Chapter 28
Schema-Driven Influences in Recovering 3-D Shape from Motion in Human and Computer Vision

Thomas V. Papathomas and Doug DeCarlo

28.1 Introduction

One of the fundamental questions in vision is how the visual system recovers a nearly veridical representation of the world, given that the retinal optic flow has an infinite number of possible interpretations, especially if one considers that there are moving objects in the environment. This one-to-many mapping from retinal optic flow onto the real-world surfaces and objects that provide the stimulation is known as the inverse problem in optics [35, 40, 45–47].

There is a long-standing debate in vision on whether visual perception, which is based on the solution to the inverse problem in optics, is influenced by schema-driven processes or it is entirely stimulus-driven and automatic. Helmholtz [26] was among the first researchers to hypothesize that perception is a process that involves "unconscious inference" at a time when it was commonly believed that perception was a purely data-driven process. This view has been adopted and extended by more recent formulations based on experimental evidence [11, 20, 22–24, 42, 52]. This view is also adopted by researchers who use a Bayesian formulation [8, 15, 19, 34, 71]; according to this view, the visual system uses "priors", such as the convexity bias [39, 60] or the "light-from-above" assumption [1, 6, 7, 36], to arrive at the most probable interpretation of the visual input, given the ambiguity of the solution to the inverse problem in optics. Proponents of this view posit that, in addition to the "bottom-up" processing that starts with data-driven sensory signals and activates progressively higher brain areas, there are also schema-driven "top-down" cognitive influences (such as experience, memory, suggestions, knowledge,

T.V. Papathomas (✉)
Department of Biomedical Engineering and Center for Cognitive Science, Rutgers University, Piscataway, NJ, USA
e-mail: papathom@rci.rutgers.edu

D. DeCarlo
Department of Computer Science and Center for Cognitive Science, Rutgers University, Piscataway, NJ, USA

S.J. Dickinson, Z. Pizlo (eds.), *Shape Perception in Human and Computer Vision*, Advances in Computer Vision and Pattern Recognition,
DOI 10.1007/978-1-4471-5195-1_28, © Springer-Verlag London 2013

etc.) that start at higher brain areas that "interpret" their input and modulate the activation of lower brain areas ([3, 17, 30, 32, 41]; but see [18] for an argument against top-down influences).

One approach to studying the interaction of bottom-up (data-driven) and top-down (prior-knowledge-driven) processes is to select stimuli in which these two processes compete against each other, such that the percept elicited by the bottom-up signals is quite different from—and often opposite from—the percept that is favored by the top-down processes. Specifically, the value of visual illusions in this effort has long been recognized [5, 22–24, 52, 69]. Illusions have been used extensively to study normal brain mechanisms and stages of processing [2, 16, 20, 38, 49]. In particular, there is a fascinating class of three-dimensional (3-D) stimuli in which the data-driven cues elicit one depth percept while the schema-driven processes elicit a strong depth-inversion illusion. Two members of this class that produce very reliable illusions are the hollow mask [21, 27–29, 43, 70] and the reverse perspective [9, 42, 43, 55, 61, 67]. In this chapter, we will consider the role of top-down influences on the hollow-mask illusion both for human and machine vision. We observe that a face-tracking algorithm that recovers the 3-D shape from animation sequences of moving faces is susceptible to the hollow-mask illusion just as humans are, when it incorporates a top-down schema of convex faces, even though the data-driven motion parallax signals are adequate to recover the veridical concave 3-D shape. We discuss the implications of this observation.

28.2 The Hollow-Mask Illusion for Humans

The hollow-mask illusion, along with reverse perspectives, is one of the best-known depth inversion illusions, where one can distinguish cues and processes that give rise to competing percepts. In the case of the hollow mask, the prior knowledge of faces being convex, based on life-long exposure to faces, is the only schema-driven influence *in favor of* the illusion. It would be instructive to summarize briefly the basic data-driven influences that provide cues for the true depth structure, *against* the illusion, with the exception of the kinetic depth effect cue that provides ambiguous information (see item 2b.2 below). (1) There are two main extraretinal signals: (1a) *Vergence* is a binocular signal: it refers to the simultaneous but opposite-directed movement of the two eyes to achieve fixation of both on the point of interest. Since it can be expressed as an angle (vergence angle) that is formed by the two eyes' lines of sight, it is a single-valued function. (1b) *Lens accommodation* is self-explanatory monocular signal: the shape of the eye lenses has to vary in order to achieve the proper optical power, also single valued, to obtain a sharp "image" on the retina. Theoretically, at least, if viewers had access to the motor signals that control the muscles affecting vergence and accommodation, they could have used them as cues to depth, provided they were derived over time, as they fixate various points on the object/surface of interest. In practice, there is a long-standing debate on whether such motor signals are indeed used as cues to depth [4, 50, 68]. (2) We next move to some of the retinal-based cues: (2a) The most important binocular signal is the *stereoscopic disparity* between the two-eyes' "images", which comprises both horizontal and vertical components; disparity provides continuous 3-D shape cues that recover the true depth ordering of a scene or object; however, disparity needs to be processed further to yield true accurate depth information [51, 64]. (2b) Some of the monocular cues that are most relevant to the hollow-face illusion are: (2b.1) *Motion parallax*, due to the observer's self-motion as he/she views a scene, produces an optic flow field that also provides continuous 3-D shape cues with properties that are similar to those of stereopsis; they recover the correct depth ordering but they need to be scaled for recovering true depth [53, 54]. (2b.2) For a stationary observer, a moving object provides depth-from-motion cues, the so-called *kinetic depth effect* or KDE [31, 62]. KDE also provides continuous 3-D shape cues, but the depth ordering is ambiguous; as an example, a rotating wire-frame globe can be perceived either veridically, rotating in the physical direction, or in reverse depth, rotating in the opposite direction. (2b.3) *Occlusion* is a powerful cue to depth but it only provides depth-ordering information. (2b.4) *Shading* can also provide cues to the 3-D shape of an object [48, 65]. (2b.5) Finally, *image blur* can be used for assessing depth relationships in a scene [25, 66]. Blur is closely related to lens accommodation because only the point that the viewer fixates on and its close surroundings are in sharp focus, whereas more distant points are blurred. The degree of blur can be used to estimate the depth differential between a blurred image point and the fixation point, but it does not inform us of the depth polarity (is it in front or behind fixation). In addition, there are other depth cues in the general case (texture gradient, atmospheric perspective, size familiarity, etc.) that may not be strongly relevant in the hollow mask illusion. Nevertheless, the plethora of depth cues makes the problem of recovering depth too complex for a thorough scientific analysis. Here, we lump together all the depth cues that provide good estimates of depth (items 1a, 1b, 2a, 2b.1, 2b.3) and observe that they are more powerful at small viewing distances.

