View Dependence of 3D Recovery from Folded Pictures and Warped 3D Faces

Patrick Cavanagh
Department of Psychology
Harvard University
Cambridge, MA 02138
patrick@wjh.harvard.edu

Michael von Grünau
Department of Psychology
Concordia University
Montreal, Quebec, H4B 1R6
vgrunau@vax2.concordia.ca

Lee Zimmerman
Department of Electrical and Computer Engineering
University of Minnesota
Duluth, MN 55812
Leezimmerman2918@aol.com

Abstract

In a popular visual illusion, the portrait on paper currency is folded into an M shape along vertical lines through the nose and the eyes. When this folded picture is tilted back and forth horizontally the face undergoes striking changes in expression. This distortion reveals two insights concerning 3D representation in the human visual system and we have explored these with experiments on simple schematic faces and observations on distortions of laser range images of faces. The observations show first that when recovering depicted depth, pictorial cues are interpreted independently of binocular depth information and second, that the recovery of facial expression is based on a scaled prototypical face structure.

1. Introduction

When we walk around a statue, we see it as a reassuringly solid, rigid object that does not change its 3-dimensional shape as we move. This indicates that our visual system has created a view independent description of the object. This representation maintains constant parameters for the object's 3-dimensional structure as we move and all that is changing is the view angle. Surprisingly, the same is true, over a more restricted range of angles, as we move in front of a picture of an object. It is surprising because the view of the object does not change as it would for the real object; instead, the projection on our retina merely compresses and expands as we move. This should correspond to an object that rotates to face us as we move and changes its 3-dimensional shape nonrigidly, thinning and fattening itself in response to our motions. We ought to find this at least mildly distracting or entertaining in the same way that we are distracted and amused by the reverse motions of a hollow mask that we see in reverse depth [1]. In contrast, we see nothing jarring or askew as we move in front of a picture and it is this relative view independence of our perception of 2D images that has allowed flat pictures and movies to take over our visual environment as an economical and convenient substitute for 3-dimensional representations.

Why are we not overwhelmed by the distortions inherent in the changing views of a 2-dimensional picture? There are two possibilities. The first is the classic theory of compensation where we sense the tilt and slant of the picture plane and correct for it so that we experience the picture as if viewed straight on [2]. There are a number of difficulties with this theory and it has been challenged many times [3] but it is still put forward as the most common explanation of the relative view independence of flat pictures. The alternative explanation is that the internal representation is not metrically 3-dimensional but has some reduced dimensionality that is faithful to the object's structure, for example, up to an affine transformation [4, 5, 6].

2. The folded face illusion

We have explored a popular but little studied visual illusion that supports the second explanation — the non-Euclidian representation — over the compensation explanation. The illusion is seen in the folded picture of a face in Figure 1. At different tilts, it is clear that the facial expression is changing. The images in Figure 1 are of course, flat projections of folded pictures so the cues to the folds are perhaps not very strong. Nevertheless, when the same folded pictures are viewed directly with binocular vision, the expression changes are the same as those seen here [4]. This observation suggests that the slants of the folded picture are not used to correct the picture plane prior to the interpretation of the local pictorial depth cues. If compensation were successful, the corrected picture should be unchanged at different tilts and the recovered 3-dimensional structure of the face would be the same. This does not happen.

Proceedings of the 2nd Intl Symp 3D Data Processing, Visualization, and Transmission (3DPVT 2004)
0-7695-2223-8/04 $20.00 © 2004 IEEE

Figure 1. When a picture of a face is folded along vertical lines through the eyes and nose and tilted around the horizontal axis, striking changes of expression are seen.

The simple description of what does happen is that the retinal projection of the folded picture is the ground image data and the recovery of the 3-dimensional structure is based only on that image, ignoring the folds and their slants. As the folded picture is tilted, the ground image changes solely because of the geometry of the projection and the recovered 3-dimensional shape changes in response. This description does not explain why we tolerate the distortions in flat, unfolded pictures with changing viewing angle. It merely rules out one explanation and leaves us with the possibility that the internal representation, that we accept as richly 3-dimensional, has in fact reduced dimensionality.

