or bending of body limbs, and expanding, streching or flowing of inani-mate objects), and looked for the solution that maximizes local rigidity and then combine the local solutions into one global solution (Ullman, 1983). In summary, the rigidity-based models assume that the solution for SFM should be projectively consistent with a rigid rotating 3D object, and their algorithm uses rigidity to constrain the inter-element distances, and build-up the local solution into a unique global solution.

Rigidity-based models used the position of elements to find the correct solution. Other position-based models include trajectory-based models, which is based on the trajectory pattern of features (James T. Todd, 1982, 1984). These models were widely used in the fields of

both human vision (Hoffman & Bennett, 1986; Dosher et al., 1989; Sperling et al., 1990) and computer vision (Tomasi & Kanade, 1992; Morita & Kanade, 1997; Anandan & Irani, 2002).

Unlike the position-based models which track positions of individual dots, the other type of SFM models is based on the analysis of the entire motion pattern (i.e., velocity field or optic flow).

There is plenty of research on the effect of motion in SFM, which explains why the velocity field is the key to solve SFM and explores how it is used in SFM. There are many studies suggesting that SFM calculation is based on the velocity field. First, when it is not possible to track individual dots, such as when there are many dots in the display and the dots correspondence is not clear (Landy et al., 1991), observers are still able to perceive the underlying 3D structure. It implies that velocity field, instead of positions of individual dots, is more robust for the visual system. Second, when the dots have limited lifetime (Sperling et al., 1990), observers are still able to solve SFM problem. Treue et al. (1991) found that the minimum dots lifetime required to perceive the 3D structure is around 50-85 ms, which is similar to the threshold for estimating 2D velocity. This also implies that the visual system uses velocity to process SFM. In addition, there is also biological evidence that motion processing is important in SFM. Siegel and Andersen (1990) monkey studies showed that areas MT and MST were involved in SFM. The fMRI study by Paradis et al. (2000) also showed that the superior occipital gyrus (SOG; presumptive V3/V3A) and the left occipito-temporal junction had specific sensitivity to the SFM displays depicting curved 3D surfaces, compared to 2D coherent motion and random motion. Since these areas are specific to motion perception, they also imply that SFM involves the analysis of the velocity field.

Optic flow field (the projection of 3D motion field) has different orders of information, such as first-order optic flow (velocity), and second-order optic flow (acceleration). A series of studies have shown that people mainly use first-order properties of optic flow in SFM (Koenderink & Van Doorn, 1986; James T Todd & Bressan, 1990; Landy et al., 1991), and they are sensitive only to the sign of accelerations (Norman & Todd, 1993). They have shown mathematically that first-order optic flow can only solve the SFM problem up to an affine solution that contains an infinite number of rigid interpretations that fall into an affine family (James T Todd & Bressan, 1990). James T Todd and Bressan (1990) also showed that when more frames

were added to the display so that theorectically the second-order optic flow information (acceleration) was available, the performance of human observers did not have much improvement, which further supports the hypothesis that people do not use higher-order properties of optic flow with SFM displays.

If the visual system is only sensitive to first-order property of the optic flow, i.e., velocity, how can it estimate 3D structure from the velocity field?

Braunstein and Andersen (1984) ran a series of experiments and proposed the constant angular velocity heuristic. Instead of testing dots' consistency directly with a rigid interpretation, they assumed that the visual system analyzed the projected image motion (i.e. 2D speed profile) of the dots to test for consistency with 3D rotation. They found that if the dots' 2D speed profile was a cosine function along each horizontal cross-section (assuming that the object is centered at the origin of the frame of reference), observers had high ratings for both depth and volumetric percepts. Therefore, they proposed that the visual system is biased to assume that 3D objects tend to rotate at constant angular speed.

There are also other velocity-based models that use the velocity field as input to re-compute the 3D structure from SFM displays. For example, the motion-perspective model use the fact that velocity is inversely proportional to the depth between object and observer under perspective projection (Sperling et al., 1989), and therefore observers can assign relative depth according to the minima and maxima of velocities. However, this method only works under pespective projection.

Another strategy for using dot's velocity is to analyze the optic flow pattern in the image by decomposing it into three differential invariants–*curl*, *shear* (*def*) and *div* (divergence) (Koenderink & Van Doorn, 1986) (Figure 3.2, from Domini and Caudek (2003b). *curl* refers to the rigid image rotation, *div* is the isotropic expansion/contraction, and *shear* is the amount of shape change over time. *curl* and *div* are also influenced by other factors (Todd, 1995), but *def* is only influenced by the object's 3D structure. In other words, *def* is very informative about the local structure such as relative slant. Domini and Caudek (1999) have proved that the component *def* is related to the surface slant and rotating speed.

In a series of empirical and computational studies, Domini and Caudek (2003a) have shown that for a planar patch rotating rigidly in depth along a vertical axis, the local velocity
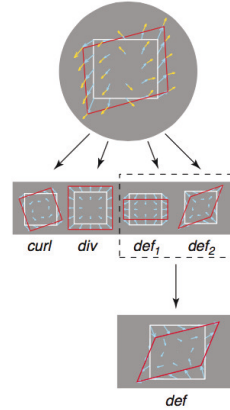
Figure 3.2: Decomposition of first-order optic flow into *curl*, *div* and *def* (from Domini and Caudek (2003a), Figure 1c)

field can be regarded as a linear field, and the deformation $def = \sigma\omega$, where σ is the local surface slant, and ω is the rotating angular speed (Domini et al., 1997; Domini & Caudek, 1998, 1999; Domini et al., 2002; Domini & Caudek, 2003b; Di Luca et al., 2004). The equation indicates that there are infinite combinations of σ and ω that can generate the same value of *def* (those pairs lie on the hyperbola defined by the above equation). They proposed a heuristic model that assumes Gaussian noise for the measurement of *def* and human observers choose the pair of (σ, ω) that maximizes the likelihood of *def*.

In summary, SFM studies use either the locations of individual dots or properties of the motion field to compute a 3D structure from image motion. Note that there are two common assumptions with those SFM studies. The first assumption is that they assume that the visual system would always choose a 3D interpretation that is projectively consistent with the SFM display. The position-based models are based on the idea that the change of dot positions should be consistent with the mapping from 3D objects to their 2D projections under the assumption of a rigid object. Velocity-based models, on the other hand, are based on the idea that the 2D motion pattern should be consistent with its 3D motion. For example, the constant angular velocity heuristic assumes that the 2D speed of dots should follow a cosine function which is projectively consistent with 3D rotation with constant angular velocity.

The second common assumption is that SFM is only based on the analysis is motion information in the display, and there is little to no room for the role of contour shape. The

position-based studies try to get the 3D structure by using some constraints on the position of individual elements. For example, the rigidity assumption uses the constraint that the pairwise distances between individual elements should remain constant throughout rotation. They build the global structure from local estimates using those constraints, and treat all elements equally within the display, mentioning no role of contour geometry. The velocity-based studies try to estimate the 3D object by analyzing the whole velocity field. For example, Braunstein's constant angular velocity heuristic is based on the fact that the projected speed of a rotating 3D object (with constant speed) should have a cosine speed profile along each slice in the direction of motion, and it did not include any contribution of the specific shape of contour. The studies based on decomposition of first-order optic flow (Koenderink & Van Doorn, 1986; Norman & Todd, 1995; Domini et al., 2002; Domini & Caudek, 2003a; Di Luca et al., 2004), build up the global structure from local analysis of optic flow deformation. They recover the local slant and angular speed from the deformation of optic flow, and gradually integrate those local measurements into one single global solution. Again, in both local measurement and global integration, no role is played by the geometry of the occluding contour.

However, there is some evidence suggesting that neither of the two assumptions may be valid. In the next section, we will present this evidence and also show how these two assumptions relate to each other.

3.1.3 Reconsidering the prevalent assumptions about projective consistency and the role of contour geometry in SFM

In the literature, since SFM displays are mostly generated from projections of rotating 3D objects, it has been taken for granted that the perceived 3D structure should always be the 3D object that generated the 2D display. It has also been the basic goal for different models: solving an SFM problem is equal to finding the 3D object that generated the 2D display, and in other words, we should use this assumption to find the projectively consistent solution.

However, there is some evidence showing that the perceived 3D structure is not necessarily to be projectively consistent with the 2D display.

One evidence comes from Thompson et al.'s empirical study (Thompson et al., 1992). They showed a display where the random-dot texture within a central rectangular region was

moving with a constant speed, whereas the random-dot texture in the surrounding area was moving at constant speed in opposite direction. They found that when the central rectangle was narrow enough, their observers reported a rotating cylinder percept. Here, the rotating cylinder percept is not projectively consistent with the constant speed.

Another piece of evidence comes from some figure-ground studies (Froyen et al., 2013; Tanrıkulu et al., 2016). They used displays of alternating light and dark vertical regions. There were random-dot texture moving horizontally at a constant speed but in opposite directions in these alternating regions. The left/right contour of the regions were either symmetric/asymmetric or convex/concave, i.e., if a region was symmetric/convex, its two neighbouring regions were asymmetric/concave. Interestingly, their observers tended to perceive the convex regions as rotating columns, despite the fact that the random-dot texture was moving at a constant speed. In this example, again the constant 2D speed induced a rotating column percept which should have a 2D cosine speed profile.

The above studies have shown that the 3D percepts induced by 2D displays did not necessarily need to be projectively consistent with the display, therefore, projective consistency may not be a necessary constraint in SFM. But note that none of them have explored the projective consistency assumption directly and systematically. Besides, one may argue that in the figure-ground studies, the competition along the boundaries between figure and ground regions could have contributed to make the rotation percept more vivid.

More direct evidence comes from our previous studies. In my master thesis (He, 2016), and also in the study in çhapter 2, unlike traditional SFM studies that generated the displays from projection of a 3D object, we manipulated the 2D speed profile and contour geometry in the displays directly. This gives us more freedom to manipulate the degree of projective consistency. Assume there is an ellipsoid rotating along its primary axis with a constant angular speed, and there are some dots attached to its surface. When the dots are projected onto a 2D plane (under orthographic projection), the projected dots should move horizontally, and their speed should follow a cosine function, i.e., the speed goes down to zero at the left and right boundaries, and reaches its maximum along the central axis of the projected ellipse. We called this the cosine speed profile. In our study, we started with a 2D ellipse-like region, and manipulated the speed profile directly. The 2D speed profile of dots was a linear combination

of the cosine speed profile and a constant speed profile, and we manipulated the weight of the cosine speed profile (what we called α). We introduced different levels of α and also manipulated the motion direction. We found that the volumetric percepts were still strong when the dot speed included as high as 40% linear motion, and even when the motion direction deviated from the horizontal by $20° - 30°$. Later on, we applied the 2D speed profile to more complex shapes (the study in chapter 2), such as asymmetric and parrallel shapes, and we found similar tolerance in the projective inconsistency in contour geometry. These results indicate that our visual system has a large tolerance for deviations from projective consistency, and the perceived structure is not necessarily projectively consistent with the 2D display.

In summary, the above studies have shown that under certain conditions projective consistency is not a necessary assumption in SFM. But why has it been taken for granted for such a long time? This is probably due to the way how SFM displays are generally generated—it is generally generated by projecting a 3D rotating object. Therefore, the display is by construction projectively consistent with the 3D object. But the above studies have shown that under certain conditions the 3D percept was not projectively consistent with the 2D display. This raises the question: why?

We know that in SFM displays, there are actually two sources of information—motion and contour shape. But the traditional models only focus on the role of speed profile, and very few of them have mentioned the role of contour geometry. Even in the literature on surface interpolation—the step to form a smooth surface after obtaining the 3D positions of individual elments, the interpolation models only consider a surface passing through all dots, ignoring the constraints brought by the 2D image contour (Hildreth, Ando, Andersen, & Treue, 1995).

The role of contour shape has been largely ignored partially because in traditional SFM displays, the percepts induced by 2D motion and by 2D contour are consistent with each other, so the effect of contour geometry is covered by 2D motion. However, it becomes more complicated and interesting when we consider the problem of aperture and occlusion. The simplest example is that when a portion of an SFM display is shown through a shaped aperture (and there are no cues to the presence of an aperture), can people still recognize the true 3D structure? Occlusion is very common in daily life, so we would expect that if the aperture is large and most of the object is still visible, we should still be able to perceive the correct

3D object. However, one interesting demo shown by Ramachandran suggests that it may not be that simple (Ramachandran et al., 1988). He reported an interesting demo that a rotating cylinder was occluded by a triangular aperture (which had the same base and height as the 2D projection of the cylinder), and the observer did not know of the presence of the aperture (there were no depth or occlusion cue). It is interesting that his observers perceived the display as a rotating cone, despite that the display was projected from a rotating cylinder. In this case, the percept induced by the contour geometry conflicts with the percept that would otherwise be induced by 2D motion, and the actual percept was dominated by the contour.

Ramachandran's demonstration suggests a potentially important role of contour geometry in SFM. Similar evidence was also reported in the figure-ground studies mentioned above (Froyen et al., 2013), where the convex regions are more likely to be seen as rotating, suggesting that different properties of the contour shape influence its dominance in SFM, and convex contour has a stronger influence on the 3D percepts.

Our previous work also provides more direct evidence for the role of contour geometry. In studies using an ellipse (He, 2016), we introduced inconsistency in motion by varying the weight of cosine speed profile. Our observers showed large tolerance to deviations from projectively consistent motion for the shape of ellipse. This may be due to the fact that the smooth curve of the ellipse was a very strong cue of being an ellipsoid, therefore the deviations from projective consistency in motion were ignored.

Overall, the above studies question the role of projective consistency and suggest the importance of contour geometry in SFM. Interestingly, the two topics seem to always occur together, i.e., contour geometry seems to play a dominant role whenenver there is projective inconsistency in the display. It suggests possible connection between the two topics. Therefore, we hypothesize that the influence of contour geometry may in part explain why SFM percepts often projectively inconsistent with the image speed profile. In particular, the presence of local boundary is a strong cue to be the 3D object's occluding contour (where the surface slant is $90°$), so when the motion and contour conflicts with each other, the influence of contour geometry can often be strong enough to overcome the inconsistency in motion, and dominate the 3D percept.

3.1.4 Current approach

In the previous study, we explored the role of projective consistency by applying the cosine speed profile to asymmetric contours. This method cannot provide clean evidence for the role of contour geometry because the speed profile was defined in terms of the 2D contour. Locally speaking(i.e. along each horizontal cross-section), the motion is consistent with a rotating 3D object, and it leads to possible correlation between the effect of motion and the effect of contour. To study the role of contour geometry, we need to keep the dot motion independent from the contour geometry. Inspired by Ramachandran's demo (Ramachandran et al., 1988), in the current study, we separated the dot motion and contour shape completely. The motion signal is always consistent with a rotating 3D object, and we manipulate the 'aperture' through which the display is seen. Therefore in most cases, the motion signal and the shape signal conflict with each other. For example, for a rotating cylinder seen through a triangular aperture (only part of the cylinder is visible), the two most likely 3D interpretations of such displays would be either dominated by the motion (cylinder) or by the contour (cone). For simplicity, let us call the two possible structures and interpretations motion-defined structure/ percept and contour-defined structure/ percept.