Thus, when observers view a hollow mask up close, the bottom-up signals of stereopsis, motion parallax, vergence eye movements and lens accommodation, among others, provide powerful signals that dominate and enable viewers to recover the veridical concave 3-D shape of the mask. These signals, however, become weaker as the viewing distance increases. At an adequately long viewing distance, the top-down influences—familiarity with convex faces in the case of a facial mask—dominate, causing depth relationships to be inverted; points that are physically further away appear to be closer than points that are physically closer. Consequently, concavities appear as convexities and vice versa, resulting in the percept of an overall convex face.

Figures 28.1a and 28.1b illustrate an essential feature of the hollow-mask illusion. Namely, even though the hollow mask of Fig. 28.1b faces to the left, the (mis-)perceived convex mask appears to face to the right (more details on this are provided later in reference to Fig. 28.2). This gives rise to two related motion illusions: (1) When a viewer moves laterally in front of a static hollow mask, the perceived convex mask appears to turn and "follow" the viewer. An explanation of

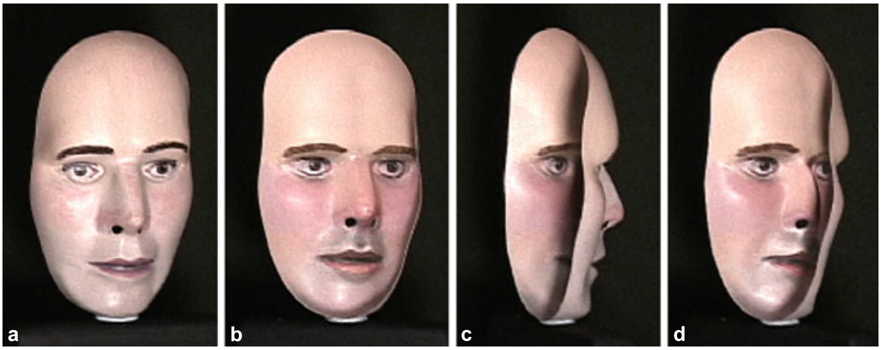

Fig. 28.1 (**a**) A convex mask that faces to the right of the viewer ($\theta = 22.5°$ in the notation of Fig. 28.2). (**b**) A concave mask that actually faces to the left ($\theta = 157.5°$) but it appears to be a convex mask facing to the right. (**c**) A mask with significant self-occlusion ($\theta = 117.3°$). (**d**) A mask that involves a marginal self-occlusion ($\theta = 135°$). See also Fig. 28.2

this illusory motion that is based on the depth inversion has been proposed by Papathomas [42]. (2) When the hollow mask is rotated in front of a stationary viewer, the perceived convex mask appears to rotate in a direction opposite to the physical direction of rotation. An extension of the explanation by Papathomas can be applied in this case. In this paper, we will concentrate on the second type of illusory motion that is elicited by a rotating mask for a stationary observer.

Figure 28.2a is a top view that illustrates the notation we use to describe the orientation of the mask. Angle θ specifies the spatial orientation of the mask, starting from zero when the convex side of the mask faces straight ahead toward the viewer, shown with solid lines, and increasing in the counter-clockwise (CCW) direction; a mask is shown in dashed lines as it rotates CCW by an angle θ_1. This mask orientation is very similar to the one used to obtain the image in Fig. 28.1a. In Fig. 28.2b the mask starts at $\theta = 180° - \theta_1$ (dashed lines) and it rotates CCW by an angle θ_1 to the straight ahead concave position at $\theta = 180°$ (solid lines). The viewer sees the concave side of the mask in this case. The image in Fig. 28.1b was obtained using a mask orientation similar to that at $\theta = 180° - \theta_1$. Notice that, if we think of the dashed-line hollow mask ($\theta = 180°$) as pointing toward the viewer, then the solid-line hollow mask of Fig. 28.2b points to the left of the viewer by an angle $\theta = -\theta_1$.

Importantly, as masks in Figs. 28.2a and 28.2b illustrate, for small rotation angles that avoid self-occlusions on the mask, and under orthographic projection, the image of the concave mask at $\theta = 180° - \theta_1$ is very similar to that of the convex mask at $\theta = \theta_1$. For example, the sizes of the left and right eyes of the masks will be roughly equal under orthographic projection. In contrast, for a perspective projection—which, after all, is what one obtains on the retina or with a camera—there are ample cues, for small viewing distances, to distinguish between the images obtained for the masks at $\theta = 180° - \theta_1$ and at $\theta = \theta_1$; this ability to distinguish between the two images decreases with increasing viewing distance.