One possible criticism of this interpretation of the folded pictures illusion is that the 3-dimensional depth of the folds is simply underestimated. The slants of the folds are not registered and so there is no compensation. In our first measurements we used schematic, wire-frame faces and had observers judge expression and depth with binocular viewing. We were able to show that the depth is seen veridically but the expressions are nevertheless judged as if the figure were almost flat. We will demonstrate in our second experiment that the faces are not judged as if they were actually flat, but rather relative to a prototypical face structure that has only moderate curvature in the mouth region. Whatever the underlying space of the representation, it is important to point out that there are simultaneously at least two: one in which the 3-dimensional information from binocular disparity is available and accurate, and one in which it is largely ignored. A separate 3-dimensional representation is constructed from pictorial cues and object knowledge while ignoring clearly visible binocular information. This dual representation is of course well known in the case of ordinary pictures where the flatness of the picture surface is clearly seen at the same time and at the same location as the 3-dimensional picture space. The only difference here with the folded pictures is that the picture surface now has a more complex 3-dimensional structure.

3. Judgments of depth and expression in wire-frame "faces"

To test judgments of depth and expression we used wire-frame faces that were simply a bent rhombus for a mouth and a pair of dots for eyes. They were presented as elements in 3-dimensional space, viewed binocularly through red-green glasses. Unlike the folded pictures, there was no picture surface and no pictorial depth cues. Nevertheless, the judgments of facial expression still showed view dependence. Specifically, when viewed straight on (0°) the schematic face appeared expressionless. However, when tilted the face took on an expression, smiling when tilted forward and frowning when tilted backward. These expression changes were seen even though the actual wire frame, was tilting rigidly without any change in its 3-dimensional structure. To measure the expression, we asked observers to adjust the corners of the mouth up and down until the face appeared to have a neutral expression at orientations from –40° to +40°. We also asked observers to make a 3D adjustment of the rhombus to determine if the depth of the shape, signaled only by the binocular disparity, was being under or overestimated. Specifically, they adjusted the same back corners up and down until they judged the top and bottom lines to have equal length in 3-dimensions.

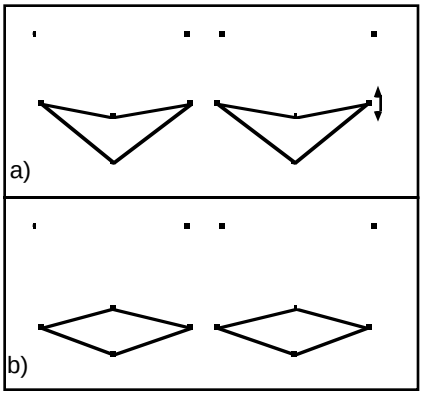

Figure 2. Left and right eye views of the wire-frame face stereogram. Observers adjusted the vertical positions of both outermost tips until the "face" appeared to have a neutral expression. The top face is tilted forward by 40 whereas the bottom, also tilted forward by 40° has the back tips of the mouth moved down to the average "neutral" position.

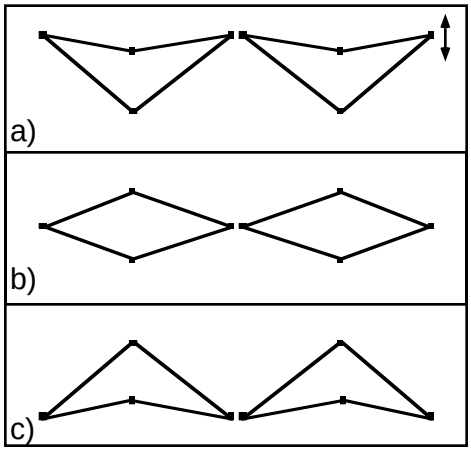

Figure 3. Left and right eye views of the bent rhombus stereogram. Viewed from the top, the rhombus has a 90° bend at the center. a) +40° b) 0°, and c) –40°. Observers adjusted the outer tips up and down until the top and bottom lines appeared of equal 3D length.

3.1 Results

Four observers were run in the two conditions and the average data for 3 of them are shown in Figure 4 (the fourth observer saw no depth from binocular disparity). The data for the 3D judgment of equal line length show good accuracy with settings reflecting an apparent depth of 75 to 90% of the actual value.

In contrast, the settings for neutral expression are not far from the settings that would be made on a 2D image. Observers brought the two tips of the mouth almost level with the midpoint between the top and bottom of the mouth viewed straight on, despite the clearly visible tilt of the top versus the bottom of the mouth and the position of the eyes (which moved forward and backward with the tilt of the mouth). The contrast between the 3D settings and the expression settings could hardly be larger. The 3D judgments of line length demonstrate a very accurate perception of the depth of the wire frame yet the judgments of expression seem to mostly ignore the 3D structure of the mouth.