We manipulated different properties of the contour shape quantatively, so that as we increase the conflict between motion and contour geometry, at some point the contour will no longer dominate the 3D percept, and we find the transition point/threshold using Method of Constant Stimuli. In particular, in the first experiment, we manipulated the horizontal width ratio (what we call ω) between the contour shape (we will call it aperture for simplicity) and the motion region, and estimated the threshold of ω which represented the transition from motion-defined percept to contour-defined percept. Our ω manipulation is analogous to our previous α manipulation in that the speed at the left/right boundaries would not decrease to zero. The results of ω thresholds in experiment 1 inform us about the range of ω that has strong contour-defined percept. In other words, the contour-defined percept can be strong when the horizontal width ratio between the aperture and the motion region is above certain value. It enables us to have more complex manipulations of contour geometry in subsequent experiments. In the second and third experiments, we separated the axis of shape away from the axis of motion either

by parallel translation or rotation. Note that in the above three manipulations, the horizontal width ratio ω is constant at different heights, which means that no matter where the observer fixates, the propotion of the cosine speed profile that is visible to the observers would be the same. For example, in Figure 3.4, for the ellipse, ω is constant across different heights. In the first three experiments we used the same shape family for the motion and aperture area (i.e., if the motion region is a triangle, so is the aperture region), so that the horizontal width ratio ω remained constant within a display. In the fourth experiment, we explored the influence of varying ω along vertical direction on the contour-defined percept. Meanwhile, we used several basic shape classes chosen from rectangle, trapezoid, diamond, hexagon, barrel, ellipse. Those shapes vary two basic features: smoothness and variation of horizontal width, so we can explore the effects of these two properties on the contour-defined percepts. In the last experiment, we also manipulated the effect of convexity on the contour-defined percepts.

3.2 Experiment 1: constant width ratio between aperture and motion

In traditional SFM displays, the whole display is visible to the observer, what if only part of it is visible? The simplest way is to 'shrink' the horizontal width of the visible region (i.e. aperture), so that only certain proportion of the display is visible. We hypothesize that the shape of contour would dominate the 3D percepts despite the inconsistency with motion, until the proportion of the horizontal width drops below some threshold. In this experiment, we will look for this transition points for different basic shape classes (rectangle, trapezoid, diamond, ellipse).

3.2.1 Method

3.2.1.1 Participants

Six Rutgers graduate students with normal/corrected-to-normal vision participated in the experiments. Three were female and three were male. They were all naive to the purpose of the experiment. All observers were paid after the experiment for their participation.

3.2.1.2 Stimuli and design

The stimuli were 2D dynamic random-dot displays occluded by different shapes of aperture, so that only part of the motion region is visible. There was no relative depth cue between the aperture and motion, so the aperture shape was seen as the bounding contour of the SFM display. We manipulated the horizontal width ratio between the aperture and the motion region, as well as the shape of aperture. The speed profile of dots was determined by the same speed profile as in our previous studies (He, 2016), also see details in the section of *Speed profile of the dots*).

3.2.1.2.1 Horizontal width ratio ω

Our main manipulation is horizontal width ratio ω between the aperture and motion region (see Figure 3.3). In Figure 3.3, the green shapes are regions that define the dot motion (i.e. the cosine speed profile). In particular, the dots are moving horizontally across the green axis, and the speed drops to zero at the boundaries. While the black shapes defined the regions of aperture. In other words, only the parts within black shapes are visible to the observers. Note that the motion region and aperture region belong to the same shape class. For each shape, if you cut a horizontal slice, ω is defined as $\frac{width\ of\ black\ region\ (aperture)}{width\ of\ green\ region\ (motion)}$. Therefore, at one extreme, when ω is 1, the two regions coincide with each other, and the whole display is consistent with a rotating 3D object. At the other extreme, when ω is 0, it means that the green region has infinite width, and the dots within the black region are moving with a constant speed. Thus, as ω decreases, there is a greater degree of conflict between the speed profile and contour geometry. In the current experiment, we defined six levels of ω, equally spaced from 0 to 1. In fact, there are two ways to manipulate ω: in one method, we can keep the green regions constant across all ω, and change the width of black regions; in the other method, we can keep the black regions constant, and vary the green regions. One major issue of the first method is that observers could figure out the correlation between the width of SFM display and our ω manipulation, and make judgements not based on the 3D percepts, but on the display width. Therefore, in the actual experiment, we used the second method that keeps the size of visible displays constant (4.7420e+04 $pixel^2$, and the dva was around 4.4°) (all the stimuli size

reported below are n dva), and vary the width of the green regions. The horizontal width of green regions should increase as ω gets smaller. For $\omega = 0$, we simply generated random dots moving with a constant speed within the black region, since theoretically speaking the green region should have infinite width.
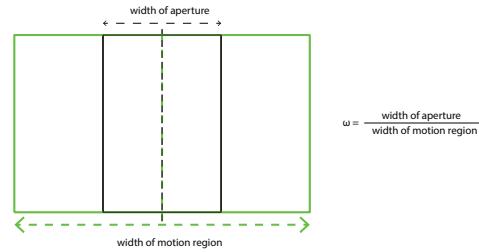
Figure 3.3: Definition of horizontal width ratio ω. The ratio is the horizontal width of aperture (black region) over horizontal width of motion region (green region).
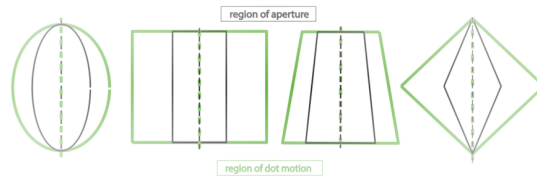
Figure 3.4: Contour shapes in Experiment 1. The green shapes define the motion region, and the black shapes define the aperture.

3.2.1.2.2 Shape of aperture

We used four different types of shape: rectangle, trapezoid, diamond and ellipse (see Figure 3.4). The rectangle is the most basic shape in our displays, since it is what was used in older SFM studies, and it does not have smooth curvature (it is a polygonal shape), and the width does not vary with height. Compared to the rectangle, the trapezoid adds some variation in horizontal width, and the diamond adds even more variation. Moreover, the diamond has two apices, so that the speed will decrease to zero at the top and bottom. The ellipse is similar to the diamond in that the speed decreases to zero at the two apices in vertical direction, but more importantly, it has smooth curvature, which makes it unique compared to the other shapes. All in all, the four shapes are simple enough, but cover basic properties such as smoothness, and variations of horizontal widths.

3.2.1.2.3 Speed profile of the dots

There were approximately 150 white dots with radius of 0.03 ° on the black screen. The dots were moving within the green regions. The initial locations of the dots were randomly sampled. The motion was induced by changing the dots' locations frame by frame. The dots moved along horizontal direction (left/right). The speed of each dot **S** is defined according to the following 2D speed profile **C(r)**:

$$C(r) = 2\pi\omega R \times cos(0.5\pi\frac{r}{R}) \tag{3.1}$$

In the above equation, let us call $C(r)$ the cosine speed profile. **r** is the distance of a dot to the shape's primary axis (see the red dotted line in Figure 3.5(b)); **R** is the radius of the horizontal cross-section. ω is the angular speed,Therefore, when the dot is on the shape's primary axis, its projected speed reaches the maximum, and when it reaches the boundary of the green region, the projected speed will reduce to zero.

(a)   (b)
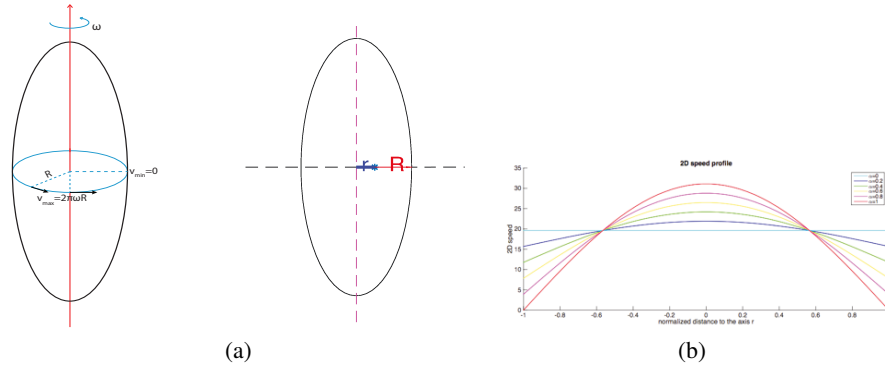
Figure 3.5: Speed profile. (a) An ellipsoid rotating with a constant angular velocity ω and its orthographic projection. R is the radius of the ellipsoid, Vs are the projected linear speed under orthographic projection, V gets its maximum V_{max} when the dot moves to the center of the projective ellipse(see b), and it gets its minimum $V_{min} = 0$ when the dot moves to the boundary. (b) The speed of dots as a function of horizontal location, 0 is at the center, 1 is at the boundary. Different lines represent different weights of cosine speed profile, i.e., α

Note that based on the above equation, the dots at the boundary have velocity of zero, and they will form an illusory but salient boundary which may introduce some artifacts. Therefore, in the actual displays, we set a very small speed threshold so that when the speed calculated by equation (1) drops below the threshold, it will move with that minimum speed.

Using this speed profile, we generated seven different displays for each condition. In

each trial, one of the seven displays was chosen randomly, and was shown through an aperture whose width was determined by the horizontal width ratio ω.

3.2.1.2.4   Other controls

In-shape contrast: since we are exploring the role of contour in SFM, we should make sure the contour cue is visible in the display. Just as our previous symmetry study, we introduced a small contrast within the aperture, i.e. the region inside the aperture was dark grey (RGB: [18, 18, 18]).

Uniform density in 2D: Sperling et al. (1989) has pointed out that uniform dot density on a 3D surface does not produce uniform dot density in 2D. The projected image has systematically varying density, which will introduce additional cues to the 3D structure. To avoid this possible confounding factor, we kept the dots 2D density uniform using the method proposed by Sperling et al. (1989). We devided the area within the shape into 10 *pixel* $\times$ 10 *pixel* small squares, and pseduo-randomly sampled dots within the whole shape, so the number of dots in each subsquare is the same. Note that dots in different locations have different speeds, which will break the uniform density over frames. We added new dots or deleted old ones at randomly chosen locations to keep the density constant across all frames. Overall, there were only around 5 such dots that appear/disappear randomly at each frame. Compared to the 150 dots, the proportion was fairly small.

Motion direction counterbalance: Besides, to eliminate possible motion direction biases, in half of the trials, the dots were moving towards left, and in the other half were moving towards right.

In summary, there were two manipulations: we had four different shapes (rectangle, trapezoid, diamond, ellipse), and we had six different width ratio ω (0, 0.2, 0.4, 0.6, 0.8, 1). For each condition, we had 30 repetitions (including motion direction counterbalance). We used Method of Constant Stimuli to estimate the threshold of ω for each type of contour shape.

3.2.1.3   Procedure

The experiment was run in a quiet dark room. Observers sat at a distance of 105 cm from an iMac screen (refresh rate: 85 Hz, resolution 1280 $\times$ 1024 pixels). The experiment was

implemented and presented using Psychtoolbox-3 on MATLAB (R2017b) MAC (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007).

In the beginning, the instruction was shown on the screen. In this experiment, the Yes/No task is a conjunction task which has two criteria. The 'yes' response needs to meet the following two criteria at the same time. First, the display should look like a rotating 3D object. In other words, it cannot be an open curved surface and it should have a closed volume in depth. Second, the perceived 3D structure should be consistent with the 2D contour. For example, if you perceive the display as a rotating cylinder, but the 2D contour is a triangle, it does not satisfy the second criterion.

After showing the instruction, the experimenter emphasized the two criteria, and then showed three examples on the screen: a flat surface, a rotating cylinder, and a rotating cube shown through a circular window, and the circle is inscribed in the square (the projection of the cube). The experimenter made sure observers could dinstingush between the flat surface and the rotating cylinder. Also, the experimenter asked observers to describe the third example, and most of them were able to tell that the 'sphere' looked 'unnatural' at the boundary. If the observers could not describe the example, the experimenter would show them two demos side by side—a real rotating sphere and the third example 'cube through circle'. She made sure the observers could understand the two criteria and could tell the difference between the two demos.

The program moved to the practice trials after the experimenter made sure that the observers understood the conjunction task even if they still perceived the third example as a rotating sphere. We think it was fine to still have the wrong percepts because in Ramachandran's demo (cylinder through triangle), their observers were also quite confident about their percepts of a rotating cone. In the practice trials, we generated nine different displays, inspired by Ramachandran's demo. The shapes that define motion and the shapes that define aperture are randomly chosen from an equilateral triangle, a square and a circle (see Figure 3.6). Therefore, the three displays lying on the diagnoal should have 'yes' responses since the 3D rotating objects were consistent with the 2D contour (i.e., sphere through circle, cone through triangle, cube through square), and for the rest of the displays, the apertures (black contours) were

inscribed to the contour that defined motion (green contours), and they should have 'No' responses. The order of the practice trials was randomized. In each trial, after a observer made his/her reponse, the correct answer, as well as the observer's response, would be shown on the screen. With the help of the three examples and the practice trials, all observers understood the two criteria.
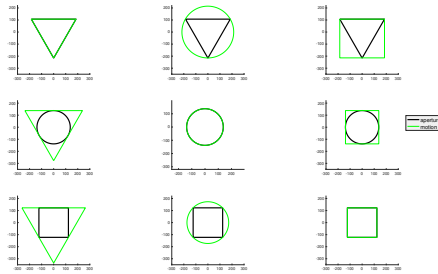
Figure 3.6: Contours of motion and aperture used in the practice trials. The green shapes define the motion region, and the black shapes define the aperture.

After the practice trials came the real experimental trials. In each trial, after a 0.5s fixation, an SFM display generated using the above manipulations was shown at the center of the screen for 3s (which is 1s longer compared to previous studies, since the task is more complicated), followed by the question with two criteria below:

A solid rotating object?

AND its 3D shape is defined by the 2D contour?

Press Left for Yes, and Right for No.

If the observers pressed any key other than Left arrrow or Right arrow, a message would show on the screen, until they made valid responses:

Invalid key. Please press "LeftArrow" for Yes, "RightArrow" for No.

There was no time limit, and it would go to the next trial after recording observer's valid response (Left/Right).

We ran 30 trials for each combination of shape and ω, yielding 720 trials in total. The order of trials were randomized, and we split the experiment into 5 blocks. The whole experiment took about 50 minutes, and observers could take a short break either between blocks or in the middle of each block (every 72 trials).

3.2.2   Results and discussion

We recorded the Yes/No responses, and then we used psignifit toolbox by Schütt, Harmeling, Macke, and Wichmann (2016) (https://github.com/wichmann-lab/psignifit) to fit the psychometric curve and also calculated the ω threshold for each shape. The psychometric curves for all the four shapes are shown in Figure 3.7, and the estimated ω thresholds as a function of shape are shown in the bar plot in Figure 3.8. The dots in Figure 3.7 indicate the proportion of 'yes' responses in each condition (each ω and shape), the vertical thinner lines show where the estimated ω thresholds were for each shape, which also corresponds to the heights of bars in Figure 3.8. The horizontal line segments indicate the 95% confidence interval for each estimation, which correspond to the error bars in Figure 3.8.
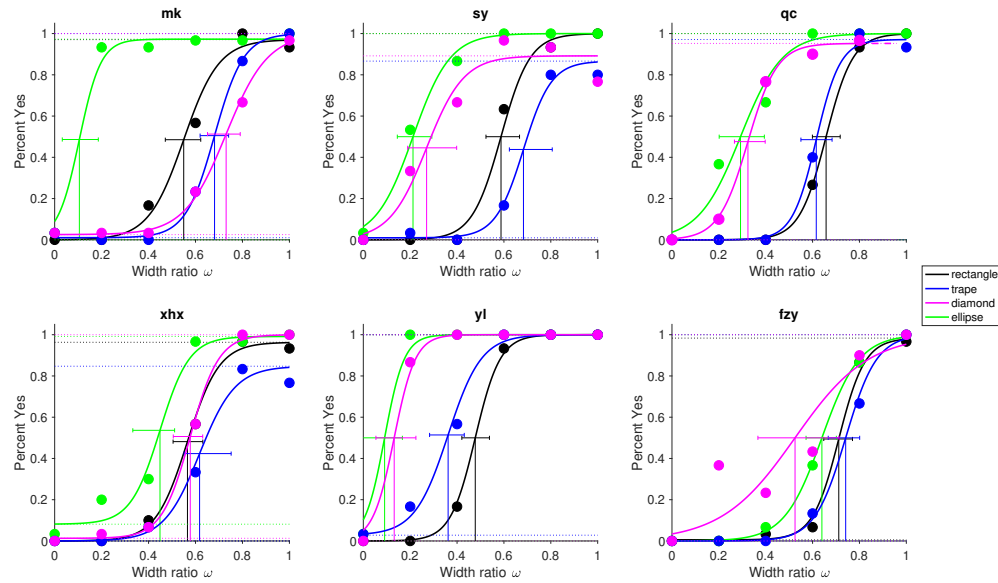
Figure 3.7: Psychometric curve for each observer in Experiment 1: constant width ratio

The logistic curves in Figure 3.7 show that for most shapes, the proportion of 'yes' responses increased as the horizontal width ratio ω increased, and plateaued before ω reached 1. The pattern conveyed a few pieces of important information.