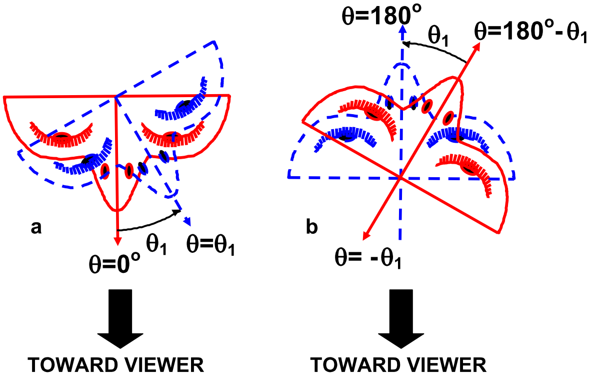

Fig. 28.2 Notation for the mask orientation. In these top views the viewer is at the bottom of the figure. For each mask orientation, the *straight-ahead solid arrow*—NOT the *thick arrow*—extending outward from the convex side is used to indicate its spatial orientation. The orientation angle θ is measured counter-clockwise from the reference position ($\theta = 0$), in which the convex mask faces the viewer. (**a**) The convex mask rotates CCW by an angle θ_1 from $\theta = 0$ (*solid lines*) to $\theta = \theta_1$ (*dashed lines*). (**b**) The mask rotates CCW by an angle θ_1 from $\theta = 180° - \theta_1$ (*solid lines*) to $\theta = 180°$ (*dashed lines*); both of these masks show their concave side to the viewer. The *dashed-line* mask in part a and the *solid-line* mask in part b were used to obtain the images of Fig. 28.1a and 28.1b, respectively

For example, for the dashed-line mask in Fig. 28.2a ($\theta = \theta_1$), the eye to the *left* of the viewer, being closer than the eye to the right of the viewer, will form a larger image. The opposite will be true for the solid-line mask in Fig. 28.2b ($\theta = 180° - \theta_1$); the eye to the *right* of the viewer will form a larger image than the eye to the left of the viewer. Of course, the size difference depends on the viewing distance of the imaging device (retina or camera) from the mask. This size difference between the left and the right eyes extends to the entire left and right sides of the face and the resulting size gradient can theoretically be used to recover the true 3-D shape. The images of Figs. 28.1a and 28.1b were obtained from a relatively large distance and, hence, these size differences are not evident.

The essence of the illusion is obtained when one compares what is perceived in the following two cases: (a) The convex masks starts from a straight-ahead position ($\theta = 0°$) and moves CCW by an angle θ_1 to position $\theta = \theta_1$, as in Fig. 28.2a. (b) The concave masks starts from a straight-ahead position ($\theta = 180°$) and moves CW by an angle θ_1, to position $\theta = 180 - \theta_1$, in the opposite direction to that shown in Fig. 28.2a. Under viewing conditions that favor the illusion, these two motions will produce the same percept, namely a convex masks that rotates CCW, because the concave mask will appear to rotate in the opposite direction to that of its physical direction of rotation. Notice that, because the kinetic-depth-effect cue (item 2b.2 in this section) is ambiguous, when we perceive the concave mask in inverted depth (convex), we perceive it rotating in the opposite direction.

The question is: will this size-differential cue, as well as other bottom-up cues (motion parallax, stereoscopic disparity, blur and possibly vergence angle and accommodation,[1] among others) overcome the schema of a convex face to recover the true concave mask shape? The answer is: it depends on several factors but primarily on the viewing distance. As explained above, the size gradient is negligible for large viewing distances and increases with decreasing viewing distance. The same is true for the differential signals provided by most of the bottom-up cues we mentioned earlier. Namely, the binocular disparity differential signals provided by mask features that are at different depths, such as the tip of the nose and lips, are very weak at long viewing distances and grow stronger as the distance decreases; ditto for motion parallax signals, and differences in vergence and accommodation. Because the strength of these bottom-up signals diminishes with increasing viewing distance, the prior experience with convex faces dominates and thus the prediction is that the illusion strength will increase with increasing distance. This is precisely what has been observed in experimental studies [21, 27–29, 43, 70].

28.3 The Hollow-Mask Illusion and Computer Vision

Most computer vision algorithms that have been developed to recover the 3-D structure of human faces include the schema for the convex form of faces as part of their knowledge base. Naturally, one would expect this schema to influence the recovery of 3-D shape when such algorithms are provided an animation sequence that involves a hollow mask. This is what we consider below for a representative face-tracking algorithm.

28.3.1 Model and Algorithm

The particular 3-D face model and tracking algorithm we used is that of DeCarlo and Metaxas [10]. The model itself is a handcrafted 3-D polygon model which has motion parameters that describe head movements (3-D translation and rotation) and facial motions (mouth movements, eyebrow raises, etc.), and shape parameters that enable the model to approximate the geometry of an individual's face. See Fig. 28.3. The 3-D face model uses about 80 spatial geometry variables (distance between eyes, length of nose, distance between upper lip and tip of nose, width of lips, etc.) that the algorithm adjusts to obtain a physical 3-D surface that conforms best to the face features that are present in the animation sequence being processed. The algorithm uses a combination of optical flow and feature alignment in order to maintain track of moving subjects. Essentially, the 3-D model, along with a simple model of image formation, is used to explain the changing appearance of a face in a series of images, in terms of its parameters.

