THE RESONANCE THEORY OF KINETIC SHAPE PERCEPTION: CONSTRAINTS ON PERCEIVING SPATIAL STRUCTURE

VICTOR KLYMENKO

RADIOLGY, UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL

ABSTRACT

Human competence and performance in perceiving three dimensional structure and motion in kinetic displays is analyzed in terms of the resonance theory of kinetic shape perception [26]. The theory, building on Todd's insight on the utility of rotation ellipses [43], and Simpson's insight on the utility of phase angles [41] in recovering structure from motion, is based on three parameters defined with respect to rotation ellipses: amplitude, trajectory separation and phase angle. A tutorial review of a number of visual phenomena, which should be considered when presenting spatial information kinetically, is analyzed within the framework of the resonance theory.

1. INTRODUCTION

Under certain conditions, human observers can voluntarily perceive three dimensional (3D) shapes from two dimensional (2D) image transformations. This capacity, referred to as structure from motion in the computer vision literature, or the kinetic depth effect or kinetic shape perception in visual psychology [13,18,19,27,42,44,45], has been the subject of much theorizing in recent years, both in terms of human competence and performance, as well as in terms of mathematical sufficiency [11]. Understanding the underlying mechanisms of human kinetic shape perception should contribute to the optimization of this source of spatial information in 3D medical imaging and in the visualization of complex data.

Consider the parallel projection of a random dot (virtual) sphere rotating about its vertical axis. Here the moving image elements, the dots, perceptually demarcate a spherical surface rotating in space (2-6). This is a remarkable feat of perceptual organization when one considers that each isolated image element is simply oscillating back and forth along a linear image trajectory that is ambiguous in terms of its location in 3-space. With respect to the third dimension each individual dot could be (a) following a random walk in 3-space, (b) following a linear trajectory in 3-space, (c) following a linear trajectory in the image plane, (d) rotating clockwise in 3-space, or (e) rotating counter-clockwise in 3-space. If the first two possibilities are eliminated on the basis of simplicity [17], and we only consider the last two or three interpretations for each dot as viable candidates, then with N dots, there are still a large number of potential perceptual organizations. Human observers tend to see one of two organizations—a rigid, coherent sphere rotating clockwise or counterclockwise, see [14]. The visual system is highly successful in filtering out all the nonrigid interpretations—the multiple 3D elastic deformations of the image in the image plane and the multiple nonrigid 3D interpretations compatible with this stimulus event, see [39]; although, of course other factors such as luminance [13], occlusion [12] and segmentation boundaries [38] may bias the interpretation. In cases where only nonrigid interpretations are possible, such as a rotating object viewed under polar projection with a large magnitude of perspective, the visual system will select the most rigid one compatible with the stimulus event. The theoretically relevant question is—what principles of organization, what metric [27], and what visual primitives does the visual system use to accomplish this? If there are random positional perturbations in the dots demarcating a rotating virtual object, the shape of the object can still be deciphered; however the object appears as if dots are moving on its surface [26]. Here the visual system appears to parse the transforming image into two motion components, see [11],—global motion specifying shape, and localized object relative motion.

The model presented below can account for some and other data of kinetic shape perception. This theory can account for human competence in kinetic shape perception and for human performance limitations. The model incorporates Simpson's insight on the utility of phase angles [41], and generalizes it in terms of Todd's insights on the utility of rotation ellipses [43], see [11]; recent data and observations are reviewed in light of the model [26].

2. THE RESONANCE THEORY

Consider the image resulting from the parallel projection of a rotating object. Each point on the rotating object traverses a circular orbit, which is projected into an elliptical trajectory, or "rotation ellipse" [22] in the image plane, see Figure 1. Three image parameters—amplitude (A), trajectory separation (S), and phase angle (P)—based on the geometry of these rotation ellipses, specify the 3D shape and motion of the object.

The amplitudes, defined as the image lengths of the major axes of the rotation ellipses, specify the (relative) lengths of the orbital diameters of the object points in space. The trajectory separations, defined as the image distances between the major axes, specify the (relative) distances in space between the orbits. Conversion to 3D units is done by dividing A by sin arcsine(r), where r is the ratio of lengths of the minor over the major axis of any ellipse. Under parallel projection r will be the same for all ellipses. For these two parameters, there is a relatively straightforward correspondence between the distal stimulus in the image and the perception. This model implies that the perception of the relative dimensions of a rigidly rotating object should be predictable on the basis of how well observers scale the object dimensions, which are parallel to the rotation axis, in terms of the dimensions which are perpendicular to the rotation axis. For rotation about an axis parallel to the image plane, a planar axis, the conversion factor is one, so in this case observers should be most accurate in perceiving the relative dimensions of an object.