First, the positive correlation between ω and the 'yes' response is consistent with the fact that the displays with larger ω contain less conflict, and are more consistent with the contour-defined interpretations. Second, note that when ω was 1, the 3D object represented
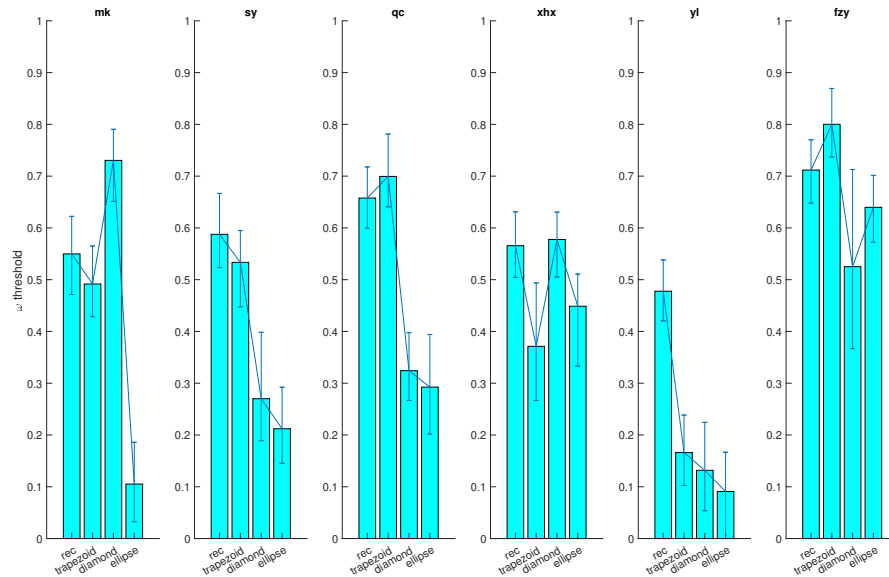
Figure 3.8: Estimated ω thresholds for each observer in Experiment 1: constant width ratio.

by the SFM display was consistent with the 2D contour, so theoretically speaking, it should have 100% 'yes' reponses; also, when ω was 0, the dots were translating with a constant speed in 2D, and it should have 0% 'yes' responses. However, the interesting finding is that the ω threshold was much smaller than 1. The average ω thresholds (with 95% confidence interval) for rectangle, trapezoid, diamond and ellipse were 0.59([0.53, 0.66]), 0.62([0.55, 0.70]), 0.43([0.34, 0.53]), 0.30([0.19, 0.38]). The ω threshold can be seen analogous to our previous α threshold (the proportion of projectively consistent speed in a display's speed profile). The similarity is that as either ω or α is getting smaller, the speed at the boundary will increase, which introduces more inconsistency with the motion-defined percepts. Therefore, just like α threshold, ω threshold can be treated as the transition where the perceived structure is no longer dominated by the 2D contour, and smaller ω thresholds mean that the 2D contour has more influence on the 3D percept. The estimated ω thresholds suggest that given a standard SFM display, when we 'shrink' the horizontal width of the area that is visible to the observer, the influence of contour was strong enough to overcome the inconsistency with the speed profile, and dominated the 3D percepts until it dropped below the ω threshold.

Figure 3.8 also shows a clear difference of ω thresholds across different shapes. Overall, the ellipse had a much smaller threshold compared to other shapes (average ω threshold =

0.30), and the diamond had a second smallest threshold (average ω threshold = 0.43) (except the second observer). The ω thresholds for rectangle and trapezoid were the largest among all observers (the average ω thresholds are 0.59, 0.61). The threshold difference in shapes also reflects the fact that 2D contour plays an important role in SFM, and different properties may have different degrees of contribution. Some shapes contain more information about the 3D structure than others. For example, the ellipse and diamond have more curvature (i.e. the most width variation along vertical direction), and therefore, the most variation in peak speed (along the shape axis). This variation in peak speed might provide more information about the relative depth in the vertical direction. In addition, compared to other shapes, ellipse has smooth curvature, and no sharp corners. Koenderink (1984) has shown that smooth 2D contour contains a great deal of information about the 3D structure. The 3D configuration may be more ambigous in the neighborhood of sharp corners in other shapes. Therefore, it makes sense that ellipse had the smallest ω threshold, and then the diamond, followed by the trapezoid and rectangle.

In summary, the results of Experiement 1 showed that the 2D contour plays an important role in SFM. The local presence of a 2D contour is a strong local cue to the occluding contour of a 3D object (where surface slant equals 90°). It appears to be a strong enough cue to override the inconsistency with the speed profile and dominates the perceived 3D structure, even when a small proportion of the display (in the horizontal direction) is visible. Also, the difference in ω thresholds of different shapes suggests that the importance of 2D contour seems to be influenced by certain shape properties, such as smoothness and variation in shape width in the vertical direction. It is worthwhile to explore in more detail other properties of shape that determine the dominant role of 2D contour in SFM. Moreover, since the results show that observers are very tolerant of ω deviating from the projectively consistent $\omega = 1$, we can continue the idea of separating motion and aperture, and use the results of Experiment 1 to do more interesting explorations of shape properties. For example, manipulating the two axes (motion axis and a shape's symmetric axis), and explore its effect on the SFM percepts.

3.3 Experiment 2: translating the shape axis

In the first experiment, we mainpulated the amount of display that was visible in the horizontal direction, which was also orthogonal to the shape's axis. The results showed that the 2D contour was able to dominate the 3D percepts even when there was large inconsistency between the speed profile and the contour-defined percepts. Its dominance vanished gradually when the width ratio drops below its ω threshold. However, note that this dominance of contour shape is just in terms of horizontal width ratio, and it seems like the easiest transition we can explore by introducing conflict between motion and contour. Next we will continue the idea of separating the motion and shape cues by separating the motion axis from the shape's axis.

Note that in tranditional SFM displays, as well as in Experiment 1, the axis of rotation and the symmetric axis of the shape coincide with each other, which is called coaxiality. The property is so 'common-sense' that people normally take it for granted and ignore the importance of it (see Morikawa (1999) for a related point). On the one hand, we know that for the bilateral symmetric shape, the symmetric axis contains vital information about the object's structure. On the other hand, if an object undergoes self-rotation and the silhouette stay fixed, it must be rotationally symmetric and rotating along its own symmetric axis. Therefore, coaxiality must be an important factor for the contour-defined percept in SFM.

In Experiment 2 and Experiment 3, we break coaxiality between motion and shape either by translating or rotating one axis away from the other. Our hypothesis is that the contour-defined SFM percept will be weaker as we increase the separation between the two axes. We used Method of Constant Stimuli to estimate thresholds of axial translation and rotation that are similar to ω thresholds, which represents the transition where 2D contour no longer dominates the 3D percepts. In each display, the shape of the aperture is in the same shape class as the motion region, but the horizontal width is smaller. For example, for a rectangular aperture, the speed profile is defined in terms of a wider rectangle. Before any translation/rotation, they are concentric, and along each rib that is orthogonal to the shape's symmetric axis, the width ratio between aperture and motion region remains constant. We used $\omega = 0.65$ for the width ratio, which was much larger than the ω threshld so that the contour-defined percept would be strong. Also it enables us to translate/rotate each smaller aperture within the corresponding motion

region to a relatively large degree, without going outside the motion region.

3.3.1   Method

3.3.1.1   Participants

Eight Rutgers graduate students participated in the experiment. Six were males and two were females. All of the observers had normal or corrected-to-normal vision. The observers were paid for their participation.

3.3.1.2   Stimuli and design

3.3.1.2.1   Shape of contours

In this experiement, we used the same rectangle and trapezoid as in Experiment 1. However, we used barrel and hexagon, instead of ellipse and diamond. We did not use ellipse and dimaond because the aperture would go outside of the motion region with even a small translation. Figure 3.9 shows the contours used in the experiment. As in Experiment 1, the green contours define the motion region, and the black contours define the visible aperture region. Different black contours represent different levels of translation (see the next section *Translation levels* for more details). The areas of all shapes remained constant (the green shapes have an area of 7.2471e+04 $pixel^2$, and the black ones have an area of 4.7106e+04 $pixel^2$, and the visual angle is around $2.7°$ in dva). Also, the range of horizontal widths of the barrel and hexagon are quite similar (visual angle: $1.09°$), so that the two shapes have nearly the same variation of horizontal width along vertical direction (54 pixels). Note that although the trapezoid has similar variation of width along vertical direction (visual angle: $1.2°$), it has the same height as other shapes, which means that the rate of horizontal width change $\frac{\Delta\ horizontal\ width}{\Delta\ vertical\ distance}$ was half of the barrel and the hexagon's, since the later two shapes are also symmetric along their horizontal symmetric axis (y = 0). Therefore, the barrel and the hexagon contain more variations of horizontal width in vertical direction (or curvature), compared to the trapezoid.
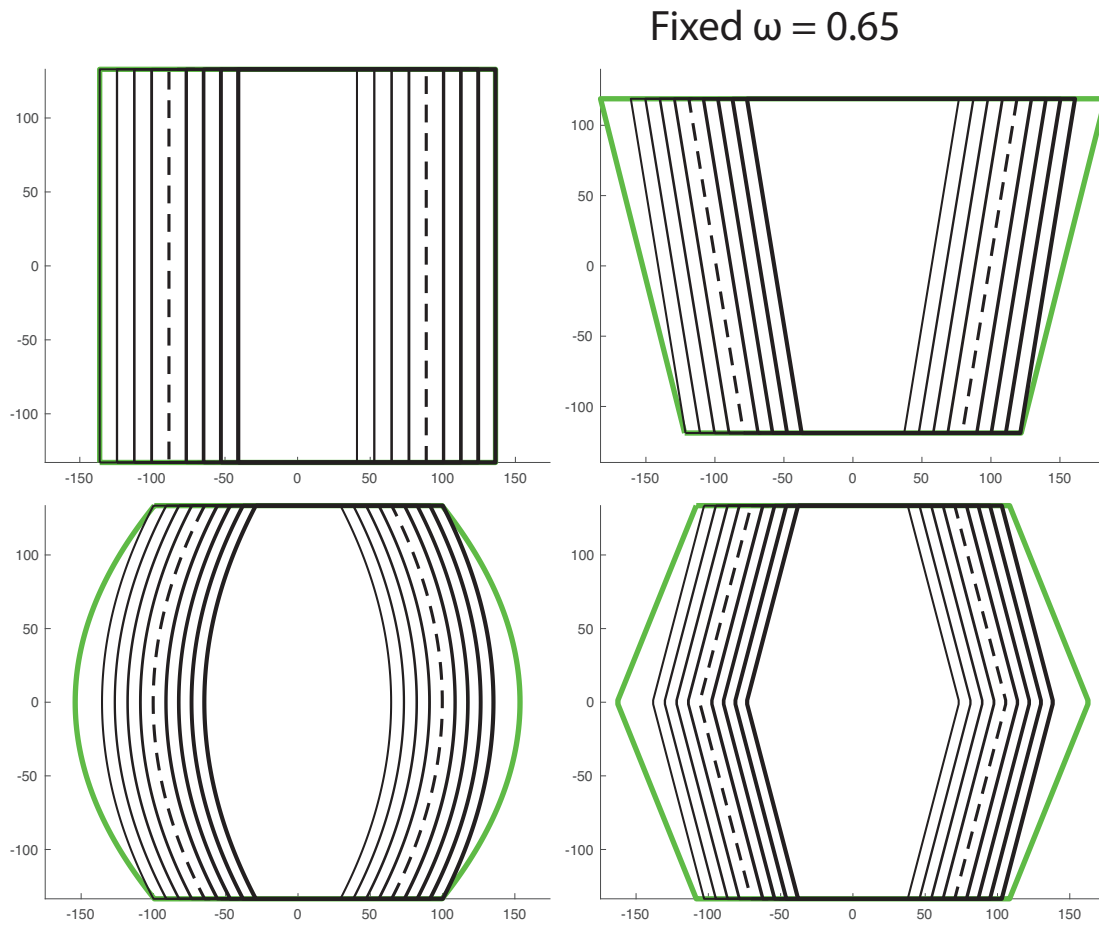
Figure 3.9: The contours used in Experiment 2. The green contours define the motion of dots, the black contours define the aperture. Different black contours represent different translation ratios, from no translation to the maximum it can reach (the aperture would be tangant to the green contour).

3.3.1.2.2 Translation levels

The apertures (the black contours in Figure 3.9) were generated from the motion region (the green contours) in the following way:

1. First, we multiplied the horizontal width of the green contours by 0.65. i.e., $\omega = 0.65$. Note that this value is larger than the ω thresholds in Experiment 1, but still enables us to do enough translation.

2. Then, we shifted the black contour horizontally (either left or right) from the common axis through different distances.

3. The shift ratio is defined as the distance of shift over the width of the aperture in the horizontal direction.

Here, we used five shift ratios. Level 0 means the coaxial case where the axes of the black region and the green region are the same; level 4 represents the most extreme case, which corresponds to the maximum translation possible while still staying within the motion (green) region. Note that for different shapes, level 4 means different shift ratios. The maximum shift ratio for the rectangle was 0.175, while the shift ratio for other three shapes was 0.115. Therefore, in the psychmetric function (Figure 3.10), the x-axis ranges were different for different shapes. Also, the shift ratio can be positive or negative (either shift left or right relative to the aperture's axis), but when we compute their psychometric functions, we did not find any bias of the shift directions and therefore we pooled over positive and negative shifts, focusing only on their magnitudes.

3.3.1.2.3 Speed profile of dots

The speed manipulation was the same as Experiment 1. We applied the cosine speed profile to the dots based on their locations in the motion region (green regions). The average linear speed was constant across different displays (0.55°/s). Therefore, the SFM displays were consistent with rotating objects whose orthographic projections were the green contours. After that, we covered the displays with the shaped apertures (black regions), so that only dots within the apertures were visible to the observers.

For each shape, we generated seven different displays, and in each trial, one display would be chosen randomly, and shown through the aperture as described above.

3.3.1.2.4   Other controls

As in Experiment 1 and 2, we also used the same uniform-density control, and we introduced the same small in-aperture contrast and same motion direction counterblanace as in Experiment 1.

Overall, we have 5 shift ratios × 4 different shapes × 2 motion directions × 2 shift directions (left/right) × 8 repetitions, adding up to 640 trials. The trials were split into 8 blocks, and each observer came on two different days (within one week) to complete the whole experiment.