[1] As Christopher Tyler commented, "vergence and accommodation would have to be derived over time by eye movements to provide shape information."

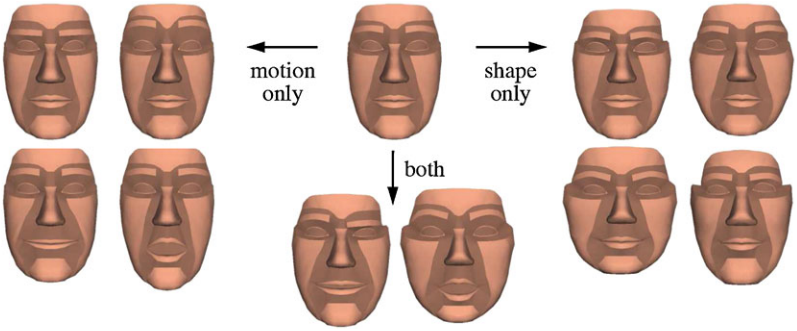

Fig. 28.3 The deformable face model from DeCarlo and Metaxas [10] has separate parameters that describe the static shape of the face and dynamic parameters that describe face and head motion

In Fig. 28.3, the light regions of the shape are the explicitly parameterized parts of the shape, e.g., a hand-crafted 3D deformable model of the lips. The dark regions connect the vertices of the deformable models together using triangles (and no new vertices), e.g., the space between the lips and nose is "filled in". These dark regions are only used for modeling occlusion and are use for predicting the locations of occluding contours. The shape and motion models are formulated the same way, except that the shape parameters are static quantities, and the motion parameters are time-varying. For instance, one particular shape parameter describes the width of the lips. One particular motion parameter is the horizontal translation of the entire head.

We developed two versions of the tracking framework, which differed only in the schema used to interpret the input images: (1) The original algorithm used the assumption that human faces are convex. It tracked faces and, naturally, the 3D face model implicitly constrained geometry to be convex overall, with depth undulations that are typical of human faces. (2) To explore the interaction of top-down and bottom-up processes on the behavior of the algorithm, the second version of the algorithm did not use the assumption that faces are convex. We allowed it to accept concave faces, as well as convex faces, and to conduct feature tracking based on this new schema. In some sense, this second version enables the algorithm to recognize a thin mask that has both a convex and a concave side, like Halloween masks.

28.3.2 Algorithm Input—Results

The input to the algorithm was a video animation of a mask that was painted realistically on both sides, as shown in Fig. 28.1. The concave side was painted to give the impression of a convex face. The mask rotated by 360° about a vertical axis in the CCW direction. The animation can be viewed in http://videos.springer.com. Examples of tracking results are shown in Fig. 28.4.

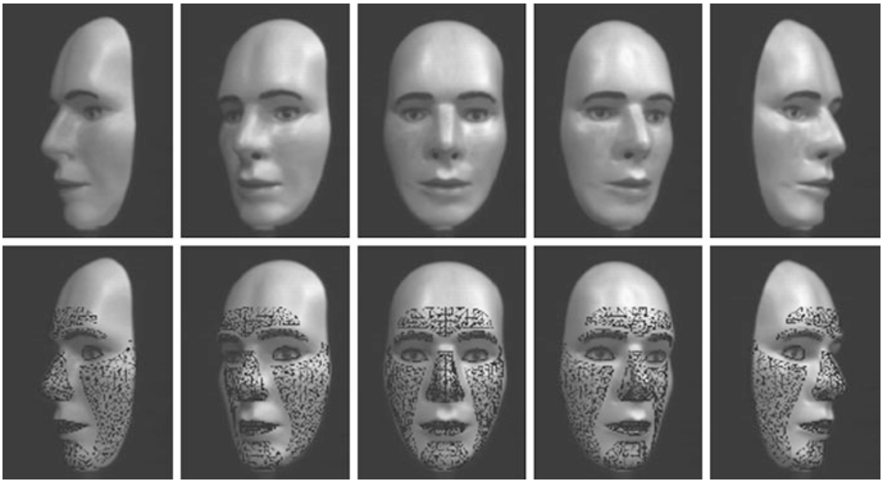

Fig. 28.4 Tracking results of the algorithm with a convex face schema (model 1). On *top* is the original sequence, and on the *bottom* the 3D model is superimposed on top of the images. The superimposed *grid lines* indicate the results of the 3D model

The results of the original version are graphed in Fig. 28.5a. The horizontal axis represents the orientation θ of the physical mask that varies from $-180°$ (concave part facing viewer) through $-90°$ (mask facing left), 0° (convex part facing viewer), 90° (mask facing right), all the way to 180° (concave part facing viewer). As θ varies continuously along the horizontal axis, the mask can be thought of as rotating CCW. The vertical axis represents the orientation that the algorithm recovers, based on the visual input, as modulated by the algorithm's schema.

If the algorithm recovers the true orientation of the mask, we would expect a line with a slope of 1. On the contrary, if the algorithm is susceptible to the hollow-mask illusion then, as Fig. 28.2 illustrates, a hollow mask at an orientation $\theta = 180° - \theta_1$ will be perceived as a convex mask at $\theta = \theta_1$; thus, as the physical concave mask rotates CCW from $\theta = 180° - \theta_1$ to $\theta = 180°$, the perceived convex mask will be seen rotating CW from $\theta = \theta_1$ to $\theta = 0°$, producing a line with a slope of -1 on the graphs of Fig. 28.5.