The remaining degree of freedom is given by the phase angle parameter. The phase angle between any two image points specifies angular separation between the corresponding object points with respect to the rotation axis, see Figure 1. The phase angle (P) between any two image points, i and j, is given by equation (1),

\[ P = \text{arccos} \left( \frac{1}{A} \right) \]

where A is defined as the image distance on the major axis to one side of the imaginary point (open dot), which is closest to the image point. \( A \) is the distance between this imaginary point and one principle vertex of the major axis. For image elements above the major axis, it is (arbitrarily) defined as the distance between the left principle vertex and the imaginary point, and for image elements below the major axis, it is defined as the distance between the right principle vertex and the imagi-
Figure 1. Resonance Parameters. Two points, i and j, on a rotating object, traverse circular orbits in space, which are projected into rotation ellipses in the image plane. Three image parameters based on the geometry of these rotation ellipses, specify the 3D shape and motion of the object. The amplitude (A) defined as the length of the major axis of a rotation ellipse, specifies the relative length of the orbital diameter of an object point in space. Trajectory separation (S), defined as the distance between two major axes, specifies the relative distance in space between the orbits. The phase angle (P), defined in terms of A and S, specifies the angular separation between the two object points in space with respect to the rotation axis [see equation (1) in text]. This is the angular separation between i and j in space when the rotation axis (R) is coincident with the line of sight, that is collapsed to a point, as indicated in the lower right.
The current analysis is limited to rotations, ignoring revolutions and other motions. There are three classes of rotation axis (R) depending on the angle of the rotation axis with respect to the image plane: point axes, where R is parallel to the image plane; plane axes, where R is perpendicular to the image plane; and space axes, which are rotation axes that are neither parallel nor perpendicular to the image plane. Except for illusory stereomagnoc effects, rotation around a point axis will, of course, produce no depth information, since the trajectory separation parameter is lost; however, amplitude and phase angles are still available. Rotation around a space axis will specify the shape and motion of the object, as described above, except for polarity of rotation direction—clockwise or counterclockwise.

Rotation around a planar axis will also produce ambiguity in the definition of phase angles, as given above, because the elliptical image trajectories will be degenerate ellipses (linear trajectories), so there will be no information as to whether the L variables, as given in equation 1, should be defined from the right or left principle vertex for each image element. However, as we shall see, this is not a problem, because of the tendency of the visual system to perceive the simplest configuration compatible with the stimulus event (29), where simplicity here is defined as the minimal variation in the phase angles over time.

Consider Johnson's (29) observations concerning one, two, or three dots traversing a common elliptical image trajectory; one dot presented alone is seen following an elliptical trajectory; however, two or more dots are spontaneously seen as rigidly connected and rotating in depth. This is because, in terms of the resonance theory, the phase angle parameter is unavailable in the single dot display. Consider again the parallel projection of two dots rotating at the same angular velocity about a vertical axis. In the image plane the dots follow a simple linear image trajectory, back and forth between two end points. The stimulus event is spontaneously perceptually organized as two dots rigidly rotating in depth; this is true whether the two dots both rotate at the same constant angular velocity or both at the same irregular angular velocity. In this case the linear image trajectory of each dot is a degenerate rotation ellipse and the only information for perceptually decoding this display configuration in terms of the third dimension is the phase angle parameter. Here the phase angle parameter specifies the third dimension and is a metric of perceptual organization in terms of which the visual system settles on the most coherent and most rigid moving three-dimensional shape compatible with a stimulus event.

The phase angle parameter is a three-dimensional analog to the Gestalt law of common fate. For many stimulus configurations it is not only accounts for why 3D rather than 2D motion is perceived, but also for the particular 2D organization seen in kinetic images.

As in the case with linear motion in the image plane, where the visual system tends to decode motion configurations as described by the Gestalt law of simplicity, or the monotone principle (see 17) and see 21 concerning the debate as to whether the common or relative motion component is minimized, the visual system also tends to perceptually decode an image in terms of the simplest or most coherent, three-dimensional interpretation compatible with the stimulus event. This is accomplished by minimizing the total phase angle variation between image elements in a display. Thus in viewing the parallel projection of a rotating random dot sphere, one would seldom see some of the dots rotating clockwise and others counter-clockwise. The dots will tend to cohere in one of the other perceptual organization, where all the dots have the same rotational parity, all dots rotating clockwise or all rotating counterclockwise, that is the phase angle deviations over time are minimal—In this case zero. This powerful tendency of the visual system to interpret an image in terms of the minimal phase angle variation compatible with the stimulus may account for a large number of perceptual phenomena. Below several phenomena are analyzed within this framework.