3.3.1.3   Procedure

The experiment was conducted in a quiet dark room. The observers sat in front of the same iMac screen at a distance of 105 cm. We used a chin rest to make sure the visual angle remained constant throughout the experiment. The displays and the experiment interface were coded and showed with Psychtoolbox-3 in MATLAB 2017a on an iMac (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The observers read through and signed the informed consent first. Then the whole procedure of the experiment was similar to Experiment 1, except that we used different displays. We used Method of Constant Stimuli to estimate the threhold of shift ratio for each shape.

3.3.2   Result and Discussion

We recorded observers's Yes/No responses, and then used psignifit toolbox by Schütt et al. (2016) (https://github.com/wichmann-lab/psignifit) to fit the psychometric curve (using logistic regression) and also calculated the threshold of shift ratio for each shape. We excluded two observers because their 'yes' responses were all around chance level (50%) regardless of condition, and we suspect that they did not understand the task. The other six observers' psychometric curves and estimated thresholds are shown in Figure 3.10 and Figure 3.11. Note

that in Figure 3.10 the x-axis indicates the actual shift ratios from the extreme case (left, larger number) to the coaxial case (right, smaller number), although for the psychmetric fitting we used negative shift ratios to maintain the shape of logistic regression. Also, since the maximum shift ratios were different for differnt shapes (0.175 for the rectangle, 0.115 for the others), the black line (rectangle) is wider than others in the horizontal direction. Figure 3.11 shows the threshold of actual shift ratios, and larger threshold means the contour has more dominance on the SFM percepts.
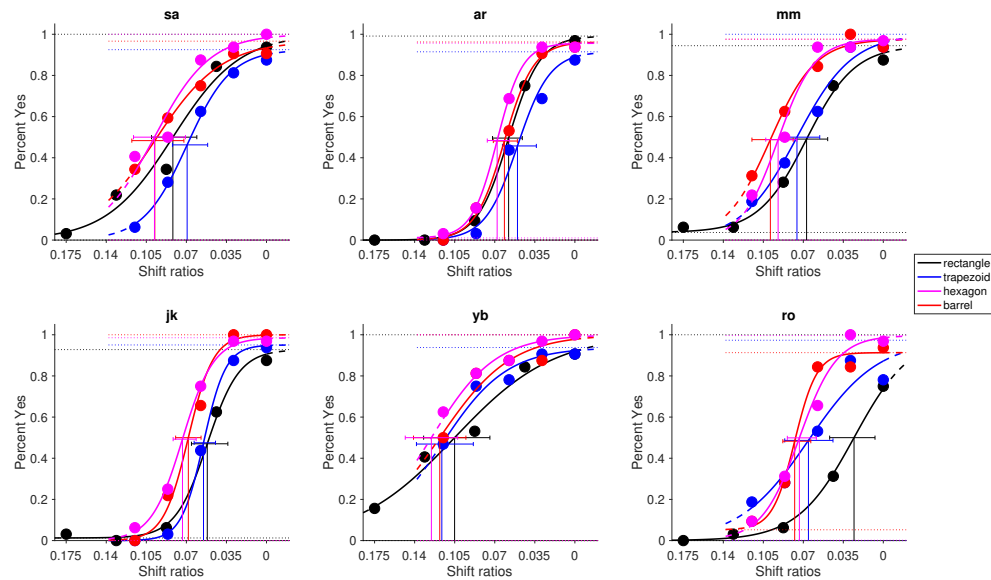
Figure 3.10: The psychometric functions of each observer in Experiment 2

The figures mainly convey two pieces of message.

First, as shift ratio gets smaller, the proportion of 'yes' responses also became smaller, and plateaued at or near 100%. It indicates that the 2D contour dominated the perceived 3D structure, and its influence got weaker as the horizontal shift between the motion axis and the aperture axis gets larger. Eventually, the inconsistency of two axes killed the contour-defined percept. The SFM percept was no longer a 3D shape whose occluding contour was defined by the image (aperture) contour. Moreover, Figure 3.11 showed that the average thresholds for rectangle, trapezoid, barrel and hexagon were: 0.065 (with 95% Confidence Interval of [0.045, 0.084]), 0.072 ([0.053, 0.087]), 0.087([0.070, 0.102]), 0.088([0.071, 0.102]). In other words, the amounts of translation represented by the thresholds were 6.5% (rectangle), 7.2%
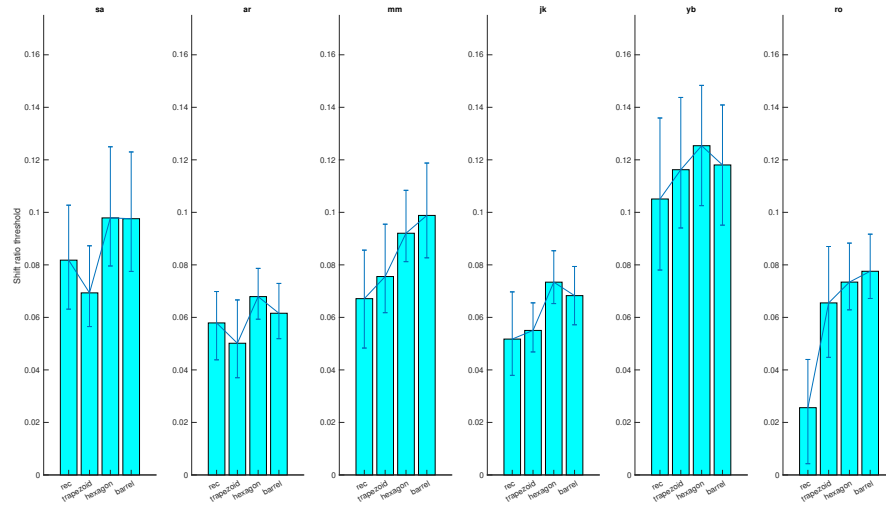
Figure 3.11: The estimated shift ratio thresholds for each observer in Experiment 2

(trapezoid), 8.7% (barrel), 8.8% (hexagon) of the corresponding shapes' widths. This means that when we shifted one axis away from the other one, the influence of the 2D contour was strong enough to tolerate the inconsistency in motion, up to 9% translation with respect to the width of the motion region (green in Figure 3.9).

Second, Figure 3.11 shows a clear increasing trend of shift-ratio threshold from the rectangle to the barrel and the hexgon. The results shows that our observers had the smallest tolerence with rectangle—when the translation was above 6.5% of the horizontal width, the contour no longer dominated the 3D percepts. The trapezoid had overall second smallest threshold, i.e., second smallest tolerance (7.2% of translation), and it was higher than the rectangle. The barrel and hexagon had overall the largest thresholds (8.7% and 8.8% of translation respectively), which means that our observers had the most tolerance with the inconsistency in those displays, and in other words, hexagon and barrel had larger influences on the 3D percepts. On the one hand, rectangle < trapezoid <barrel or hexagon in terms of variations of horizontal width in vertical direction, and the decreasing thresholds suggest that the more variation there was, the more influence the 2D contours had on the 3D percepts. On the other hand, the similar thresholds between barrel and hexagon also suggests that smoothness of barrel's left/right sides seemed to have little influence on the 3D percept.

Overall, the results are consistent with Experiment 1 in that the rectangle consistently

has smaller tolerance in the deviation from projective consistency, which means that the rectangle has less dominance on the 3D percepts, compared to other shapes. Also, the results are consistent with that the ω thresholds for ellipse and diamond (it was barrel and hexagon in Experiment 2) were similar in most observers. It suggests that variation of horizontal width and smoothness may have made the contour more vital in determining the 3D percepts in SFM.

3.4 Experiment 3: rotating the shape axis

The idea behind Experiment 2 and 3 is to separate the motion axis from the shape axis, and explore its influence on the perceived 3D structure. In experiment 2, we have already explored the separation by translating one axis away from the other one. The results of experiment 2 showed that the increased parallel distance between the two axes led to a decreasing dominating role of the 2D contour, and that observers were quite sensitive to the horizontal offset of the two axes. Similarly, in experiment 3, we are going to test the idea by rotating one axis away from the other.

3.4.1 Method

3.4.1.1 Participants

Seven Rutgers graduate students participated in the experiment. Two were females and five were males. All observers had normal or corrected-to-normal vision. They were paid after each of the two sessions of the experiment for their participation.

3.4.1.2 Stimuli and design

3.4.1.2.1 Shape of aperture

In this experiment, we used five types of shapes to define the aperture. We used the same rectangle, trapezoid, hexagon and barrel as Experiment 2 (see Figure 3.12), and we also used ellipse with the same size as other shapes. The reason to add ellipse is that none of the other four shapes has a completely smooth contour. Perhaps this was less relevant in previous experiments, since the motion direction was always orthogonal to the axis of shape, which means that
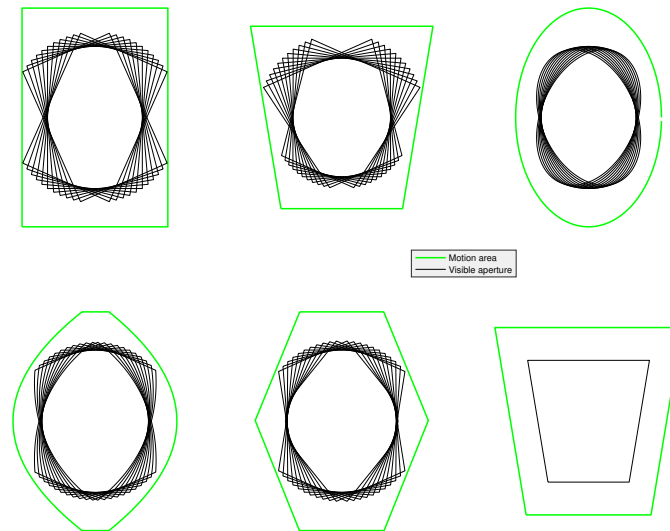
Figure 3.12: Contours of aperture (black) and dots motion (green) in Experiment 3. First five figures show all rotation conditions with all shapes, and the last one shows how we get the green region by expanding the black region.
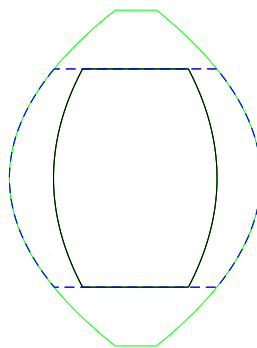
Figure 3.13: The way to generate the motion region (green) of barrel from aperture (black) in Experiment 3.

the moving dots would only disappear at left and right boundaries, and for the barrel, the dots would disappear at the smooth parts of the boundary. However, in the current experiment, the rotation of one axis means that the motion direction would no longer be orthogonal to the shape axis, and there would be dots disappearing at the corner between the straight and the smooth edges of the barrel. Therefore, we added ellipse where the contour was smooth everywhere.

3.4.1.2.2 Shape of dot motion region

Figure 3.12 shows both aperture contours (in black) and the regions that generate the dot motion (in green). In previous experiments, we first generated the motion region, and then contracted and translated it (horizontally) to get the aperture (the vertical heights were the same between motion region and aperture). However, in the current experiment, the motion region has larger height than the aperture, since we want to rotate the apeture and also make sure it is still within the motion region. Therefore, we first generated the aperture shape with the size of 4.7106e+04 $pixel^2$ (visual angle was around $4.4°$, same as previous experiments), and then expanded the apertures to get the motion region.

As in Experiment 2, when there was no rotation between the axes, the horizontal width ratio between the aperture and the motion region was $\omega = 0.65$. We chose the value of 0.65 because: 1) based on the results of experiment 1, 0.65 is much larger than the width ratio ω thresholds of all shapes, so based on conventional rigidity principle, observers should always perceive the display (such as the last one in Figure 3.12) as a rotating 3D object whose ortho-graphic projection is the 2D aperture; 2) it provides enough degree of freedom to rotate the aperture (black region) but also make sure it is still in the motion region (green region).

The ways we used to make sure ω was 0.65 were a little different for different shapes. For the rectangle, trapezoid, hexagon and ellipse, we simply uniformly expanded the horizontal and vertical dimensions of the aperture by a factor of $\frac{1}{0.65}$. For the barrel (Figure 3.13), we first multiply the x-coordinates of left and right splines by $\frac{1}{0.65}$, and then interpolated the extended splines whose vertical height was $\frac{1}{0.65}$ of the aperture. Finally we closed up the top and bottom by straight lines to get the motion region.

3.4.1.2.3 Rotation of axis

We separated the two axes by rotating the axis of the aperture. We defined 5 levels of rotation, equally spaced from $0°$ to $24°$, including both clockwise and counterclockwise rotation (Figure 3.12). $24°$ was the maximum rotation we could get and still make sure all the apertures still lie within their respective motion regions. In addition, we added a small random rotation angle ranging from $-6°$ to $6°$ to the whole SFM display (but it did not change the angle between the two axes), to make sure that the display would never be perfectly horizontal (even in the $0°$ rotation condition).

3.4.1.2.4 Speed profile of dots

We applied the same cosine speed profile to the motion (green) regions as before. Therefore, if the whole green region is visible to the observers, the SFM display would be consistent with a 3D object rotating with a constant angular speed(the average linear speed is $0.45°/s$), and the 2D contour is the projection of the 3D object. As before, we also ensured that the SFM displays had uniform dot density.

3.4.1.3 Procedure

In summary, we had five shapes (rectangle, trapezoid, hexagon, barrel and ellipse), and five levels of rotation for each shape. For each condition, we also counterbalanced the motion direction (left/right), and rotation direction of the aperture shape (clockwise and counterclockwise), and eight repetitions (but the random rotation of the entire display was different for each trial). Therefore, we had 32 trials for each main condition, and adding up to 800 trials in total. We split the 800 trials into two sessions, and five blocks in each session. Observers finished each session on two different days within a week. The instructions, task, practice trials and the experimental procedure was similar to Experiment 2. We recorded the 'yes'/'No' response to the same question with the two criteria.
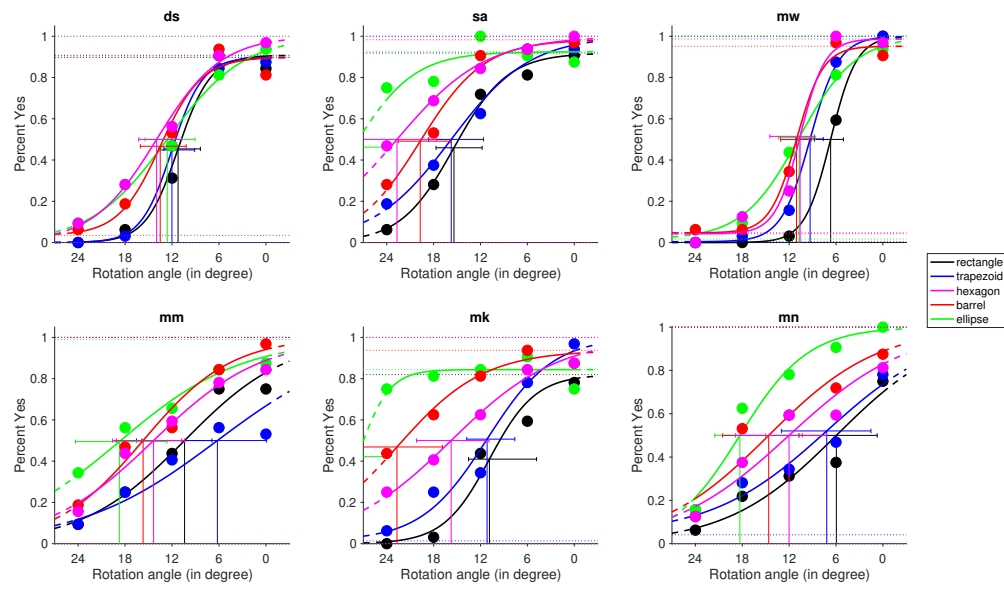
Figure 3.14: Psychometric curve of 'yes' response as a function of degree of rotation in experiment 3.
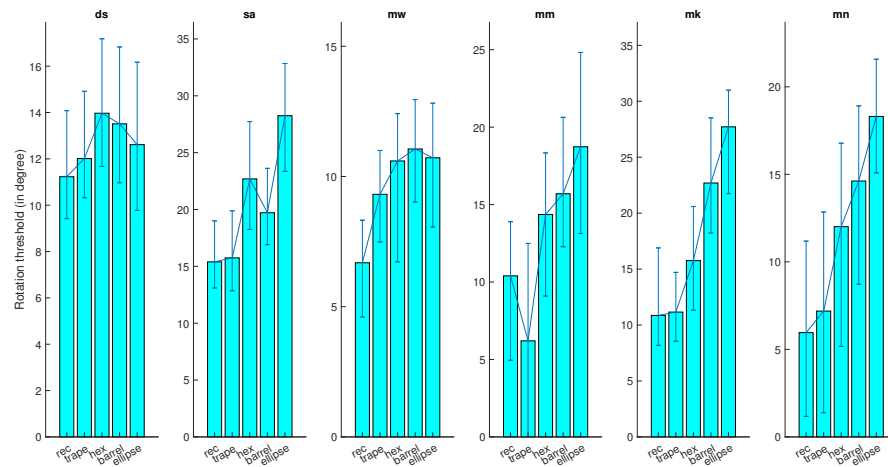
Figure 3.15: The fitted rotation threshold for each observer in experiment 3. The error bar shows 95% confidence interval.