This is precisely what the results of the algorithm indicate: The algorithm recovers the orientation of the convex mask perfectly for $-90° \leq \theta \leq 90°$. However, in the approximate range $132° < |\theta| \leq 180°$, the data indicate that the algorithm "experiences" the hollow-mask illusion as evidenced by the slope of -1; see Fig. 28.6. Interestingly, but predictably, the algorithm does not recover a face at all in the approximate range $90° < |\theta| < 132°$. This is the range for which there is a significant extent of self-occlusions, as illustrated in Fig. 28.1c. The marginal orientation of 132° is shown in Fig. 28.1d, for which the algorithm barely recovered a face.

The results of the modified algorithm that recognizes the existence of both convex and concave faces are shown in Fig. 28.5b. As expected, the rich optic flow signals, without the constraint of a convex face, allow the algorithm to recover the true 3-D

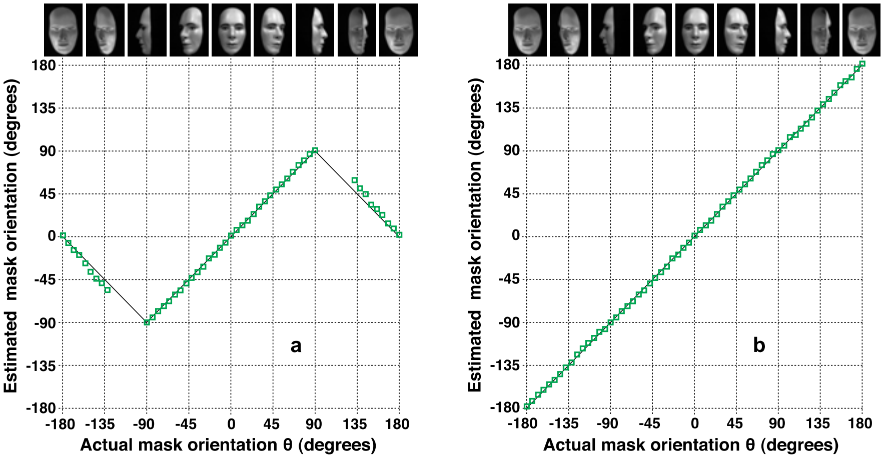

Fig. 28.5 The results of the algorithm under two different 3-D schemata, as applied to the animation sequence. The *horizontal axis* denotes the actual mask orientation of the stimuli during the rotation; *icons at the top* display the mask view for the corresponding orientation (the stimuli are identical in the two cases). The *vertical axis* denotes the mask orientation estimated by the algorithm. (**a**) Results with model 1 that assumes a convex mask schema. The algorithm tracks the mask well when the convex part is visible ($-90° \leq \theta \leq 90°$). However, the estimate is in the opposite direction from the actual motion (it approximately has slope -1), when the concave part is visible without any self-occlusions ($135° \leq |\theta| \leq 180°$). The algorithm experiences tracking problems and is unable to explain the image when parts of the mask occlude other parts (roughly in the range $90° < |\theta| < 135°$). (**b**) Results with model 2, a schema that accepts both convex and concave faces. The estimated mask orientation approximately matches the actual mask orientation for all 360 degrees (it has slope 1)

shape and spatial orientation of the mask over the entire range of 360°. Indeed, the slope of the line is 1 for $-180° \leq \theta \leq 180°$.

28.4 Discussion

In some sense, these findings can be thought of as evidence for the existence of schema-driven influences in visual perception for the particular case of human faces. Liberated from these top-down influences—in this case the knowledge that faces are convex—the algorithm uses the bottom-up signals to recover exactly the true shape and orientation of facial masks. However, when the algorithm is imbued with the knowledge of faces being convex, it behaves just as humans in being “fooled” by the hollow-mask illusion. As Theo Pavlidis [44] commented, “I have not seen any other demonstration where machine vision algorithms also suffer from optical illusions.”

There are at least two pieces of evidence for top-down influences. The first is the inversion effect, that is, the reduced strength of the illusion when the hollow mask is displayed upside-down [27, 28, 43]. Apparently, the cause for the reduced strength

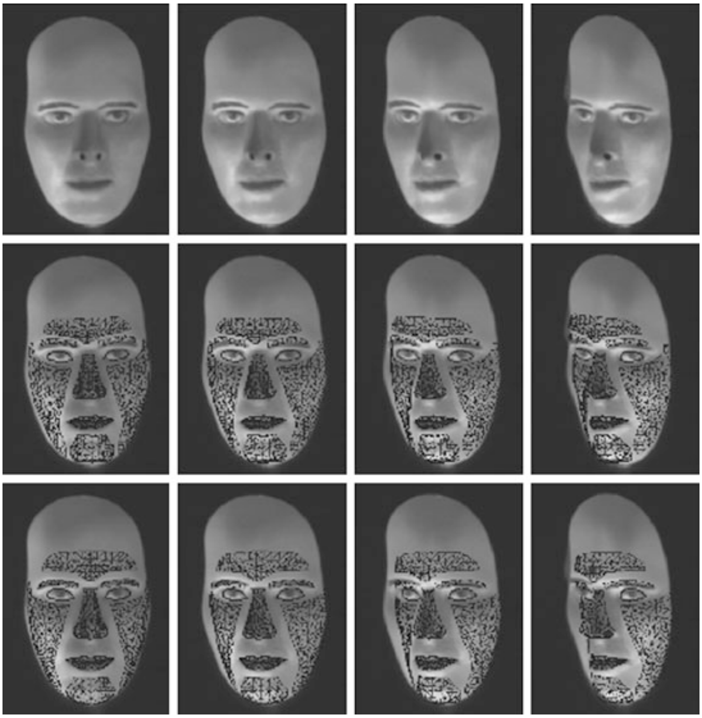

Fig. 28.6 Tracking results for the concave side of the mask. On *top* is the original sequence. In the *middle* are tracking results using model 2 (the convex-concave 3D model), which estimates the mask position correctly; it recovers a hollow mask rotating CW. On the *bottom* are tracking results using model 1 (the convex-face 3D model), which has the rotational direction reversed; it recovers a convex mask rotating CCW