6.1. The Pulsing Illusion

The tendency to see perceptual organizations with minimal phase angle variation may account for the pulsing illusion (15,19,24). When two square arrays of dots move with a linear image velocity in opposite directions, so that the two arrays periodically coincide, a temporal illusion occurs during the moment of coincidence. This is due to the point in time all the dots in the display appear to stop moving; they appear to stick or pause before continuing. While the effect is discernible for small numbers of dots, it is most noticeable for arrays with a large number of dots. If these dots were rigidly connected, then the linear image velocities will be perceptually decoded as a variable angular rotation, see Figure 2. The dots departing from one another will correspond to an angular acceleration and the dots approaching will correspond to an angular deceleration. The moment in time when the dots coincide corresponds to the moment when the angular velocity is a minimum—If one assumes that the phase angle between the dots is constant. Thus the apparent pause may be due to the visual system adapting to the lowest phase angle variation—a constant phase angle—compatible with this stimulus. (see 15) for an alternative explanation.

6.2. The Shrinking Cylinder Effect

Recently a demonstration where two coaxial random dot cylinders rotating about the Y axis appear to change their relative dimensions has been reported (30). Although the two coaxial cylinders have identical diameters, when the rotational velocity of one of the cylinders is increased, it is extremely difficult to perceive two identical cylinders of identical diameter spanning at different speeds. Instead one sees a small cylinder inside a large cylinder, where the two cylinders are rotating with identical angular velocity. This may be due to the visual system's strong tendency to minimize the phase angle variations. The tendency may be so strong that the visual system perceptually shrinks the slowly rotating cylinder. Conversely a physically smaller coaxial cylinder spinning at a higher rotational velocity will appear to rotate at the same angular velocity as the larger cylinder (31). Thus the visual system shrinks or expands perceived size in its attempt to minimize phase angle variation (see 36 for an alternative explanation).

6.3. The Rubber Pencil Illusion

The rubber pencil illusion, a well known magician's trick, occurs when one wiggles a pencil that is held loosely off-center between the thumb and index finger (36). The pencil, though rigid, appears to become rubbery and to bend as it moves. Visual persistence has been suggested as the source of the illusion (36). Although intensity, known to affect visual persistence, does not seem to affect it. The illusion occurs when the moving pencil has both a rotational motion component and a translational motion component as shown in Figure 2. The combination of the two motion components results in elliptical motion trajectories. If the visual system interprets these elliptical trajectories as rotation ellipses, then the phase angles registered in terms of these ellipses will continuously vary and therefore will specify a continuously bending non-rigid object.

6.4. Further Considerations

Psychophysical research is needed to determine if the resonance theory is a determining or at least a contributing factor in each of the above phenomena, as there are other plausible explanations. The resonance theory unlike some structure from motion theories (44,45) does not assign a special status to rigid objects. Rigid and non-rigid objects are equally recoverable. The theory predicts three canonical classes of non-rigidity, which can be seen by independently introducing variability into each of the three parameters. Unlike Todd's theory (43), the resonance parameters can recover structure from an object undergoing a variable rotation rate, which is in accordance with human performance. The theory may also account for several perceived effects.
Figure 2. Pausing Illusion. When two rectangular arrays of dots each move along linear image trajectories, with a linear image velocity, in opposite directions, then at the moments in time when the dots coincide, the entire display appears to come to a standstill, to pause. The bottom of the figure shows two of the dots at five points in time, and the top of the figure shows how the two dots are perceptually organized in terms of rotational motion. The moment in time when the dots appear to pause (t4) is the moment when the dots are moving at a minimal rotational velocity (see text). An alternative explanation in terms of the cancellation of velocity vectors in the image plane has been proposed (15).

Figure 3. Rubber Pencil Illusion. When one applies a rotational motion component to a pencil, left, and adds a translational component, middle, illusory bending of the pencil will occur. The result of these two motion components are the elliptical image trajectories shown on the right. If the visual system interprets these as rotation ellipses, then for each set of points on the pencil, the phase angles, as computed by equation (1), will continuously vary over time, as determined by their location on their rotation ellipses. A phase angle between two points for one of the positions of the pencil is indicated by dashed lines.
3. Perspective

The resonance theory implies perceptual mechanisms, or smart devices [37,401], tuned to registering geometric invariants of the rotation ellipse, under parallel projection. If anything in the real world looks in perspective, then why don't all real objects rotating in depth look non-rigid? First, in the real world there is normally an overabundance of converging information available to specify shape. Second, except for unusual situations, such as being very close to a large rotating object, the magnitude of perspective is usually fairly small and so the differences between a parallel and a polar projection are minor. Also, in this case much of the transforming image will be registered in the visual periphery, where resolution is lower. Third, in the absence of other sources of information, while perspective projections of rotating objects look less rigid when directly compared to parallel projections, it's not really the case that they look less rigid than real objects. It may simply be that kinetic parallel projections are hyper-rigid, that is they appear more rigid than real objects. Next the differences between parallel and perspective projections are discussed within the framework of the resonance theory and several perspective effects are analyzed.