Is it possible that the visual system has a separate representation that is used for judging expressions, one that is so primitive that it operates directly on the 2D retinal image before any 3D structure is generated? There is some relevant work in studies of the function of the amygdala, a primitive part of the brain linked to the evaluation of emotion. These studies suggest that the amygdala responds to information that is routed to it through subcortical areas bypassing analysis in the visual cortex where binocular disparity could play a role [7]. Emotional responses to fearful faces are registered in the amygdala even though the subjects do not report seeing the faces. The separate representation of face and facial expression is also indicated by neuropsychological findings. In particular, some patients may lose the ability to judge facial expression following neurological damage but retaining face recognition [8]; others may lose face recognition but retaining the ability to make judgments of expression [9].

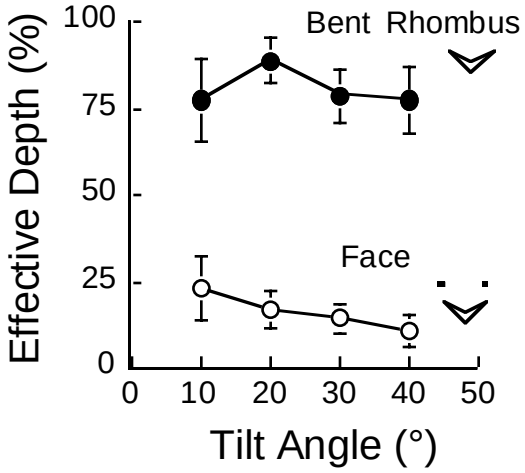

Figure 4. Settings for the neutral face (outline circles) and the equilateral bent rhombus (filled circles) as a percent of appropriate 3D depth. Vertical lines show ± 1.0 SE. The adjustments for the rhombus demonstrate that accurate 3D information is registered by observers. The settings for the neutral face show that this information is largely ignored in determining facial expression.

On the other hand, there is evidence against the simple view that expressions are judged solely on the features of the 2D projection of the face. First, the data for judgment of neutral expression in Figure 4 do not lie around the value for strictly 2D analysis (0% depth). They are somewhat above this, suggesting an effective depth of about 20% of the presented value. Given that our wire-frame mouths had an extreme bend of 90°, 20% of that depth corresponds to a bend of about 22°. This is about half of the value we measured on a sample of 5 human mouths. It is possible that the judgments of emotion are based on a default curvature for human mouths, but the unnaturally sharp central bend of the wire-frame mouth changes the way the default is applied. It is possible then, that the expression judgments are based on a default mouth curvature that overrides binocular disparity information that indicates much more extreme curvature.

A second piece of evidence suggesting that the judgment is not 2D-based comes from observing real

faces. If a person tilts his or her face backwards while holding a neutral expression, the curve of the mouth takes on a decidedly downward arc that ought to correspond to a frown if seen in isolation. However, the person's apparent expression seems unchanged. This could only be the case if the judgment of expression were taking head tilt and mouth curvature into account.

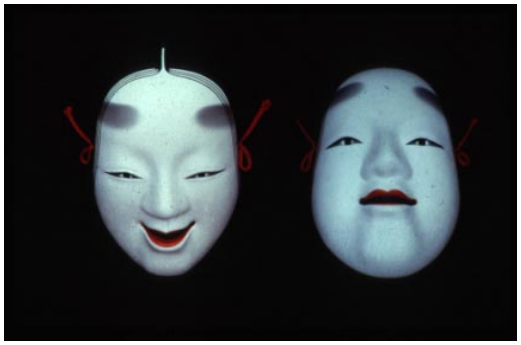

Figure 5. A Noh mask tilted forward and backward. Most observers notice that the mask when viewed binocularly appears to smile when tilted forward but not when tilted backward. The expression changes are similar to those seen in the 2D views here.

Even though a real face does not undergo a change in apparent expression when the person tilts his or her head forward or backward, one particular kind of mask does show this property. Noh masks like that in Figure 5 seem to smile when tilted forward and frown when tilted backward [10]. Interestingly, these tilts of the mask are taken to represent exactly the opposite emotions in the Noh tradition, where, for example, tilting forward indicates sadness. The effects of tilt on the emotions in these masks has been investigated in more detail by Lyon et al. [11]. The property of these masks that is relevant to our argument here is that the curvature of the mouth is very exaggerated even though the curvature of the face is not. We next explored whether the view dependence of the expression resulted from a deviation from prototypical curvature in the region of the mouth.