is the lack of familiarity with inverted faces. The second piece of evidence is the reduced strength of the hollow-mask illusion in cases where the cognitive influences are impaired. Examples of such impairment are observed with subjects who are sleep-deprived [63] or under the influence of cannabis [13, 14, 56] or alcohol [58]. Schizophrenia (SZ) patients also experience a weaker illusion than controls, i.e., they tend to perceive the hollow mask as hollow [13, 33, 37, 57, 59]. One possible explanation is that SZ patients have a reduced ability to exert top-down influences in perception. Evidence for such weak feedback connections in SZ patients was presented by Dima et al. [11], based on fMRI data; Dima et al. [12] presented additional evidence on the basis of reduced P300 and P600 electro-encephalography (EEG) components, which signal late-stage processing, in SZ patients.

In addition to the explanation that the illusion is based on face familiarity, another possibility is that the illusion may not be stimulus-specific but instead may owe to a bias in favor of convexity [39, 60]. This hypothesis was tested by Hill and Bruce [28]. They reported that a hollow human mask produces a much stronger illusion than a "hollow potato" by comparing the switching distance, namely the average of the viewing distance at which the illusion breaks down on approach—starting with the illusory percept from a long distance—and the viewing distance at which the illusion sets in on retreat—starting with the veridical percept at a very close distance. Thus, the human mask enhances the bias to see concave surfaces as convex.

A third possibility, raised by Barlow [5], is that the sensory mechanisms analyze the redundancy that exists inherently in the sensory signals and the associations between input variables. According to Barlow, neural mechanisms respond to the "established associative structure in the input messages by recoding them … [thus] making new structure more easily detectable" [5]. In the case of faces, he argues that our vast prior experiences resulting from moving past normal faces—as they themselves move or remain stationary—have resulted in an efficient neural coding scheme for the representation of our own motion and the visual motion signals generated by the 3D geometry and the parts of a normal face. Thus, when we move past a stationary hollow mask, this recoding scheme now works in the wrong direction to compensate for the visual motion that it expects. The result is that the scheme interprets the visual motion signals as elicited by a normal (convex) face that moves. The weaker illusion for an upside-down hollow mask is accounted by Barlow's [5] hypothesis as resulting from less exposure to upside-down faces and therefore less recoding mechanisms.

What happens for $90^\circ < |\theta| < 132^\circ$? Figure 28.1c shows a mask oriented at $\theta = 117.3^\circ$. The algorithm, as well as the human visual system, is unable to recognize a normal face in this and other cases where there are extensive self-occlusions. Apparently, the algorithm that has a convex-face schema cannot resolve the discontinuities in the 3-D shape that are recovered by the optic flow and provides no output for a face. When the extent of the self-occlusions is limited, as in the marginal orientation of Fig. 28.1d ($\theta = 132^\circ$), the algorithm can still—barely—recover a face. The same occlusion that acts as noise for tracking a convex face in the convex-face schema algorithm can be used as signal in the thin-mask schema algorithm.

At this point, we can conjecture about how a computer vision algorithm for scene perception would operate when presented with an animation sequence of a rotating reverspective. A reverspective is a 3-D piece that is constructed and painted realistically such that the painted perspective cues depict a depth structure that is exactly opposite to the physical structure, which is recovered correctly by binocular disparity and motion parallax signals [61, 67]. A computer vision scene analysis system that is driven mainly by data-driven signals (motion parallax, stereoscopic disparity, etc.) would recover the veridical 3-D structure. We conjecture that, if such a scene analysis system is endowed with schema-driven modules for perspective, it will likely experience the illusion that human experience, that is, perceiving the direction of rotation to be the opposite from the physical direction, in analogy with the behavior of the face tracking algorithm (see pertinent video animation in http://videos.springer.com).

This raises an interesting issue. Ideally, computer vision systems need to be endowed with some schemata that help them process the visual input more efficiently and accurately. For example, face-processing systems benefit greatly from a built-in deformable model of a generic face that is characterized by several deformation parameters. This allows the systems, when provided the input video sequence of a particular face, to optimize the parameters so as to fit the best possible model to that particular face. The price they pay is that such systems are fooled by the hollow-mask illusion, as humans are.

Humans, however, have the advantage of changing the viewing conditions (shorter viewing distance, binocular viewing, etc.), not to mention handling the stimulus by touch, to gain more knowledge about the true 3-D structure of the stimulus. We can see two extreme options for computer vision systems. The first—easy—approach involves endowing such systems with additional schemata, such as the schema of a thin facial mask in the case of face perception. The second—difficult—approach is to enable these systems to explore possibilities that are not covered by the initial repertory of schemata they were provided with at inception. In the case of face tracking algorithms, the system would have to analyze the optic flow in the video, recognize that it fails to arrive at a solution in some instances and search for alternative schemata that, if successful, it would add it to its own repertoire. The first approach is one of "spoon feeding" schemata to the system and it requires constant supervised learning; the second approach is one of "learning from experience" and it appears more promising but requires fundamental advances before it can be implemented. Of course, these two approaches are applicable to cases where we are limited to images derived from the visible spectrum only. Obviously, systems equipped with range finders can provide 3D signals to recover the true depth structure.

