Stereomotion is processed by the third-order motion system: reply to comment on "Three-systems theory of human visual motion perception: review and update"

Zhong-Lin Lu

Laboratory of Brain Processes, Department of Psychology and Neuroscience Graduate Program, University of Southern California, Los Angeles, California 90089-1061

George Sperling

Human Information Processing Laboratory, Departments of Cognitive Sciences and Neurobiology and Behavior, and the Institute of Mathematical Behavioral Sciences, University of California, Irvine, Irvine, California 92697

Received April 23, 2002; revised manuscript received April 29, 2002; accepted May 2, 2002

Two theories are considered to account for the perception of motion of depth-defined objects in random-dot stereograms (stereomotion). In the Lu–Sperling three-motion-systems theory [J. Opt. Soc. Am. A **18**, 2331 (2001)], stereomotion is perceived by the third-order motion system, which detects the motion of areas defined as figure (versus ground) in a salience map. Alternatively, in his comment [J. Opt. Soc. Am. A **19**, 2142 (2002)], Patterson proposes a low-level motion-energy system dedicated to stereo depth. The critical difference between these theories is the preprocessing (figure–ground based on depth and other cues versus simply stereo depth) rather than the motion-detection algorithm itself (because the motion-extraction algorithm for third-order motion is undetermined). Furthermore, the ability of observers to perceive motion in alternating feature displays in which stereo depth alternates with other features such as texture orientation indicates that the third-order motion system can perceive stereomotion. This reduces the stereomotion question to "Is it third-order alone or third-order plus dedicated depth-motion processing?" Two new experiments intended to support the dedicated depth-motion processing theory are shown here to be perfectly accounted for by third-order motion, as are many older experiments that have previously been shown to be consistent with third-order motion. Cyclopean and rivalry images are shown to be a likely confound in stereomotion studies, rivalry motion being as strong as stereomotion. The phase dependence of superimposed same-direction stereomotion stimuli, rivalry stimuli, and isoluminant color stimuli indicates that these stimuli are processed in the same (third-order) motion system. The phase-dependence paradigm [Lu and Sperling, Vision Res. **35**, 2697 (1995)] ultimately can resolve the question of which types of signals share a single motion detector. All the evidence accumulated so far is consistent with the three-motion-systems theory.

OCIS code: 330.4150.

1. INTRODUCTION

In 1995, Lu and Sperling[1] proposed that the human perception of the direction of visual motion is served by three separate motion systems: a first-order system that responds to moving luminance patterns; a second-order system that responds to moving modulations of *feature activity*, e.g., stimuli in which an area of higher contrast or of flicker moves; and a third-order system that computes the motion of locations marked as figure (versus ground) in a "salience map" of visual space. Based on the temporal-frequency-tuning function and the susceptibility to static pedestals, Lu and Sperling concluded that the motion of depth-defined objects in dynamic random-dot stereograms (stereomotion) is processed by the third-order motion system.[1]

Recently, Lu and Sperling[2] reviewed the new evidence generated subsequent to the publication of the original three-systems theory. Various challenges were resolved, yielding a more clearly defined and significantly strengthened theory. In particular, they considered the evidence for Patterson's assertion that "stereoscopic motion is processed by a motion-sensing system composed of special-purpose mechanisms that function like low-level motion sensors" (Ref. 3, p. 3329). They found that all the evidence considered by Patterson admits of alternative interpretations. They also pointed out that rivalry motion[4] is a common confound in stereomotion experiments. They concluded that in stereomotion, "the typical third-order temporal tuning function, the ease with which motion standstill can be produced, and the lack of evidence to the contrary, suggest that stereomotion is perceived only by the third-order motion system" (Ref. 2, p. 2353).

In a comment, Patterson again asserts that "stereomotion is processed by a low level [non-feature-tracking] mechanism" (Ref. 5, p. 2143). He supports his assertion by reiterating some "old" evidence and advancing two lines of "new" evidence.[6,7] He also asserts that "stereomotion has not been confounded by rivalry" (Ref. 5, p. 2143).

Here, we first clarify the differences between a low-level (non-feature-tracking) mechanism, tracking mechanisms, and third-order motion. We show what psychophysical paradigms can and cannot discriminate between these mechanisms. We demonstrate the following:

1. Patterson's interpretation of his own experiments reflects an incorrect intuition about motion processing systems—in fact, third-order motion easily accounts for the results of the cited experiments.
2. Computational analysis of the two lines of "new" evidence cited in Patterson's comment shows that they are, in fact, fully consistent with the three-motion-systems theory.
3. Rivalry motion can and—unless great care is taken to avoid it—usually does occur as a confound in stereomotion experiments, at both small and large disparities.

2. SOME CLARIFICATIONS OF THE THREE-MOTION-SYSTEMS THEORY

A. Three-Motion Systems: Preprocessing versus Motion Processing

The three-motion-systems theory explicitly specifies first-, second-, and third-order motion computations. At the point of motion extraction, the first-order and the second-order systems use a similar algorithm that can be modeled either as a Reichardt detector[8] or, equivalently, as a motion-energy detector.[9] Probably, the third-order system also uses a motion-energy detector, but this has not been established.