3.1 Perspective and Rigidity

Under parallel projection, the rays generating the image of an object are parallel. Under parallel projection, the rays originate at a point a finite distance from the object. As the magnitude of perspective in the image of a rotating object increases, the perceived rigidity of the object decreases. This effect has been replicated by a number of researchers [261]. Computationally changing the magnitude of perspective is accomplished by holding the virtual object a fixed distance from the image plane, and moving the virtual projection point towards the virtual object to increase perspective and away to decrease it. When observers view images of objects viewed from the same distance as the virtual projection point, they will see the object as it would appear in the real world. However, there is no special perceptual status to this point, and in fact the perspective-perceived rigidity remains more or less the same (monotonically decreasing for higher perspective) regardless of the observer's physical observation distance from the display.

Figure 4 shows the parallel and the polar projection of a rotation ellipse. It illustrate how perspective affects the amplitude parameter. Under parallel projection the major axis of the rotation ellipse (the amplitude) projectively correspond to the diameters of the orbits of the object points in space. When perspective is introduced into the image, the orbital diameter is instead represented by a chord of the ellipse parallel to the major axis. The displacement of this diameter chord from the major axis is not monotonic related to the magnitude of perspective in the image. The major axis is a salient feature of an ellipse. Without introducing artificial constraints such as constant angular velocity, there is no salient procedure for locating the diameter chord. If the visual system defaults to interpreting the major axis as the amplitude, then by equation (1), perturbations will be introduced into the phase angles. These perturbations are perceived as specifying a less rigid object. The greater the perspective, the greater the perturbations, and the less rigid the object will appear. This is because the phase angle parameter is geometrically based on the major axis as projected orbital diameter [261]. This explains the findings regarding increasing attenuation of rigidity by increasing perspective; although, it should be noted that the change in relative eccentricity across rotation ellipses introduced by perspective, may play a minor role [43]. Figure 5 illustrates how perspective introduces perturbations into the phase angle between two dots rotating in depth around a common rotation ellipse.

3.2 Rotational Polarity

We now know that in kinetic shape perception the human visual system is not a projective decoder, in the sense that it performs the inverse projective transform to recover the object, as was once thought. In other words, the visual system does not take into account its viewpoint in computing structure. If it did so, then when viewing a display of a rotating object at say, one half or twice the computed distance, the non-rigid deformations in perceived shape would be enormous. Fortunately for the movie industry, all the perceptual information is defined within the image plane and remains invariant with respect to observation distance. If the visual system is not a projective decoder in this sense, then how does the visual system infer rotational polarity? We know that it does because a rotating object viewed under polar projection is perceived as less rigid when a perceptual reversal occurs, where it is seen as rotating in the incorrect direction [11,351]. The visual system might accomplish this by making a binary decision for each rotation ellipse as to which way the diameter chord is displaced. It if assumes that it is displaced in the direction shown in Figure 5, above the major axis, the phase angle perturbations will be attenuated and perceived rigidity will be enhanced—closer to a rigid object. If it makes the wrong assumption, that the diameter chord is below the major axis, the phase angle perturbations will be greater and the object will be perceived as even less rigid. It is a well-established observation that ambiguously rotating objects under polar projection will be seen a higher percentage of the time rotating in the vertical direction. This is another example of the visual system's tendency to minimize phase angle perturbations. The visual system appears to loosely incorporate perspective in kinetic shape perception.

3.3 Ames Trigonal Illusion

The loose incorporation of perspective may account for the Ames illusion, which occurs when an appropriate vertically oriented planar trigonal object rotates about a vertical axis and is viewed from a sufficient distance. It appears as if it were oscillating back and forth, rather than rotating through complete cycles, see Figure 6. The planar rotation axis and the viewing distance cause the rotation ellipses to be indistinguishable from degenerate ellipses or linear trajectories. The visual system assumes a larger degree of perspective than actually present by interpreting the figure as rectangular. For example, if the observer is situated a finite distance from the Ames object, the observer positions the virtual projection point, the Ames illusion would be enhanced—closer to a rigid object. The Ames illusion would occur for complete rotations, which are large rotation ellipses shown in the bottom of the figure. Since the object only traverses the inner portion of these ellipses, it is perceived as oscillating rather than rotating.