3.4.2   Results and discussion

The proportion of 'yes' response were calculated and we used psignifit toolbox by Schütt et al. (2016) (https://github.com/wichmann-lab/psignifit) to fit the psychometric curve and then calculated the rotation threshold for each shape. Just like the shift ratio, the rotation threshold is the transition where the perceived 3D structure is no longer determined by the aperture. The psychometric fits are shown in Figure 3.14. The horizontal axis shows the rotation angle in degrees (from $0°$ to $24°$). Note that larger rotation angle indicates larger misalignment between the shape axis and the motion axis, and therefore larger rotation angle should have lower 'yes' response. Different curves indicate different shapes. The horizontal line segements indicate the 95% confidence intervals, and the center of the line segment indicates the estimated threshold of rotation angle. The top horizontal dotted lines are the plateau of each psychometric curve.

Figure 3.14 shows clear increasing trends of 'yes' response for all observers as rotation angle decreased (note that the rotation angle is decreasing along the horizontal axis). In the current study, rotation of zero means that the shape axis and the motion axis are aligned with each other. Compared with the other rotation angles, this condition is the 'basic' case that contains the least conflict between the motion-defined structure and the contour-defined structure. Therefore, the observers should have the most contour-defined percepts with this condition, which was shown in the figure that the 'yes' response was the highest and close to 100% for most observers. However, note that even when the rotation angle was zero, the displays were not completely consistent with rotating objects defined by their 2D contours, because we applied width ratio of 0.65 in the displays, i.e., the width ratio between each aperture and its corresponding motion area was 0.65 (see Figure 3.12). This may explain why for some shapes and for some observers, the 'yes' response did not reach 100%. Also, the proportion of 'yes' responses was below 0.1 for the basic shape (rectangle) when rotation angle is $24°$. It suggests that $0 - 24°$ was a reasonable range of rotation that gives us nice psychmetric curves ranging from approximately 0% to 100%. The estimated thresholds also confirm this. Figure 3.15 shows that the thresholds ranged from $6°$ to $15°$, depending on the shape and observer. According to De Bruyn and Orban (1988)'s study on motion direction discrimination threshold on differcent velocities, our display (with a speed of $0.45°/s$) has a direction discrimination

threshold around 8°. Our rotation thresholds for rectangle and trapezoid are somewhat larger than the this direction discrimination threshold (and it is even larger for ellipse and barrel). The small thresholds indicate that when there is a small orientation misalignment between the shape axis and the motion axis, the 2D contour would no longer dominate the SFM percepts. In other words, it seems that observers are also sensitive to the misalignment of axis orientation for the basic rectangle, and have more tolerance for barrel and ellipse.

Table 3.1: Averaged rotation threshold in degrees across all observers: mean and 95% confidence interval

rectangle	trapezoid	hexagon	barrel	ellipse
10.09	10.27	14.90	16.22	19.39
(6.28,13.27)	(6.23, 14.47)	(10.96, 19.42)	(12.19, 19.75)	(15.58, 23.06)

In addition to the effect of rotation angle, the results also showed an interesting effect of shape. The average rotation thresholds for different shapes, as well as their 95% confidence intervals, are shown in Table 3.1. The two main different properties of all shapes are variation of horizontal width, and smoothness. First, we can treat rectangle as the basic shape given it does not have either of the properties. And the threshold of rectangle was the lowest (10.09°), which means that it has the lowest tolerance to the misalignment of the two axes. In other words, even when the shape axis was rotated slightly away from the motion axis, the contour-defined percepts would be broken. Then we can group the other shapes by either of the two properties. On the one hand, if we separate the shapes by variation of horizontal width, the variation gets larger from rectangle to trapezoid to hexagon/barrel/ellipse, introducing more variations in the peak speed along the central vertical axis. The interesting finding was that ellipse/barrel/hexagon (the average threshold is 16.83°) has higher rotation thresholds compared to trapezoid (10.27°) and rectangle, which means that they had higher tolerance on the misalignment of axis orientation. On the other hand, if we group the shapes by smoothness, rectangle/trapezoid/hexagon contains only straight boundary, while barrel has two smooth edges on the left and right sides, and ellipse has completely smooth contour. The results showed that the ellipse(19.39°) almost always had the highest threshold, the barrel (16.22°) had the second highest threshold (in four observers). It suggests that as the 2D contour contained more smooth curvature, the rotation threshold became larger, meaning that apertures with smooth contours had more tolerance to

the misalignment of axis orientation, and had more dominance on the 3D percepts. All together, the pattern of rotation thresholds across different shapes suggests that both variation of horizontal width and smoothness play important roles on the contour-defined percepts in SFM.

Together with Experiment 2, the results of the two experiments show that observers are sensitive to the misalignment of shape axis and motion axis, either in the form of translation or rotation. Small misalignment would break contour dominance soon, especially the shift ratio. Also, both smoothness and variation of horizontal width ratio promote the contour-defined percepts.

3.5 Experiment 4: variation of width ratio $\Delta\omega$

In the previous three experiments, the aperture regions were generated from the motion regions by contracting their horizontal dimension by some constant ratio, meaning that each horizontal width ratio between the aperture and motion region remained the same along vertical direction (in Experiment 3, only the 'base' cases with rotation angle of zero had this property). However, in real life, when an object is occluded, the occluder does not necessarily have a shape similar to the object. In the language of geometry, it means that the horizontal width ratio would vary in vertical direction. Ramachandran et al. (1988) reported an interesting demo where an SFM display of a rotating cylinder shown through an aperture with a triangular shape. The bottom edge of the triangle had the same width as the cylinder. Instead of perceiving it as a rotating cylinder seen through an aperture, the display looked like a rotating cone whose 2D projection is a triangle. This suggests that the effect of contour shape can be strong enough to dominate the SFM percept even if the horizontal width ratio is not constant in the display.

The current experiment was inspired by their display. We manipulated the amount of variation of horizontal width ratio (let us call it $\Delta\omega$) between the aperture region and the motion region. We used the same rectangular motion region for all stimuli in this experiment, but varied the shape of the aperture (see Figure 3.16). To explore the effect of smoothness and convexity, we used a two by two facor design of smooth/straight $\times$ convex/concave shapes.

3.5.1 Method

3.5.1.1 Participants

Seven Rutgers graduate students participated in the experiment. Two were male and five were female. All of them had normal or corrected-to-normal vision. The observers were paid after the experiment for their participation.

3.5.1.2 Stimuli and design

The dot motion and aperture were generated independently. We used the same dot motion within a rectangular region and showed the motion through different apertures.

3.5.1.2.1 Dot motion

In all displays, we applied the cosine speed profile to a rectangular region (width and height: 272.63 pixels, height: 265.81 pixels, 5.5° in visual angle), decipting a rotating cylinder with a constant angular speed of 0.65 °/s, and the peak linear speed was 0.77 $°/s$. The same density control method was applied here, and on average, the dot density was $0.002/pixel^2$. We generated ten random SFM displays using the above parameters, and in each trial, one of the ten displays was chosen randomly.

3.5.1.2.2 Shape of aperture: smoothness and convexity

The apertures always shared the same top and bottom edges with the motion region (see Figure 3.16). We used a two by two factor design to define the left and right edges.

3.5.1.2.2.1 Smoothness   The first factor was smoothness. The results of previous experiments suggest that smoothness can affect the dominance of contour in SFM. Therefore, we include both hexagon and barrel. As shown with the red shapes in Figure 3.16, the left/ right edges of the hexagon was defined using three points. The points $(x_i, y_i$ (i = 1,2,3)) satisfies
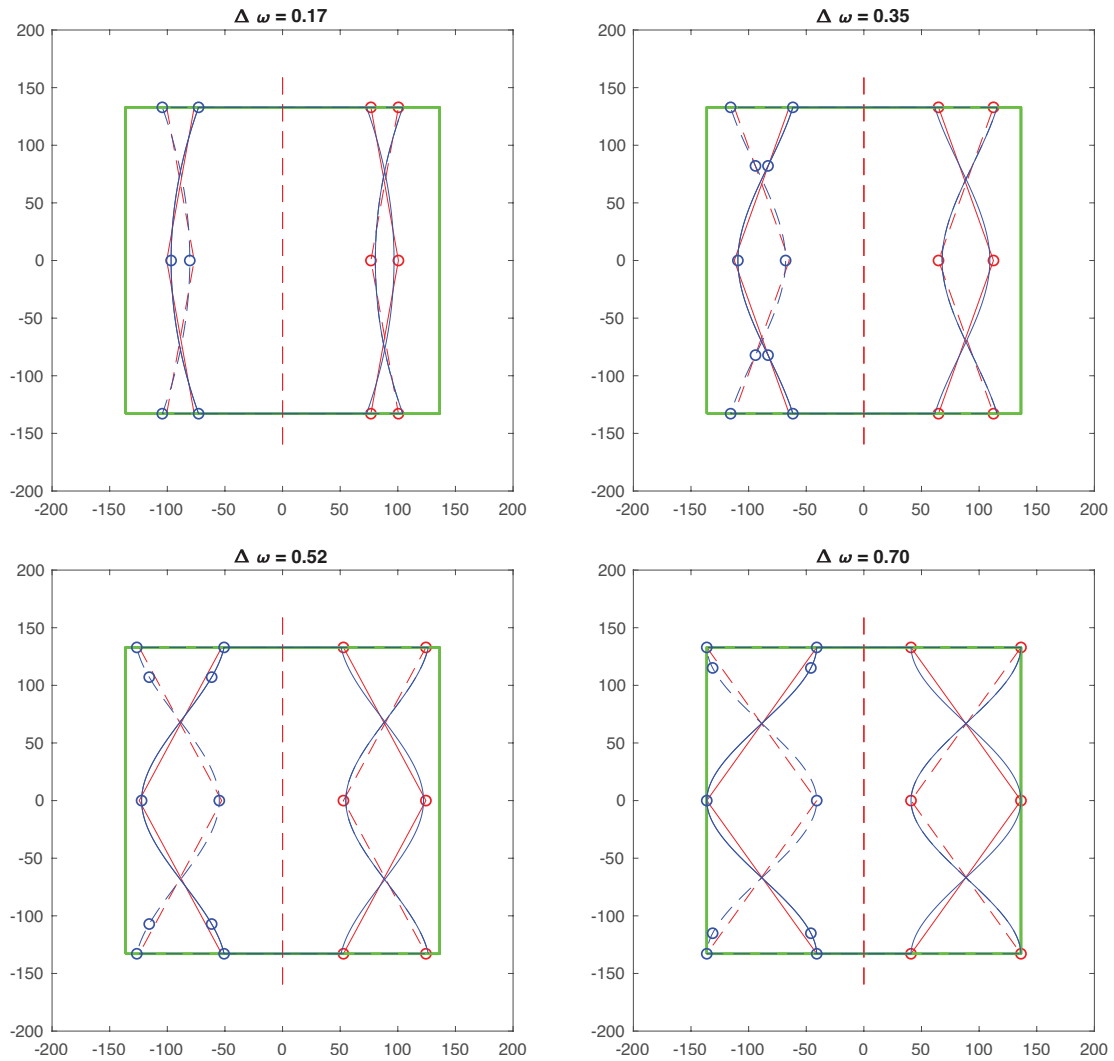
Figure 3.16: The contours in experiment 4. The green shapes show the motion region, and the blue (smooth curves) and red shapes (straight curves) show the aperture regions. The dots indicate the points that define each curve. The red dotted line indicates the axis of shape, as well as dot motion.

the following equations:

$$\frac{\sum_{i=1}^{3} x_i}{3} = \bar{\omega} * w$$

$$|x_3 - x_1| = \Delta\omega * w$$

$$y_1 = h \tag{3.2}$$

$$y_2 = 0$$

$$y_3 = -h$$

$\bar{\omega}$ and h are constant. $\bar{\omega}$ is the average width ratio ω, which equals to 0.65. The value was chosen based from the results of experiment 1, when ω was 0.65, the display already looked as vivid as a 3D object defined by the 2D contour. h and w are the height and width of dot motion region. $\Delta\omega$ is the other main manipulation—variation of ω. In other words, the hexagon would meet two criteria: it would have a desired variation of ω ($\Delta\omega$) as well as a fixed $\bar{\omega}$.

The smooth curves were created by interpolating cubic spline curves from either 3 or 5 dots (Figure 3.16). If $\Delta\omega$ is lower than 0.15, we used three points, otherwise, we used five points. The purpose was to make sure the curve did not have inflection points (i.e, it will be purely convex or purely concave). The three points were found by fixing their y coordinates as h, 0, and $-h$, and searching through the space of x-axis to find the points that achieve the desired levels of $\bar{\omega}$ and $\Delta\omega$: if you evenly sampled dots (x_i, y_i) along the curve, the mean width ratio of x coordinates $\bar{\omega} = 0.65$, and $\Delta\omega$ is the same as the corresponding hexagon. Similarly, for the five points, we simply added two symmetric points in-between the three points (see Figure 3.16). The difference is that we did not fix the y-coordinates and searched through a 2-D space (x-y space) to find the best solution satisfying the above two criteria.

3.5.1.2.2.2 Convexity The second factor was convexity. For the convex shape, there is greater inconsistency between the motion and aperture near the top and bottom (since the width ratio is smaller there), but the concave shape has greater inconsistency near the center. Research has shown that people would fixate more at the center of the shape (Henderson, 1993; Vishwanath & Kowler, 2003, 2004). If this is also true with our SFM displays, the convex shape would have more 'yes' responses to our task. Therefore, we included both convex and concave shapes. The convex shapes were generated by the above method. The concave shapes

were generated by flipping the convex curves horizontally, so they would still have the same $\Delta\omega$ and $\bar{\omega}$.