4. Judgments of expression in distorted 3D faces

We took laser range data for faces, reconstructed the faces and computed shading for lighting from the top right. These shaded values were then used to color the faces as we tilted and distorted them. The face data were in cylindrical coordinates so that for every position, y, along the central axis of the head, the surface of the head was specified as a radial distance function around the central axis. Each surface point was given the luminance value computed from the shading step. To tilt the head, we inclined the vertical axis of the head forward or backward and took two 2D views of the head, offset for a typical binocular view from a standard viewing distance. The head information was clipped on the sides and top.

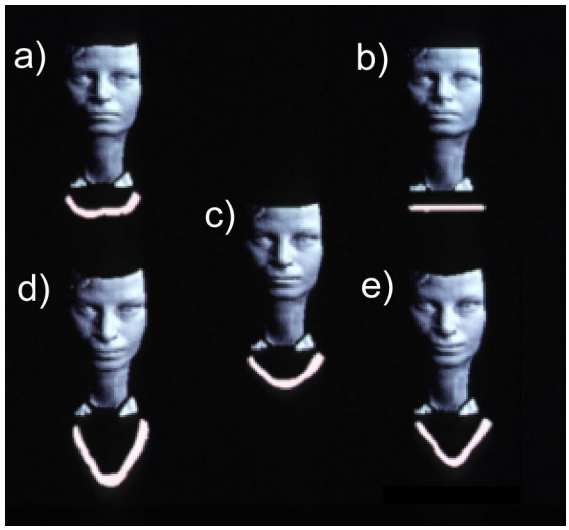

Figure 6. 3D heads with different depth modulations. All are tilted forward 15° and below each face is a profile of curvature across the face at the level of the mouth. a) local compression; b) isotropic compression; c) normal depth; d) isotropic stretching; e) local stretching.

To create the curvature distortions we manipulated the amplitude of the front-back dimension of the head's radial distance function prior to tilting and taking the two binocular views. The left right amplitudes were unchanged; these specify the horizontal separation of face features including the width of the face. These front-back and left-right dimensions of the face surface correspond to the sine and cosine components of the radial distance function. In three isotropic distortions the front-back variations were either flattened (to produce a 2D picture, Figure 6b), left undistorted (Figure 6c), or stretched by a factor of 2 (Figure 6d). In two anisotropic distortions, the front-back variations were either stretched (Figure 6e) or compressed (Figure 6a) locally within the vertical swath that included the chin, the mouth, the nose and the center of the forehead. The factor of distortion changed smoothly from a factor 1.0 (no distortion) just to the left and right of the mouth to a maximum of 1.5 (stretch) or to a minimum of 0.66 (compression) right along the center of the face.

The effect of these anisotropic distortions was to exaggerate the mouth curve in the local stretch condition and flatten the mouth in the local compression version. The curvature of the mouth region is about the same in the isotropic stretch version and the local stretch version. The difference is that in the isotropic case, the highly curved mouth is in the context of a head that is uniformly stretched back to front whereas in the locally stretched version, only the central region of the face has high curvature. Similarly, the mouth curvature is about as flat in the isotropically flat version as in the locally compressed version. In the isotropic version, the whole face is flat whereas in the locally compressed version, the face has normal curvature everywhere but in the flat mouth area (and the vertical band above and below it).

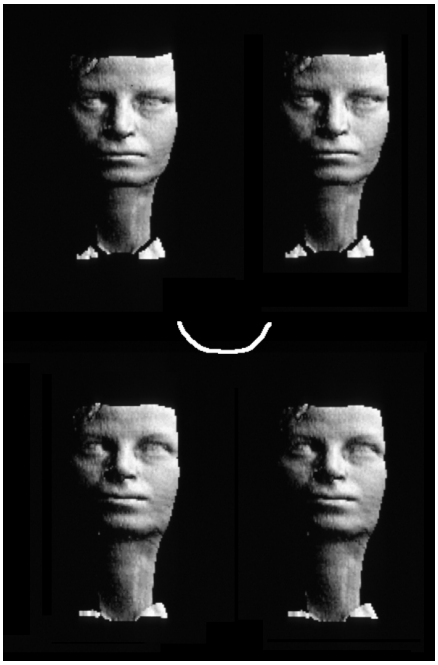

Figure 7. Stereoscopic pairs of locally compressed faces tilted forward 15° on top and backward 15° on the bottom. The left and right views are appropriate for crossed fusion. Most observers report that the bottom face has a more positive expression.