3.4 Rotation Axis Effect

Recently a failure of kinetic shape constancy was reported to occur in a polar projection of a random dot sphere rotating in depth, where the observer was situated at the correct viewpoint [30,331]. The sphere appears to increasingly flatten and to lose its depth as the inclination of the space rotation axis increasingly deviates from that of a planar axis. In terms of the resonance theory, the change in apparent shape may be due to the fact that for these displays, the differences in elliptical trajectories between a parallel and a polar projection are greater for greater inclinations of the rotation axis for the spherical object.

3.5 Further Comments

The theory predicts that objects undergoing partial rotations, that is oscillations, under polar projection will look less rigid if most of the image elements are located near the principle vertices of the rotation ellipses, due to the two binary decisions that need to be made the direction of displacement of the diameter chord described in Section 3.2, and whether the viewpoint should be defined with respect to the right or left principle vertex as described in Section 3.2. How these two factors might be related is a question for further research.
Figure 4. Perspective. The circular orbit of two dots projected into a rotation ellipse under parallel projection (top) and under polar projection (bottom). Under parallel projection, a diameter of the circular orbit projectively corresponds to the major axis (A) of the ellipse. Under polar projection, this diameter instead corresponds to a chord (C) of the ellipse. If the phase angle parameter in equation (1) is tuned to interpreting the major axis as the orbital amplitude, then perturbations will be introduced into the phase angle over time producing an attenuation of perceived rigidity.

Figure 5. Phase Angle Perturbations. On the left, the parallel projection of two rigidly connected dots rotating around a rotation ellipse are shown at three points in time. The phase angle (P) between the dots, based on equation (1) remains constant. The major axis, (A), and center of the ellipse, which projectively corresponds to the diameter and center of the circular orbit, are shown. For the polar projection, shown on the right, the diameter of the circular orbit is instead represented by a chord of the ellipse (C). The change in computed phase angle over time, introduced by the displacement of the diameter chord from the major axis, is shown for three points in time. The dashed lines indicate the perceived phase angle based on a parallel projection decoding.
4. Kinetic Shape Perception

4.1 The Motion-Induced Contour

The motion-induced contour is an illusory contour seen in the image of an object rotating in depth (22,24,26,27). It is seen where a dihedral edge would normally be located on a variety of rotating virtual objects under parallel or polar projection, see Figure 7. It is not perceptually salient in static views of the object. The visual system appears to segment the homogeneous image regions within the boundaries of the object into planes oriented in depth. This is analogous to the illusory spherical surfaces demarcated by random dots rotating in depth, except where a spherical surface is a continuous change in orientation, the virtual dihedral edge, the motion-induced contour, is an abrupt change in orientation between two planar surfaces. The visual system appears to interpolate surfaces based on the 3D kinetic shape information of local visible contours. The motion-induced contour is only seen in stimulus conditions in which kinetic shape can be recovered. For example, it is not seen in the looming transformation, where human observers are unable to recover structure from motion.

4.2 Limitations

The image of an object translating in depth undergoes the transformation known as looming. Although, looming sequences exhibit size-distance constancy, human observers are unable to recover the specific 3D shape of an object from these motion sequences. This is because the rotation ellipses upon which the resonance parameters are based are unavailable in looming, and because the human visual system does not appear to contain mechanisms sensitive to kinetic shape as specified by projective invariants, such as the cross ratio (21). In looming (20),

Computer generated images of virtual four dimensional objects rotating in four space under parallel or polar projection appear to be nonrigid 3D objects rotating in depth (31). For parallel projection, the difference between 3D and 4D motion, in terms of the resonance theory, is simply a difference in location of rotation ellipses, which are shifted along the dimension parallel to the major axis so that the minor axes of the ellipses are no longer aligned (details in (31)). The problem of 4D recovery reduces to one of perceptually realigning these axes via an additional source of visual information (31).

4.3 Final Comment

Recent work has investigated perceptual organization of figure-ground perception in terms of visual mechanisms sensitive to different spatial and temporal scales (26,28-30). Current work is investigating the spatial and temporal sensitivity of visual mechanisms to the resonance theory parameters in progress.

Acknowledgements. The writing of this paper was supported by NIH PO1-CA-47982. Thanks to secretary Joy Telle and photographer Bo Strain.

REFERENCES
(1) Proceeding First International Conference on Computer Vision. IEEE Computer Society Press

229