3.5.1.2.3 Shape of aperture: variation of width ratio $\Delta\omega$

Variation of width ratio $\Delta\omega$ was the third manipulation in the current experiment. It is defined as the maximum change of width ratio ω along vertical direction: $\Delta\omega = max(\omega_i) - min(\omega_i)$, and $\omega_i = \frac{width\ of\ aperture}{width\ of\ motion\ region}$. We defined four levels of $\Delta\omega$. The maximum $\Delta\omega$ was 0.7, which also made sure that the aperture was within the motion region, and $\bar{\omega} = 0.65$ so that all the shapes would have the same area. The other three levels were 0.17, 0.35, 0.52, as shown in Figure 3.16.

In summary, there were three manipulations of the aperture shape: smoothness, convexity and variation of width ratio $\Delta\omega$, adding up to 16 ($2 \times 2 \times 4$) conditions. For each condition, we also had motion direction (left/right) counterbalance, and 12 repetitions (the dot motion was selected from one of the ten displays randomly, so it was not exactly the same), adding up to 384 trials. The order of trials was randomized for each observer.

3.5.1.3 Procedure

The experiment was conducted on the same iMac machine with the same Psychtoolbox-3 in MATLAB 2017a on a MacOS (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007). The room was dark and quiet. Observers first signed the informed consent. The experiment consisted of three parts: instruction and three examples (a flat 2D surface, a rotating cylinder, a rotating cube shown through a circular aperture), nine practice trials (same as previous experiments), then the actual experimental trials. The experimental interface was similar to the previous experiments. The order of the 384 trials was randomlized for each observer, and they were split into 6 sessions. Observers could take breaks in between sessions. The whole experiment took about 50 minutes.
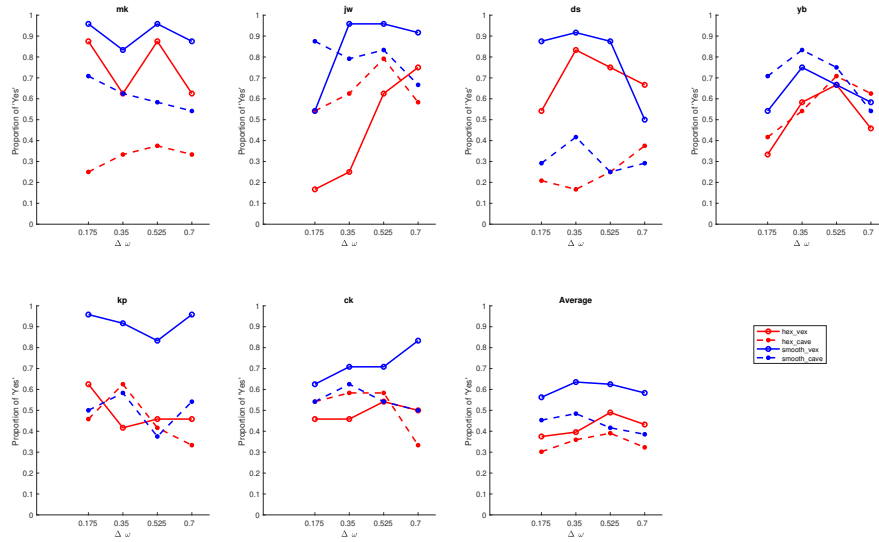
Figure 3.17: The results of experiment 4. The blue curves represent smooth shapes, and the red ones represent the hexagons. The solid lines represent convex shapes, and the dotted lines represent concave shapes. The last plot shows the average results.

3.5.2 Results and discussion

One observer's 'yes' responses in each condition never exceeded 50%, so we did not include her data in further analysis. Figure 3.17 shows the proportion of 'yes' reponses as a function of variation of width ratio $\Delta\omega$ for each condition. The last plot shows the average results, while the others show the individual results. The smooth shapes are shown in blue, and the hexagons are shown in red. The solid lines represent convex, and the dotted lines represent concave shapes.

We also performed Binomial Logistic Regression using the response as dependent variable, and smoothness, convexity, and $\Delta\omega$ as independent variables. The results showed significant effects of smoothness ($p < 0.0001$) and convexity ($p < 0.0001$), but no significant effect of $\Delta\omega$ ($p = 0.596 > 0.05$). The results of Binomial Logistic Regression further support the trends in Figure 3.17. However, there are some individual differences (see Table 3.2. Five out of six observers (except DS) indicate significant effects of smoothness, half observers (MK, DS, KP) show significant effects of convexity, and only one observer (JW) shows a significant effect of $\Delta\omega$. The individual data also shows strong effects of smoothness and convexity, but little effect of $\Delta\omega$.

Table 3.2: p-value statistics of Binomial logistic regression for individual observer in Experiment 4

name	smoothness	convexity	Δω
mk	0.000*	0.000*	0.262
jw	0.000*	0.130	0.000*
ds	0.085	0.000*	0.569
yb	0.009*	0.171	0.422
kp	0.000*	0.000*	0.150
ck	0.007*	0.146	1.000

* means significance

Figure 3.17 shows that in most cases the blue lines were above the red lines, indicating a clear effect of smoothness. It means that smooth shapes have a greater dominance on the 3D percepts. It is consistent with the previous experiments in that smooth displays such as ellipse and barrel tend to have greater tolerance to deviations from projective consistency, and the contour-defined percept was also stronger. Koenderink (1984) has shown that more information about 3D structure can be inferred from the smooth 2D contour line. Smooth contour may contains more information about the 3D shape compared to the other polygonal contours.

Figure 3.17 also shows that the solid lines were almost always above the dotted lines, which indicates that convex contours had more 'yes' reponses compared to the concave contours. It suggests a bias towards convexity, and that the convex contours have more dominiance on the perceived structure. Studies have shown that convexity is a strong cue to figure. Peterson and Salvagio (2008) have shown that in the context of figure-ground, the convex regions are much more likely to be seen as figural. Froyen et al.'s studies have also provided evidence for this (Froyen et al., 2013). Here, our results also show that convexity is a strong shape cue in the context of SFM. In addition, as mentioned before, concave shape has more inconsistency around center area because it is narrower part at the center, so it has smaller width ratio; whereas convex shape has less inconsistency around center area because the width ratio is larger. The bias towards convexity may also be due to the fact that people are more likely to fixate near the center of a shape (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007), and that is where the width ratio is the highest for the convex shapes, so convex displays contain 'less' inconsistency.

However, Figure 3.17 shows little effect of $\Delta\omega$, with the lines almost flat across different $\Delta\omega$. One possible reason is the competition between the two cues. As $\Delta\omega$ incresases, the curvature of left/right edges increases, which leads to more variation of horizontal width along vertical direction, so the contour shape would be predicted to have stronger influence and hence more 'yes' responses. However, as $\Delta\omega$ increases, the inconsistency in motion also increases (has lower horizontal width ratio), which might predict fewer 'yes' responses. The two cues have opposite influence on the responses, therefore are likely to cancel out each other's influence.

Another possible explanation is the location of fixation. The width ratio ω varies across different heights, so the fixation would influence the vividness of contour-defined percepts. For example, for Ramachandran's demo of the triangular aperture, the inconsistency in motion is more obvious near the top apex compared to the bottom edge, and one's fixation location will influence the percept of a cone vs cylinder. Indeed, this is the reason why in all previous experiments we kept the width ratio constant.

In summary, the current experiment has shown the important role of contour geometry in SFM in the context of displays that contained varying horizontal width ratio. Smooth and convex contours have more influence on the perceived 3D structure, perhaps because they contain more information about the potential 3D structure. However, $\Delta\omega$ has surprisingly little effect. This may due to the competition of the two cues, or due to the location of fixation.

3.6   General discussion

The traditional SFM literature focuses on the role of speed profile of the dots, leaving little room for the role of contour geometry. However, some demos (Ramachandran et al., 1988) and studies (Froyen et al., 2013) have shown that when there is an inconsistency between motion and contour geometry in SFM, a 3D percept consistent with the contour geometry can be still vivid, which points to the possible contribution of contour geometry. Our previous work also provides evidence for the role of contour shape (He, 2016), but we did not explore whether the 3D percept was consistent with the contour geometry, making it less convicing. The current study aims to fill the gap of the role of contour geometry in SFM as well as the conditions where it dominates the 3D percept. Our basic idea is similar to Ramachandran's demo (Ramachandran et al., 1988), in that we independently manipulated the motion region that defined the projectively consistent SFM display and the shape of the aperture through which the display was shown. We used a 'Y/N' task about contour-defined percept that the perceived 3D structure is consistent with the 2D contour. In the first three experiments, the aperture and its corresponding motion region belonged to the same shape class (with constant horizontal width ratio ω). In experiment 1, we first explored how much inconsistency in motion our visual system could tolerate and still have the contour geometry dominate the 3D percept. We found that the contour dominated the 3D percept until ω dropped to around 0.6 for the rectangle and 0.3 for the ellipse. In experiments 2 and 3, we explored the role of coaxiality (motion and aperture shape share the same axis) by disrupting coaxiality using translation or rotation. We found that our observers were quite sensitive to the deviations from coaxiality in either form, suggesting that coaxiality is an important factor for contour dominance. In the last experiment, the motion region was consistent with a rotating cylinder, but we used different aperture shapes, introducing different levels of variations in horizontal width ratio along the vertical direction. However, this had little effect on the contour-defined percept. In addition, throughout the four experiments, we also found that, the smooth contour as well as the ones with more variation of horizontal width (i.e. curvature), had stronger influence on the SFM percept. Overall, the results support our hypothesis that contour geometry plays an important role in SFM, its influence is strong enough to override the inconsistency in motion,

and may provide an explanation for why one can readily obtain 3D SFM percepts that are not projectively consistent with the motion of the dots.

3.6.1 The strong influence of contour geometry

The most important implication of our results is the importance of 2D contour in SFM, which should not be ignored. The presence of an image contour provides a strong cue to the occluding contour of a 3D object, where the surface slant is $90°$. The perceived depth that is due to the presence of an image contour is strong enough to override the inconsistency in motion (such as the non-zero speed at the boundary) and dominated the percept. This allows us to make sense of some previous reports in the literature. For example, in Ramachandran's demo of a cylinder shown through a triangular aperture, the triangular shape is a strong cue to a rotating cone, though the motion was generated from a rotating cylinder. In the figure-ground study by Froyen et al. (2013), the regions with constant speed were perceived as rotating columns because of the existence of the convex contours. Thompson et al. (1992) also showed an illusion of rotating cylinder from random-dot texture within a central rectangular area translating with constant speed, surrounded by texture translating in the background region with opposite speed. They showed that as the width of the central region decreased, the report of a cylinder percept increased, suggesting a strong boundary effect. However, they only used rectangular regions and did not manipulate contour geometry.

In the related phenomenon of kinetic depth effect (KDE) (Wallach & Oconnell, 1953), a dynamically changing silhouette (e.g. resulting from the rotation of a 3D object) can be readily perceived as a 3D structure. Thus contour geometry is recognized to play an important role in this context, where the contour is dynamically changing (and there are no moving dots). In the SFM literature, by contrast, the role of contour geometry is standardly ignored. There are two possible reasons for this. First, a typical SFM display contains moving dots, but no explicit contour—though there is clearly a contour defined implicitly by the region filled with the dot texture. Second, and more importantly, the shape of this implicit contour is typically static in SFM displays. What our results show is that even when the contour geometry is static in SFM displays, it still plays a strong role in 3D interpretation—indeed so much so that it can easily override the influence of dot motion.

3.6.2 The conditions for a contour-defined percept

Since the role of contour geometry cannot be ignored, the next question is what geometric properties determine whether the aperture shape will dominate the 3D SFM percept (i.e. "pull" the perceived 3D structure to be consistent with the contour). The current study explored some basic factors that influence the dominance of 2D contour.

3.6.2.1 Horizontal width ratio ω

First of all, we explored the most basic case where the aperture and motion region belong to the same shape class such that the aperture shape is simply a "compressed" version (in the horizontal direction) of the motion region, and only part of the motion region is visible. We found that the value of ω threshold is quite similar to our previous α threshold. α is the weight of the cosine speed profile in a linear combination with the constant speed profile. As the aperture gets narrower, the proportion of the whole cosine speed profile also become smaller, which is similar to adding a constant speed component to the actual 2D speed as we did for the α manipulation. Therefore, as ω (or α) gets smaller, the inconsistency of the speed profile will become larger. More importantly the small ω thresholds showed that when the horizontal width ratio is constant, the geometry can override the lack of consistency with speed profile (for example the non-zero speed at the boundary), and dominate the 3D percept when the inconsistency was as high as 40% for the rectangle (and it was even higher for some shapes, e.g. 70% for the ellipse).

3.6.2.2 Role of coaxiality

Second, our results indicate the importance of coaxiality in having the contour define the shape of the perceived 3D object in SFM. When the coaxiality between motion axis and shape axis was disrupted, the contour-defined percept diappeared quite soon. Studies have shown that shape's axis of symmetry contains rich information about the shape's structure, and it is very important in shape perception (Tyler, 1995; Kayaert & Wagemans, 2009; Machilsen, Pauwels, & Wagemans, 2009). In our displays, as the motion axis deviated from the shape's symmetric axis, the speed profile was not symmetric along the shape axis, in other words, the shape axis

no longer served as the symmetric axis of motion. Therefore, the structure conveyed by the shape's axis conflicted with the motion-induced structure, and it killed the contour-defined 3D percept. In addition, our results showed that the two basic transformations of axis, translation and rotation within the projection plane, both killed contour dominance once the transformation exceeded some thresholds. Since other complex transformation can be decomposed into these two factors, our results suggest that any deviations from coaxiality within the projection plane would be likely to kill contour dominance in SFM.

3.6.2.3   Role of variation of horizontal width

In addition, we used several basic shape classes—rectangle, trapezoid, hexagon, barrel, diamond and ellipse. The difference of thresholds across different shape classes suggests that certain shape geometries influence the role of contour in SFM.

First of all, the strength of contour-defined percept increases as the variation of horizontal width (i.e. curvature) increase. The variation in horizontal width increases from rectangle, to trapezoid, to hexagon and barrel, to diamond and ellipse. However, the inconsistency in motion is the same for different shapes because 1) with constant horizontal width ratio ω, the proportion of cosine speed profile that is visible is constant along each horizontal slice (regardless of the absolute width); 2) all displays had the same average linear speed and the same area. Since variation of horizontal width does not influence the motion, the difference in performance suggests that contour shape influences the percept, and its influence is stronger with more variation of horizontal width.

Note that the variation in horizontal width introduces non-zero velocity gradient in vertical direction, aside from the gradient in the horizontal direction. A study by Braunstein and Andersen (1984) showed that the velocity gradient in the vertical dimension is perceived as curvture in depth, and is important for depth judgments. Therefore the vertical gradient provides extra information about relative depth, and leads to a higher proportion of contour-defined percepts in our experiment. Koenderink and Van Doorn (1986) showed that the optic flow created by the motion of a 3D object can be decomposed into three differential invariants *curl*, *deformation*, and *divergence*, and deformation is very informative about the local slant. In vertical direction, the velocity gradient results in the deformation of optic flow and suggests

non-zero local slant, which may provide extra information about the contour shape. It may explain why variation of horizontal width makes the contour-defined percept stronger.

3.6.2.4   Role of smoothness

Second, our results show that smooth countours have greater influence on the SFM percept. The most obivous case is that the ellipse consistenly had the strongest contour-defined percept, followed by the barrel. Koenderink (1984) has shown mathematically that smooth contours convey a lot of information about the 3D surface curvature in the case of 2D line drawing. However, the other shapes have sharp corners where the 3D interpretation may be ambiguous. Thus the ellipse is more informative about the 3D structure. Another possible reason is the accidental-view of straight top and bottom edges of other shape classes. The projection of an ellipsoid would always be a smooth ellipse, regardless of viewpoint. However, for other shapes, such as the cylinder, the projection of a vertical cylinder is a rectangle only if the line of eyesight is within the central horizontal slice of the cylinder (otherwise, the top and bottom edges will have some curvature). Therefore, the ellipse is a more typical representation of a rotating 3D object compared to other shapes, and it has stronger influence on the SFM percept.