Stereo pairs were generated for each of the 5 versions, one set tilted forward by 15° and the other tilted back by 15°. Two of the authors as well as other lab members viewed the images stereoscopically. None of the isotropic distortions produced any impressions of emotion change between the +15° and –15° heads (see Figure 9 for isotropic stretching). In contrast, the locally stretched and compressed versions did produce an apparent change of expression between the +15° and –15° cases. The shift in emotion was subtle but consistent (see Figures 7 and 8 for local compression).

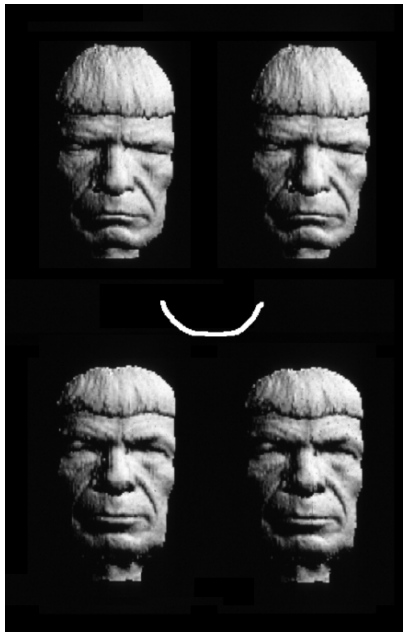

Figure 8. Same as Figure 7 but for a different face.

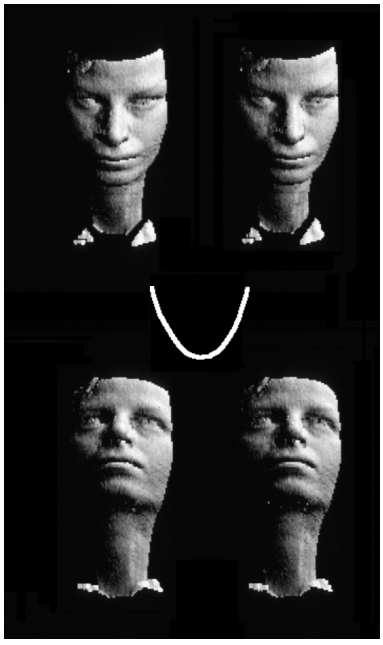

Figure 9. Stereoscopic pairs of isotropically stretched faces tilted forward 15° on top and backward 15° on the bottom. Most observers report that the top and bottom face have similar, neutral expressions.

The lack of effect on apparent emotion for the tilted undistorted heads differs from the report of Lyon et al. [11]. Their observers did find a change of emotion as an undistorted head was tilted at different angles. It is not clear why we do not find this result but the absence of emotional changes seen on real tilted heads in everyday life would appear to support our observations here. Moreover we did not find a change on the isotropically stretched face either.

Note that the degree of mouth curvature that was seen as a slight smile or frown in the locally stretched or compressed faces was the same as that seen as a neutral expression in the isotropically stretched or compressed faces, respectively. In other words, the expression is determined not by the curvature per se but the curvature in relation to the apparent structure of the face. Finally, the apparent shift in emotion was similar whether the faces were viewed in stereo or as 2D images. This suggests that the pictorial cues to depth were sufficient to establish the facial structure and that the binocular information did not contribute.

5. Conclusions

If a folded picture of the front view of a face is tilted, the apparent expression changes dramatically. Our results here suggest that two factors contribute to this illusion. First, pictorial depth cues are interpreted from the 2D (retinal) projection of the folded picture without any correction for any slant of the picture plane. The projections of the folds onto the retina distort the curvature of the mouth (and other facial features) and these distortions become part of the facial expression. Second, whatever the distortion of the facial features, they are interpreted in terms of a prototypical, scaled face that overrides binocular depth information. Strong upward or downward curvature of the mouth that would be seen as smiles or frowns on a typical face may be seen as a neutral expression if the face appears to be strongly stretched in the front-back axis.