Acknowledgements We would like to thank Manpreet Kaur for conducting experiments on the hollow-mask illusion with human observers. We thank Christopher Tyler who reviewed the manuscript and offered valuable suggestions.

References

1. Adams WJ, Graf EW, Ernst MO (2004) Experience can change the 'light-from-above' prior. Nat Neurosci 7:1057–1058
2. Aglioti S, DeSouza JFX, Goodale MA (1995) Size contrast illusions deceive the eye but not the hand. Curr Biol 5:679–685
3. Ahissar M, Hochstein S (2004) The reverse hierarchy theory of perceptual learning. Trends Cogn Sci 8(10):457–464
4. Banks MS, Backus BT (1998) Extra-retinal and perspective cues cause the small range of the induced effect. Vis Res 38:187–194
5. Barlow HB (1997) The knowledge used in vision and where it comes from. Philos Trans R Soc Lond B, Biol Sci 352(1358):1141–1147
6. Berbaum K, Bever T, Chung CS (1983) Light source position in the perception of object shape. Perception 12:411–416
7. Berbaum K, Bever T, Chung CS (1984) Extending the perception of shape from known to unknown shading. Perception 13:479–488
8. Caudek C, Fantoni C, Domini F (2011) Bayesian modeling of perceived surface slant from actively-generated and passively-observed optic flow. PLoS ONE 6(4):e18731

9. Cook ND, Hayashi T, Amemiya T, Suzuki K, Leumann L (2002) Effects of visual field inversions on the reverse-perspective illusion. Perception 31:1147–1151
10. DeCarlo D, Metaxas D (2000) Optical flow constraints on deformable models with applications to face tracking. Int J Comput Vis 38(2):99–127
11. Dima D, Roiser JP, Dietrich DE, Bonnemann C, Lanfermann H, Emrich HM, Dillo W (2009) Understanding why patients with schizophrenia do not perceive the hollow-mask illusion using dynamic causal modeling. NeuroImage 46:1180–1186
12. Dima D, Dillo W, Bonnemann C, Emrich HM, Dietrich DE (2011) Reduced P300 and P600 amplitude in the hollow-mask illusion in patients with schizophrenia. Psychiatry Res Neuroimaging 191:145–151
13. Emrich HM, Leweke FM, Schneider U (1997) Towards a cannabinoid hypothesis of schizophrenia: cognitive impairments due to a dysregulation of the endogenous cannabinoid system. Pharmacol Biochem Behav 56:803–807
14. Emrich HM, Weber MM, Wendl A, Zihl J, Von Meyer L, Hanisch W (1991) Reduced binocular depth inversion as an indicator of cannabis-induced censorship impairment. Pharmacol Biochem Behav 40:689–690
15. Feldman J, Singh M (2006) Bayesian estimation of the shape skeleton. Proc Natl Acad Sci USA 103(47):18014–18019
16. Franz VH, Gegenfurtner KR, Bulthoff HH, Fahle M (2000) Grasping visual illusions: no evidence for a dissociation between perception and action. Psychol Sci 11:20–25
17. Gilbert CD, Sigman M (2007) Brain states: top-down influences in sensory processing. Neuron 54:677–696
18. Goldberg I, Harel M, Malach R (2006) When the brain loses its self: prefrontal inactivation during sensorimotor processing. Neuron 50:329–339
19. Goldreich D, Peterson MA (2012) A Bayesian observer replicates convexity context effects in figure-ground perception. Seeing Perceiving 25(3–4):365–395. doi:10.1163/187847612X634445
20. Gregory RL (1968) Perceptual illusions and brain models. Proc R Soc B 171:279–296
21. Gregory RL (1970) The intelligent eye. McGraw-Hill, New York, pp 126–131
22. Gregory RL (1980) Perceptions as hypotheses. Philos Trans R Soc Lond B, Biol Sci 290:181–197
23. Gregory RL (1997) Knowledge in perception and illusion. Philos Trans R Soc Lond B, Biol Sci 352:1121–1128
24. Gregory RL (2005) The Medawar lecture 2001 knowledge for vision: vision for knowledge. Philos Trans R Soc Lond B, Biol Sci 360(1458):1231–1251
25. Held RT, Cooper EA, Banks MS (2012) Blur and disparity are complementary cues to depth. Curr Biol 22:426–431
26. Helmholtz H (1910/1867). Handbuch der Physiologischen Optik, vol 3. Voss
27. Hill H, Bruce V (1993) Independent effects of lighting, orientation, and stereopsis on the hollow-face illusion. Perception 22(8):887–897
28. Hill H, Bruce V (1994) A comparison between the hollow-face and ‘hollow-potato’ illusions. Perception 23:1335–1337
29. Hill H, Johnston A (2007) The hollow-face illusion: object-specific knowledge, general assumptions or properties of the stimulus? Perception 36:199–223
30. Hochstein S, Ahissar M (2002) View from the top: hierarchies and reverse hierarchies in the visual system. Neuron 36:791–804
31. Jain A, Zaidi Q (2011) Discerning nonrigid 3D shapes from motion cues. Proc Natl Acad Sci USA 108(4):1663–1668
32. Jones MJ, Sinha P, Vetter T, Poggio T (1997) Top-down learning of low-level vision tasks. Curr Biol 7:991–994
33. Keane BP, Silverstein SM, Wang Y, Papathomas TV (2013) Reduced depth illusions in schizophrenia: Evidence for a weakened, state-dependent convexity prior. J Abnormal Psych 122(2):506–512