3.6.2.5   Role of convexity

In addition, Experiment 4 showed a clear bias for convex shapes, i.e., convex shapes have more dominance on the SFM percept. Liu, Jacobs, and Basri (1999) showed that the convex contour was more likely to be perceived as a single object in amodal completion. Concave extrema are usually the place to segment parts (Hoffman & Richards, 1984; Hoffman & Singh, 1997), while convex contour means the continuation of single part (Rosin, 2000; Barenholtz, Cohen, Feldman, & Singh, 2003; Bertamini & Farrant, 2005; Cohen, Barenholtz, Singh, & Feldman, 2005; Feldman & Singh, 2005). In the rotating column studies by Froyen et al. (2013), the convex regions were more likely to be perceived as figural (rotating objects). Our results are consistent with those findings in that convexity is a strong cue to a single object, and its dominance is stronger than concave contours. Besides, in our case, the possibility that observers tend to fixate the a shape could also have played a role, since the central area for convex shapes has less inconsistency (larger horizontal width ration).

3.6.3 Future directions

All in all, the current study provides direct evidence for the important role of contour geometry in SFM. The local presence of a 2D contour is a strong local cue to the occluding contour of a 3D object (where surface slant equals $90°$). The influence of occluding contour can be strong enough to override the inconsistency of the speed profile, hence dominating the SFM percept. We also found that certain shape properties influence the dominance of 2D contour, such as horizontal width ratio, coaxiality, variation of horizontal width (curvature) and smoothness.

Our study introduces a new approach to study SFM. First, unlike traditional displays that generated by projecting a 3D rotating object, we generated the displays directly by manipulating the 2D speed profile and contour shape in the image; and second, we independently manipulated the speed profile and the contour geometry, so we can study the combined effect of any speed profile and any contour shape as needed. The two innovations enable us to manipulate quantitatively the inconsistency between the two sources of information, and explore more deeply the role of contour geometry, as well as the interaction between the image motion and contour geometry. The current study is just a start to explore the role of contour geometry in SFM. Future work should examine more complex shape properties and their interaction with the image motion.

Our statement that projective consistency is not necessary sounds a little surprising at first. One way of understanding this may be in terms of a probabilistic model. The current study has shown that both motion and contour geometry are vital in SFM percepts. It suggests that we can reframe SFM as a cue combination problem. We can describe the perceived 3D surface mathematically as a slant and a tilt map of locations (x, y) on two likelihoods, one is in terms of motion profile, and the other is about contour geomtry. The conflict between the two cues speaks to the violation of projective consistency: the 3D percept violates projective consistency in terms of the image motion. However, the effect of contour pulls the 3D percept to be more consistent with contour geometry. The relative strength of the cues will determine the final percept. However, the slant and tilt at neighbouring (x,y) locations cannot be independently determined because there are non-local constraints imposed by surface smoothness. Therefore, the combination of the two cues is unlikely to be captured by a simple linear combination based

on their respective 'reliabilities' (as in standard cue combination models, REF????Eners' and banks, 2002, Haptics and vision, Robbis and Jacobs in vision, binocular, textures...?????). It is more likely to be determined by a complex host of factors, such as the ones studied here (width ratio, coaxiality, etc.). More empirical studies are needed to show how contour geometry and image motion interact with each other, including the measurement of their relative strength, the measurement of their inconsistency, and the concrete way in which they combine the two cues. Therefore, more empirical work would be needed before a full mathematical model can be developed. However, the results reported here (along with further empirical studies) will surely inform any attempt to mathematically model this combination problem in the future.

Chapter 4

General discussion

4.1 The role of projective consistency

In the SFM literature, it is always assumed that the 'recovered' 3D structure should be projectively consistent with the image motion. For a long time, this common assumption has existed implicitly in most SFM models. For example, the rigidity-based models look for a rigid 3D structure that is consistent with the image frames (Ullman, 1979, 1983). Some velocity-based models assume that the recovered 3D object should have a cosine speed profile which is projectively consistent with a rotating object (Braunstein & Andersen, 1984). Projective consistency is also widely assumed in many computer vision models, such as the factorization method (Tomasi & Kanade, 1992; Morita & Kanade, 1997), and the feature tracking models (Engels, Fraundorfer, & Nistér, 2008; Zhang, Dong, Jia, Wong, & Bao, 2010).

For a long time, the role of projective consistency has not been questioned, because the 2D display is normally generated by projecting an actual (or simulated) 3D object, and it is natural to think that the recovered 3D object should be consistent with that 3D object. However, some studies suggest that the 3D percept does not require projectively consistent 2D displays. Under certain conditions, constant 2D speed profile can trigger rotating percepts, such as the figure-ground study by Froyen et al. (2013), and the SFM study by Thompson et al. (1992).

These studies suggest that projective consistency may be a too strong assumption in SFM. However, note that none of these studies have explored systematically the role of projective consistency in SFM. The current work was aimed at systematically exploring projective consistency and reconsidering its role in SFM.

Unlike the traditional way of starting with a 3D object, we directly manipulated the speed profile of the 2D dot display, and used the speed profile to determine the speed of any dot

based on its location in the display. In the first study (Chapter 2), the speed profile was a linear combination of a cosine speed profile (which is projectively consistent with a 3D object rotating with a constant angular speed) and a constant speed profile (which represents 2D translation with linear motion). By controling the weight of the cosine speed profile α, we can control the degree of projective consistency in motion in any given SFM display. In He (2016), we applied the above speed profile to simple ellipse shape, and the results showed that observers could tolerate large deviations from projective consistency in motion. When α is as low as 0.6 (with 40% inconsistent motion), the 3D percept was still strong. That study also showed that observers were insensitive to projective inconsistency stemming from motion direction. In that study, the inconsistency arose from motion. One can argue that though the speed profile is not perfectly cosine, a 2D ellipse can be a projection of an ellipsoid, and observers are still seeking for a projectively consistent solution (an ellipsoid which is projectively consistent with the 2D contour), but perhaps allowing for some noise in motion.

To provide stronger evidence questioning the role of projective consistency, we introduced stronger deviations from projective consistency in terms of contour shape as well as motion profile in Chapter 2. We applied the speed profile to symmetric, asymmetric and parallel shapes. Note that there is no simple 3D object that is consistent with the asymmetric and parallel displays, since any asymmetric 3D object undergoing self-rotation will change its 2D silhouette over time. We found that the volumetric percepts were still robust when the motion contained as high as 40% inconsistency. Also, although there was an effect in the predicted direction (symmetric displays had slightly stronger volumetric percepts than asymmetric/parallel displays), the influence of symmetry was surprisingly small. This suggests that the global shape property symmetry has little effect on the 3D percept, and that global inconsistency in contour shape still cannot break the vividness of the volumetric percepts. Overall, these results provide stronger evidence calling into question the necessary role of projective consistency in terms of contour shape, as well as motion profile.

In the asymmetric shapes, globally speaking, the contour shape is not consistent with a rotating object. However, the speed profile along each horizontal slice is defined in terms of the contour shape, so locally speaking, the infomation contained in motion is correlated with the information provided by contour shape. For example, when α is 1, the speed at the boundary is

zero, which can be a cue to be an occluding boundary of a 3D object. So that locally speaking, perhaps symmetry or asymmetry does not make a difference as long as the speed profile is defined in terms of the local contour shape. This may provide an explanation for the similar performance for the symmetric and asymmetric displays.

In chapter 3, we introduced more dramatic deviations from projective consistency by separating the shape that defined the motion, and the shape of aperture through which the motion is seen. Now since the contour geometry and motion profile can be manipulated independently, the inconsistency exists both globally and locally, and it provides much more freedom to introduce any kind of inconsistency. As long as the motion region is not the same as the aperture, the display will not be projectively consistent with a 3D rotating object. The more difference we introduce between the two regions, the more projective inconsistency it introduces. The results of chapter 3 showed that when the aperture that defined the contour geometry differed from the motion region that defined the speed profile, the 3D percepts were still strong in some conditions.

4.2   The role of contour geometry

The results of chapter 2 as well as our previous work on projective consistency (He, 2016) raise the question why observers are so tolerant to large deviations from projective consistency in SFM. In SFM displays, there are actually two sources of information—motion and contour shape, but in the literature, most theoretical and empirical studies only focus on the role of motion, and essentially assume that contour geometry is not relevant on SFM. The rigidity-based models use local rigidity as a constraint to get the 3D structure of each local patch, then combine these local solutions into one global interpretation (Ullman, 1979, 1983). The velocity-based models build up the slant and angular speed locally from the deformation of optic flow, then integrate into a global solution (Domini et al., 2002; Domini & Caudek, 2003a). No matter what the model is based on, either position or velocity, they first build the 3D structure locally, then integrate the local 3D depth into one single solution. This implies that all dots within the display contribute equally to the SFM percept, regardless of their locations (near or away from the bounding contour). Noe of these studies has considered the role of contour geometry.

However, some studies have suggested that the role of contour geometry cannot be ignored in SFM. As mentioned earlier, in Ramachandran et al. (1988)'s demo, the rotating cone percept was consistent with the 2D triangle, in spite that the motion pattern was consistent with a rotating cylinder. Thompson et al. (1992) created a display of a rectangular texture region, surrounded by the background region, and the two regions were identical except that the textures were moving with different constant speeds in opposite directions. Thompson found that the region near the boundary had an illusion of rotation, but the rotating percept became weaker further away from the boundary. Also they found that when the width of the rectangle was narrower, the rotation percept was stronger.

These studies show that under certain conditions, people can have a vivid 3D percept with a display that is not projectively inconsistent with a 3D rotating object. More importantly, the 3D percept is determined by the contour geometry. This suggests the importance of contour geometry in SFM, despite the inconsistency with the motion profile. Connecting back to the issue of projective consistency, our hypothesis is that both motion and contour geometry contributes to the SFM percept, and the influence of contour geometry can often override the inconsistent motion. The role of contour geometry may provide a possible explanation for why projective consistency is not a necessary requirement as previous thought in SFM.

One difficulty in exploring the influence of contour geometry is that it is difficult to separate the influence of motion and the influence of shape since the motion is usually consistent with contour shape, and locally speaking it is consistent with a 3D object. As we discussed above, this may be the reason why in chapter 2 we did not find large difference between symmetric and asymmetric displays. The idea suggests that instead of exploring the global shape properties, it seems more important to focus on the local presence of a contour.

To separate the effect of motion and shape completely, in chapter 3, we use a new approach that separates the region that defines motion and the aperture which defines the bounding contour of the SFM display. Since the motion is no longer defined by the 2D contour, the motion is not even locally consistent with the contour geometry, and we can explore their separate contributions. In addition, in previous studies, we only asked about a general 3D percept, which could be ambiguous because the percerived 3D structure may be more complex and not be consistent with the contour geometry. Here we asked the specific question about contour-defined

percept, so the 'yes' response only captured the percept that was dominated by the contour geometry (i.e. a perceived 3D structure consistent with the image contour).

With this novel setup, we ran four experiments and explored the role of contour geometry step by step.

In the first experiment, the motion region and the aperture belonged to the same shape classes such that the horizontal width ratio (what we called ω) was constant along each horizontal slice. When ω is 1, the motion is consistent with the contour geometry. When ω decreases, the inconsistency between motion and contour increases, so the contour-defined percept should get weaker. The important finding is that the contour-dominance was still strong when ω dropped to 0.6 (and even smaller for some shapes). ω can be regarded as analogous to our previous α manipulation (the linear weight of projectively consistent speed, relative to the constant speed component). When ω is smaller than 1, only part of the whole cosine speed function is visible, which is quite similar to adding a constant speed component to the overall speed when α is not 1. This suggests that the non-zero speed at the boundary does not kill the 3D percept. Overall, the small ω thresholds show that the contour plays an important role in the 3D percepts, its influence is strong enough to overcome the inconsistency with motion, so when only part of the cosine speed profile was visible, the contour-defined percept was still strong.

In the second and third experiments, we broke the coaxiality between the motion region and the contour geometry (defined by the aperture) . The results showed that observers were quite sensitive to the deviation of coaxiality, either by parallel shift or rotation. We know shape's axis of symmetry contains vital information about the structure of the shape (Tyler, 1995; Kayaert & Wagemans, 2009; Machilsen et al., 2009). An object undergoing self-rotation along its symmetric axis should have a symmetric velocity field relative to its axis (if we ignore the motion direction). When the motion axis deviates from the shape's symmetric axis, the velocity field is no longer symmetric along the shape's symmetric axis, so that the contour-dominance disappeared as the deviation between the two axes increased.

In the last experiment, the horizontal width ratio between the motion region and aperture was no longer constant within the display. We introduced variation of horizontal width ratio $\Delta\omega$ in the display. We used the same rectangular region to define motion (which decipts

a rotating cylinder), but we changed the aperture shapes by using convex/concave polygons and smooth barrels with different values of $\Delta\omega$. We found that convex contours have stronger contour-defined percepts. This is consistent with previous studies that convexity is a strong cue to be a single object (Hoffman & Richards, 1984; Hoffman & Singh, 1997; Barenholtz et al., 2003; Bertamini & Farrant, 2005; Cohen et al., 2005). This is also consistent with the fact that people tends to fixate near the center of an object (Henderson, 1993; Vishwanath & Kowler, 2003, 2004), and convex displays have less inconsistent motion near the center compared to the concave displays.

However, the contour-defined percept was quite similar across apertures with different values of $\Delta\omega$, showing little effect of $\Delta\omega$. One possible reason is that when we increase $\Delta\omega$, the curvature of the left/right edges also increases. Larger curvature has greater dominance on the 3D percept, it will lead to more 'yes' responses to the contour-defined percepts. But as $\Delta\omega$ increases, the proportion of motion that is visible also decreases, introducing larger inconsistency in motion, and it will lead to fewer 'yes' responses. The effect of motion and shape go in on opposite directions and both get stronger with the increasing $\Delta\omega$, so their effects may cancel out. Another possible reason is that, since the variation of width ratio ω introduced different degrees of inconsistent motion at different heights, the strength of the 3D percept would depend on where observer looks within the shape. (This is also the reason why, in the first three experiments, we only used displays with a constant ω within each shape).

Also, throughout the experiments, we used several basic shape classes to define the motion regions as well as the apertures—rectangle, trapezoid, hexagon, barrel, diamond and ellipse. Those shapes mainly differ in the following aspects: variation of horizontal width (curvature) and smoothness. Note that since we equated the average linear speed as well as the size of aperture area across different displays, the inconsistency in motion is the same under same level of ω, shift ratio, or degree of rotation. Therefore, the difference in performance is due to these two shape properties. We found that as the shape curvature or smoothness increased, the contour-defined percept became stronger. It suggests that both shape properties make the influence of contour shape more salient in the SFM percept.

In particular, the variation of horizontal width increases from rectangle to trapezoid, to hexagon/barrel, and to diamond/ellipse. The variation of horizontal width introduces non-zero

velocity gradient in the vertical direction, aside from the gradient in horizontal direction. It thus introduces extra information about the shape's 3D structure. Koenderink and Van Doorn (1986) have shown that optic flow can be decomposed into three differential invariants *curl*, *deformation*, and *divergence*, and one can get the local slant from the decomposition of *deformation*. The non-zero velocity gradient in the vertical direction introduces the vertical component of the deformation of optic flow, and may provide more information about the local slant in that direction and therefore is more informative about the 3D structure.