The first factor, that pictures are interpreted without any initial compensation for slants of the picture plane, suggests that flat pictures are successful because the human visual system represents space in a non-Euclidian manner. We are tolerant of, though not indifferent to [3], the stretching of the image with changing viewpoint. The idea that there could be a compensation step prior to the interpretation of pictorial cues is an odd one. How could such a process evolve given that we would never have to deal with viewing pictures from an angle until very recently? It is far more likely that our recognition of objects should be tolerant of stretching and compression as these are the main dimensions of variability across members of a category. Some trees are broad and some narrow, some faces wide and some skinny. Whatever the reason for the lack of a Euclidean internal representation of 3D space, the representation we do have is sufficient for workable recognition and navigation. More practically, the tolerance to certain distortions has allowed humans to exploit 2D representations in pictures and movies as convenient surrogates for 3D scenes that can be enjoyed from a wide range of viewing locations.

The second factor, that pictorial cues are interpreted in terms of a prototypical scaled face, is closely related to previous demonstrations of the effectiveness of known shape in determining perceived depth. Specifically, when a particular view of a random 3D shape resembles the 2D view of a familiar object, the shape is most often seen as the familiar object even though that would override clear binocular depth information to the contrary [12]. Similarly for the wire-frame shapes of our first experiment, the depth was registered with reasonable accuracy, but the expressions were judged as if the curvature of the mouth were much closer to the prototypical curvature of a mouth than to the actual curvature. This represented a reduction in the effective depth by a factor of about 400%. Our observations of the distorted 3D heads revealed a further level of sophistication in the use of prototypical object depth. Apparently, the depth that is imposed on the object is scaled to the image data and the scaling involves at least two independent dimensions for heads, front-back and side-to-side.

In cases of familiar objects, then, the representation of accurate 3D information is often redundant. Moreover, when the 3D information disagrees with the prototype, it will not only be ignored, it will also lead to distracting illusions of nonrigidity.

6. Acknowledgements

This work was supported in part by funds from NSERC and NEI to PC. The authors would like to thank Gaile Gordon of the Harvard Robotics Lab for kindly providing the 3D laser scanned head data.

7. References

[1] R. L. Gregory. *The Intelligent Eye*. Weidenfeld & Nicolson, London, 1970.

[2] M. H. Pirenne, *Optics, Painting, and Photography*, Cambridge University Press, Cambridge, UK, 1970.

[3] J. B. Deregowski, and D. M. Parker, On a Changing Peerspective Illusion within Vermeer's *The Music Lesson*, *Perception, 17*, 1988, pp. 13-21.

[4] P. Cavanagh, S. C. Peters, and M. von Grünau, Rigidity Failure and its Effect on the Queen. *Perception 17*, 1988, p. A27.

[5] J. E. Cutting, Rigidity in Cinema Seen from Front Row, Side Aisle, *Journal of Experimental Psychology: Human Perception and Performance, 13*, 1987, pp. 323-334.

[6] J. J. Koenderink, and A. J. van Doorn, What is "Pictorial Relief"?, In H. Hecht, R. Schwartz, and M Atherton, *Looking into Pictures*, MIT Press, Cambridge, MA, 2003, pp. 239-300.

[7] J. S. Winston, P. Vuilleumier, and R. J. Dolan, Effects of Low-Spatial Frequency Components of Fearful Faces on Fusiform Cortex Activity, *Current Biology, 14*, 2003 pp. 1824-1829.

[8] A. W. Young, D. J. Hellawell, C. Van De Wal, and M. Johnson, Facial Expression Processing after Amygdalotomy. *Neuropsychologia, 34,* 1996 pp. 31-39.

[9] R. Bruyer, C. Laterre, X. Seron, P. Feyereisen, E. Strypstein, E. Pierrard, and D. Rectem, A Case of Prosopagnosia with Some Preserved Covert Remembrance of Familiar Faces, *Brain and Cognition, 2,* 1983 pp. 257-284.

[10] P. Cavanagh, and M. von Grünau, 3-D Objects That Appear Nonrigid During Rotation. *Investigative Ophthalmology and Visual Science Supplement, 30,* 1989, p. 263.

[11] M. J. Lyon, R. Campbell, A. Plante, M. Coleman, M. Kamachi, and S. Akamatsu. The Noh Mask Effect: Vertical Viewpoint Dependence of Facial Expression Perception. *Proceedings of the Royal Society, London B Biological Science, 267,* 2000, pp. 2239-2245.

[12] P. Sinha, and T. Poggio, Role of Learning in Three-Dimensional Form Perception. *Nature, 384,* 1996 pp. 460-463.