34. Kersten D, Yuille A (2003) Bayesian models of object perception. Curr Opin Neurobiol 13:1–9
35. Kersten D, Mamassian P, Yuille A (2004) Object perception as Bayesian inference. Annu Rev Psychol 55:271–304
36. Kleffner DA, Ramachandran VS (1992) On the perception of shape from shading. Percept Psychophys 52:18–36
37. Koethe D, Gerth CW, Neatby MA, Haensel A, Thies M, Schneider U, Emrich HM, Klosterkotter J, Schultze-Lutter F, Leweke FM (2006) Disturbances of visual information processing in early states of psychosis and experimental delta-9-tetrahydrocannabinol altered states of consciousness. Schizophr Res 88:142–150
38. Krekelberg B, Lappe M (2001) Neuronal latencies and the position of moving objects. Trends Neurosci 24:335–339
39. Langer MS, Bülthoff HH (2001) A prior for global convexity in local shape-from-shading. Perception 30:403–410
40. Palmer S (1999) Vision science: from photons to phenomenology. MIT Press, Cambridge
41. Papathomas TV (1999) The brain as a hypothesis-constructing-and-testing agent. In: LePore E, Pylyshyn Z (eds) What is cognitive science? Blackwell, Oxford, pp 230–247
42. Papathomas TV (2007) Art pieces that 'move' in our minds—an explanation of illusory motion based on depth reversal. Spat Vis 21:79–95
43. Papathomas TV, Bono L (2004) Experiments with a hollow mask and a reverspective: top-down influences in the inversion effect for 3-d stimuli. Perception 33:1129–1138
44. Pavlidis T (2012) Personal communication. December 15, 2012
45. Pizlo Z (2001) Perception viewed as an inverse problem. Vis Res 41(24):3145–3161
46. Poggio T, Torre V, Koch C (1985) Computational vision and regularization theory. Nature 317:314–319
47. Purves D, Lotto RB (2003) Why we see what we do. Sinauer, Sunderland
48. Ramachandran VS (1988) Perception of shape from shading. Nature 331:163–166
49. Ramachandran VS (1995) Anosognosia in parietal lobe syndrome. Conscious Cogn 4(1):22–51
50. Regan D, Erkelens CJ, Collewijn H (1986) Necessary conditions for the perception of motion in depth. Investig Ophthalmol Vis Sci 27:584–597
51. Richards W (2009) Configuration stereopsis: a new look at the depth-disparity relation. Spat Vis 22(1):91–103
52. Rock I (1983) The logic of perception. MIT Press, Cambridge
53. Rogers B (2009) Motion parallax as an independent cue for depth perception: a retrospective. Perception 38(6):907–911
54. Rogers B, Graham M (1979) Motion parallax as an independent cue for depth perception. Perception 8(2):125–134
55. Rogers BJ, Gyani A (2010) Binocular disparities, motion parallax, and geometric perspective in Patrick Hughes's 'reverspectives': theoretical analysis and empirical findings. Perception 39:330–348
56. Semple DM, Ramsden F, McIntosh AM (2003) Reduced binocular depth inversion in regular cannabis users. Pharmacol Biochem Behav 75(4):789–793
57. Schneider U, Borsutzky M, Seifert J, Leweke FM, Huber TJ, Rollnik JD, Emrich HM (2002) Reduced binocular depth inversion in schizophrenic patients. Schizophr Res 53:101–108
58. Schneider U, Dietrich DE, Sternemann U, Seeland I, Gielsdorf D, Huber TJ, Becker H, Emrich HM (1998) Reduced binocular depth inversion in patients with alcoholism. Alcohol Alcoholism 33:168–172
59. Schneider U, Leweke FM, Sternemann U, Weber MM, Emrich HM (1996) Visual 3D illusion: a systems-theoretical approach to psychosis. Eur Arch Psychiatry Clin Neurosci 246:256–260
60. Sherman A, Papathomas TV, Jain A, Keane BP (2011) The roles of perspective, angle polarity, stereo and motion parallax in perceiving 3D objects. Seeing Perceiving 25:263–285
61. Slyce J (2011) Patrick Hughes: perverspective, 3rd edn. Momentum, London

62. Sperling G, Landy MS (1989) Kinetic depth effect and identification of shape. J Exp Psychol Hum Percept Perform 15(4):826–840
63. Sternemann U, Schneider U, Leweke FM, Bevilacqua CM, Dietrich DE, Emrich HM (1997) Pro-psychotic change of binocular depth inversion by sleep deprivation. Nervenarzt 68(7):593–596. [Article in German]
64. Tyler CW, Kontsevich LL (1995) Mechanisms of stereoscopic processing: stereoattention and surface perception in depth reconstruction. Perception 24(2):127–153
65. Uttal WR, Liu N, Kalki J (1996) An integrated computational model of three-dimensional vision. Spat Vis 9(4):393–422
66. Vishwanath D, Blaser E (2010) Retinal blur and the perception of egocentric distance. J Vis 10(10):26
67. Wade NJ, Hughes P (1999) Fooling the eyes: trompe l'oeil and reverse perspective. Perception 28:1115–1119
68. Welchman AE, Harris JM, Brenner E (2009) Extra-retinal signals support the estimation of 3D motion. Vis Res 49:782–789
69. Westheimer G (2008) Illusions in the spatial sense of the eye: geometrical-optical illusions and the neural representation of space. Vis Res 48(20):2128–2142
70. Yellott JI Jr (1981) Binocular depth inversion. Sci Am 245:118–125
71. Yuille A, Kersten D (2006) Vision as Bayesian inference: analysis by synthesis? Trends Cogn Sci 10:301–308