Smoothness, on the other hand, seems to influence the dominance of contour geometry. The degree of smoothness increases from rectangle/trapezoid/hexagon/diamond, to barrel, whose boundaries on the left/right sides are smooth, and ellipse has completely smooth contour. The reason for the stronger dominance of smoooth contour may be due to the nongeneric view of the straight top/bottom edges as well as the corners between straight and smooth edges. For example, the rectangle is likely to be the projection of a cylinder or a cuboid. But the cylinder only works if the line of sight is within the plane that goes through the central horizontal cross section of the cylinder. And the cuboid only works under more strict condition that the line of sight should also be perpendicular to the vertical cross section of the cuboid. However, the class of ellipse can be the projection of an ellipsoid, regardless of view angle. In other words, ellipse is more likely to be the projection of 3D object than other shapes. Therefore, smooth contour dominates the 3D percept more strongly. This may also be the reason why smooth occluding contours carry so much information about 3D geometry (Koenderink, 1984).

4.3   Summary and future direction

The current dissertaion provides some new insights to the study of SFM. One innovation is the way that we create SFM displays. We manipulate the 2D speed profile (speed as a function of location) directly to create the SFM displays, rather than projecting 3D objects, as is typical in the SFM literature. The main advantage of our approach is that our displays need not to be necessarily projectively consistent with actual 3D objects, and this also enables us to define speed profile for asymmetric 2D contours. Therefore, we can manipulate projective consistency quantitatively by manipulating the weight of cosine speed component in the speed profile. The

second innovation is the complete separation of motion region and aperture (which defines the bounding contour of the SFM display). This gives us even broader freedom to manipulate the magnitude of deviation from projective consistency. More importantly, since the motion is defined independently from the 2D contour, now we can study the contribution of contour geometry without worrying about its possible confounding influence of the 2D motion.

With this new approach, the current work has provided direct evidence for the two themes mentioned in the Introduction chapter: projective consistency and the role of contour geometry.

First, we have shown that when an SFM display contains projective inconsistency either in motion or in both motion and shape, the 3D percept can be still strong. The study in chapter 2 showed that globally inconsistent 2D contours did not kill the 3D percept, and the 3D percept for asymmetric displays was almost as vivid as symmetric displays (though there is a small but consistent effect). Chapter 3 further confirms this statement by introducing much greater projective inconsistency with the separation of motion region and aperture. Both studies indicate that SFM percepts are not necessarily projectively consistent with image motion.

The statement that projective consistency is not necessary sounds a little surprising at first. One way of understanding this may be in terms of a probabilistic model. As a probabilistic problem, our final percept is a posterior that depends on both likelihood and prior. The measurments in the 2D display has its likelihood distribution, and projective consistency means that our percept is consistent with the likelihood, i.e, the projectively consistent solution should be the MLE. However, because of the prior, the posterior, i.e., our final 3D percept, will often shift away from the MLE (i.e., the MAP estimate), which could be one way to understand the violation of projective consistency.

Secondly, our work has also shown that the role of contour geometry cannot be ignored in SFM. In chapter 2, we found surprisingly little effect of global asymmetry on SFM percepts. There might be two reasons. First, locally speaking, the local speed profile along each horizontal cross-section is still consistent with 3D rotation because it is defined in terms of contour shape (the shape width along that cross-section). Second, the presence of contour is a cue to the occluding boundary of a 3D object. Both factors promote a 3D percept, and override the inconsistency arising in the process of intergrating local measurements into a global percept.

The results of chapter 2 suggest that we should focus on the local contour geometry (or the local presence of a contour), and also we should find a way to separate the contributions of motion and contour shape. In chapter 3, we separated the region that defined motion and the aperture, so that the two factors could be manipulated independently of each other, and we could explore the effect of contour geometry. Our results showed that contour geometry plays an important role in the SFM percepts, and several factors, such as horizontal width ratio, coaxiality, smoothness, variation of horizontal width, as well as convexity influence the strength of contour dominance. This further proves the importance of contour geometry in SFM.

Our results, as well as previous literature (Ramachandran et al., 1988; Froyen et al., 2013), have shown that the violation of projective consistency seems to always occur simultaneously when the contour geometry plays a dominant role in SFM. This suggests that the two themes are closely related, and moreover, the role of contour geometry may be the reason why 3D SFM percepts can be projectively inconsistent with image motion. To be more specific, both motion and contour geometry play an important role in SFM. When the two sources of information are in conflict, one source of information will often override the inconsistency in the other source of information and dominate the final percept. This suggests that we can reframe SFM as a cue combination problem. The experiments in Chapter 3 investigated and documented some of the ways in which the image motion and contour geometry interact, and the conditions under which the contour appears to dominate. However, further investigation will be needed to fully map out the way in which these two cues combine to produce a 3D structure.

This cue-combination way of looking at SFM may provide another way of understanding why the projective-consistency constraint is often violated in SFM, such that the 3D percept is often not projectively consistent with the image motion. If contour geometry plays an influential role in SFM, then (as we have seen in the experimental results) it can 'pull' or 'capture' the 3D percept, thereby resulting in a 3D percept that is not projectively consistent with the image motion.

In conclusion, the current dissertation has shown that projective consistency is not necessary in SFM, whether the inconsistency comes from motion or from contour shape. Moreover, we have shown the important role of contour geometry in SFM. This role is strong enough

to overide the inconsistency with image motion and dominate the SFM percept. We have also explored several basic factors that influence the dominance of contour geometry. Overall, the results show that the contribution of contour geometry cannot be ignored, and they calls for further investigation of how image motion and contour geometry combine to produce a perceived 3D structure in SFM.

Bibliography

Anandan, P., & Irani, M. (2002). Factorization with uncertainty. *International Journal of Computer Vision*, *49*(2/3), 101–116.

Asada, H., & Brady, M. (1986). The curvature primal sketch. *IEEE transactions on pattern analysis and machine intelligence*, (1), 2–14.

Barenholtz, E., Cohen, E. H., Feldman, J., & Singh, M. (2003). Detection of change in shape: An advantage for concavities. *Cognition*, *89*(1), 1–9.

Bennett, B. M. [B. M.], & Hoffman, D. (1985). The computation of structure from fixed-axis motion: Nonrigid structures. *Biol. Cybern*, *51*, 293–300.

Bertamini, M., & Farrant, T. (2005). Detection of change in shape and its relation to part structure. *Acta Psychologica*, *120*(1), 35–54.

Blum, H. (1973). Biological shape and visual science (part i). *Journal of theoretical Biology*, *38*(2), 205–287.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial vision*, *10*, 433–436.

Braunstein, M., & Andersen, R. A. (1984). Shape and depth perception from parallel projections of three-dimensional motion. *Journal of Experimental Psychology:Human Perception and Performance*, *10*(6), 749–760.

Cohen, E. H., Barenholtz, E., Singh, M., & Feldman, J. (2005). What change detection tells us about the visual representation of shape. *Journal of Vision*, *5*(4), 3–3.

De Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision research*, *28*(12), 1323–1335.

Di Luca, M., Domini, F., & Caudek, C. (2004). Spatial integration in structure from motion. *Vision Res*, *44*(26), 3001–13. doi:10.1016/j.visres.2004.07.004

Domini, F., & Caudek, C. (2003a). 3-d structure perceived from dynamic information: A new theory. *TRENDS in Cognitive Sciences*, *7*(10), 444–449.

Domini, F., & Caudek, C. (1998). Distortions of depth–order relations and parallelism in structure from motion. *Perception & Psychophysics*, *60*(7), 1164–1174.

Domini, F., & Caudek, C. (1999). Perceived surface slant from deformation of optic flow. *Journal of Experimental Psychology:Human Perception and Performance*, *25*(2), 426–444.

Domini, F., & Caudek, C. (2003b). Recovering slant and angular velocity from a linear velocity field: Modeling and psychophysics. *Vision Res*, *43*(16), 1753–64.

Domini, F., Caudek, C., & Proffitt, D. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology*, *23*(4), 1111–1129.

Domini, F., Vuong, Q., & Caudek, C. (2002). Temporal integration in structure from motion. *Journal of Experimental Psychology:Human Perception and Performance*, *28*(4), 816.

Dosher, B., Landy, M., & Sperling, G. (1989). Ratings of kinetic depth in multidot displays. *Journal of Experimental Psychology:Human Perception and Performance*, *15*(4), 816–825.

Engels, C., Fraundorfer, F., & Nistér, D. (2008). Integration of tracked and recognized features for locally and globally robust structure from motion. In *Visapp (workshop on robot perception)* (pp. 13–22).

Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences*, *103*(47), 18014–18019.

Feldman, J., & Singh, M. (2005). Information along contours and object boundaries. *Psychological review*, *112*(1), 243.

Froyen, V., Feldman, J., & Singh, M. (2013). Rotating columns: Relating structure-from-motion, accretion/deletion. *Journal of vision*, *13*(10), 6.

He, X. (2016). *Structure from motion without projective consistency* (Master's thesis, Rutgers University-Graduate School-New Brunswick).

Henderson, J. M. (1993). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *47*(1), 79.

Hildreth, E., Ando, H., Andersen, R. A., & Treue, S. (1995). Recovery of 3-d structure from motion with surface reconstruction. *Vision Res*, *35*(1), 117–137.

Hoffman, D., & Bennett, B. (1986). The computation of structure from fixed-axis motion: Rigid structures. *Biological Cybernetics*, *54*(2), 71–83.

Hoffman, D., & Richards, W. A. (1984). Parts of recognition. *Cognition*, *18*(1-3), 65–96.

Hoffman, D., & Singh, M. (1997). Salience of visual parts. *Cognition*, *63*(1), 29–78.

Jain, A., & Zaidi, Q. (2011). Discerning nonrigid 3d shapes from motion cues. *Proc Natl Acad Sci U S A*, *108*(4), 1663–8. doi:10.1073/pnas.1016211108

Kayaert, G., & Wagemans, J. (2009). Delayed shape matching benefits from simplicity and symmetry. *Vision research*, *49*(7), 708–717.

Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., Broussard, C., et al. (2007). What's new in psychtoolbox-3. *Perception*, *36*(14), 1.

Koenderink, J. (1984). What does the occluding contour tell us about solid shape? *Perception*, *13*(3), 321–330.

Koenderink, J., & Van Doorn, A. (1986). Depth and shape from differential perspective in the presence of bending deformations. *JOSA A*, *3*(2), 242–249.

Koendreink, J. (1991). Affine structure from motion. *JOSA A*, *8*(2).

Landy, M., Dosher, B., Sperling, G., & Perkins, M. E. (1991). The kinetic depth effect and optic flow–ii first and second order motion. *Vision Res*, *31*(5), 859–876.

Liter, J. C., Braunstein, M. L., & Hoffman, D. (1993). Inferring structure from motion in two-view and multiview displays. *Perception*, *22*(12), 1441–1465.

Liu, Z., Jacobs, D. W., & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision research*, *39*(25), 4244–4257.

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of vision*, *9*(12), 11–11.

Morikawa, K. (1999). Symmetry and elongation of objects influence perceived direction of translational motion. *Perception & psychophysics*, *61*(1), 134–143.

Morita, T., & Kanade, T. (1997). A sequential factorization method for recovering shape and motion from image streams. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(8), 858–867.

Norman, J. F., & Todd, J. T. [James T]. (1995). The perception of 3-d structure from contradictory optical patterns. *Perception & Psychophysics*, *57*(6), 826–834.

Norman, J. F., & Todd, J. T. [James T.]. (1993). The perceptual analysis ofstructure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics*, *53*(3), 279–291.

Paradis, A., Peres, C., J., D., Van de Moortele, P., E.Lobel, Berthoz, A., … J.B., P. (2000). Visual perception of motion and 3-d structure from motion: An fmri study. *Cerebral Cortex*, *10*, 772–783.

Pelli, D. G. (1997). The videotoolbox software for visual psychophysics: Transforming numbers into movies. *Spatial vision*, *10*, 437–442.

Peterson, M. A., & Salvagio, E. (2008). Inhibitory competition in figure-ground perception: Context and convexity. *Journal of Vision*, *8*(16), 4–4.

Prins, N. et al. (2016). *Psychophysics: A practical introduction*. Academic Press.

Prins, N., & Kingdom, F. A. (2018). Applying the model-comparison approach to test specific research hypotheses in psychophysical research using the palamedes toolbox. *Frontiers in psychology*, *9*.

Ramachandran, V., Cobb, S., & Rogers-Ramachandran, D. (1988). Perception of 3-d structure from motion: The role of velocity gradients and segmentation boundaries. *Attention, Perception, & Psychophysics*, *44*(4), 390–393.

Rosin, P. L. (2000). Shape partitioning by convexity. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *30*(2), 202–210.

Schütt, H. H., Harmeling, S., Macke, J. H., & Wichmann, F. A. (2016). Painfree and accurate bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vision Research*, *122*, 105–123.

Siddiqi, K., Tresness, K. J., & Kimia, B. B. (1996). Parts of visual form: Psychophysical aspects. *Perception*, *25*(4), 399–424.

Siegel, R., & Andersen, R. A. (1990). The perception of structure fkorn visual motion in monkey and man. *jounal of cognitive neuroscience*, *2*(4), 306–319.

Sperling, G., Dosher, B. A., & Landy, M. S. (1990). How to study the kinetic depth effect experimentally.

Sperling, G., Landy, M., Dosher, B., & Perkins, 1., M.E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology:Human Perception and Performance*, *15*(4), 826–840.

Tanrıkulu, Ö. D., Froyen, V., Feldman, J., & Singh, M. (2016). Geometric figure–ground cues override standard depth from accretion-deletion. *Journal of vision*, *16*(5), 15–15.

Thompson, W. B., Kersten, D., & Knecht, W. R. (1992). Structure-from-motion based on information at surface boundaries. *Biological cybernetics*, *66*(4), 327–333.

Todd, J. T. [James T.]. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, *36*(2), 97–103.

Todd, J. T. [James T.]. (1982). Visual information about rigid and nonrigid motion: A geometricanalysis. *Journal of Experimental Psychology:Human Perception and Performance*, *8*(2), 238–252.

Todd, J. T. [James T], & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, *48*(5), 419–430.

Tomasi, C., & Kanade, T. (1992). Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, *9*(2), 137–154.

Treue, S., Andersen, R. A., Ando, H., & Hildreth, E. (1995). Structure-from-motion: Perceptual evidence for surface interpolation. *Vision Res*, *35*(1), 139–148.

Treue, S., Husain, M., & Andersen, R. A. (1991). Human perception of structure from motion. *Vision Res*, *31*(1), 59–75.

Tsai, R., & Huang, T. (1984). Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions*, *1*, 13–27.

Tyler, C. W. (1995). Empirical aspects of symmetry perception. *Spatial Vision*, *9*(1), 1–8.

Ullman, S. (1983). Maximizing rigidity: The incremental recovery of 3-d structure from rigid and nonrigid motion. *Perception*, *13*(3), 255–74.

Ullman, S. (1979). The interpretation of structure from motion. *Proceedings of the Royal Society of London B: Biological Sciences*, *203*(1153), 405–426.

Vishwanath, D., & Kowler, E. (2003). Localization of shapes: Eye movements and perception compared. *Vision Research*, *43*(15), 1637–1653.

Vishwanath, D., & Kowler, E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. *Journal of Vision*, *4*(6), 4–4.

Wallach, H., & Oconnell, D. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*(4), 205–217.

Zhang, G., Dong, Z., Jia, J., Wong, T.-T., & Bao, H. (2010). Efficient non-consecutive feature tracking for structure-from-motion. In *European conference on computer vision* (pp. 422–435). Springer.