What differentiates the three systems is the computation prior to motion extraction. All three systems are preceded by stages of light adaptation, division into spatial-frequency channels, and contrast gain control. At this point, the first-order system computes motion energy in each spatial-frequency channel. The inputs to first-order motion are represented as positive and negative values relative to the mean luminance. The second-order system computes motion direction from the output of texture grabbers (spatial–temporal channel filters followed by full-wave rectification). Whereas the input to first-order motion is point contrast directly, the input to second-order motion is approximately the variance of point contrast.[10] The third-order system computes motion direction from modulations of stimulus salience, i.e., changes of location of visual areas marked as "important" or as "figure" versus a background that is marked as "ground."[11,12] The different properties of the first- and second-order motion systems stem almost entirely from different preprocessing prior to motion extraction. Preprocessing is also the main factor that distinguishes the third-order motion system from first- and second-order motion systems. Additionally, the parameters of third-order motion extraction may differ from those of first- and second-order motion extraction.[2]

B. The Third-Order Motion Computation Is Not Feature Tracking

1. Feature Tracking

When our eyes track an object moving in a smooth, predictable trajectory, there may be very little retinal slip. Yet, despite the almost stationary retinal image, we are perfectly aware of the true motion because we are aware of the eye movement. When we are viewing some third-order-motion stimuli, attentionally tracking a physical feature as it moves produces a sensation analogous to tracking by eye movements.[13] This is feature tracking.

2. Motion Standstill

When the third-order system was originally proposed, the distinction between a third-order motion computation and feature tracking was still unclear to us.[1] Soon afterward, we developed a procedure to make physically moving stimuli appear to be standing still.[14] Standstill occurs with slowly or rapidly moving stimuli when the stimulus is contrived so that all three motion systems fail but texture, shape, and color systems continue to function. The phenomenon of motion standstill brought about a realization that feature tracking typically requires a prior motion computation. Without a functioning motion system, when a feature moves within a range that is of the same order of size as the feature itself, the feature is represented by the shape, color, and texture systems as an invariant feature at an invariant location. At some point, when it has moved sufficiently, it is re-represented. But there is no sensation of motion.

Motion standstill demonstrates that the shape, form, and color systems are designed to extract an invariant representation of a (stationary) object from the haphazard involuntary and voluntary movements of the eye. To know that an object has moved (in order to track it) requires a motion signal. Without a functional motion system, one can perform feature search but not smooth feature tracking.

3. Third-Order Motion

The third-order system automatically extracts motion from the spatiotemporal modulation of salience (figure–ground). An explicit computational model of salience extraction and third-order motion (as it applies to the perceived movement of color and texture stimuli and the effects of attention on these perceived movements) was published in 1999.[12]

4. Can an Attentional Feature Search Produce a Sensation of Motion?

On the basis of the phenomenon of motion standstill, we propose that without a prior motion computation, one can perform a feature search but not feature tracking (because without a motion computation, the feature would seem to remain in the same place for an extended period and then reappear at a new place). In such a case, can attentional feature search (without eye movements) produce a sensation of feature motion when no third-order motion is computed? We would guess not, but the question has not been answered.

3. EVIDENCE PROPOSED IN FAVOR OF A DEDICATED MOTION-ENERGY COMPUTATION FOR STEREOMOTION

In this section we reconsider the evidence that Patterson[3,5] adduced in favor of a low-level stereomotion mechanism. We first summarize our previous analysis of the evidence cited by Patterson in Ref. 3. We then analyze the "new" evidence cited by Patterson in Ref. 5.

A. The Distinction between Motion-Energy and Third-Order Motion Theories of Stereomotion Perception

1. *Preprocessing versus Motion Detection*

There are two issues: (1) the question of the nature of the motion computation itself and (2) the nature of the preprocessing. The third-order motion theory proposes that motion is computed on the salience map where figure–ground is represented. Thus all the factors that determine figure–ground segregation act prior to and are combined prior to the third-order motion computation [Fig. 1(a)]. Stereo is an important factor in determining figure–ground. Typically in a stereo scene, the foreground is perceived as figure and the background is perceived as ground, although this is not an absolute rule. Small, distinctive patches on large homogeneous surround typically are perceived as figure. This applies to stereo images, whether the small patches are in front of (crossed disparity) or behind (uncrossed disparity) the background. The third-order motion theory is neutral as to whether third-order motion is processed by a motion-energy detector or some other kind of motion detector. Because the intensity resolution of the salience map seems to be rather coarse, it has not been possible to make a definitive determination of the third-order motion algorithm.

The motion-energy theory of stereo motion holds that there is a low-level motion-energy computation dedicated to stereomotion, i.e., a low-level motion-energy system that is unshared with other systems [Fig. 1(b)]. This is similar to arguments that have been made in favor of a low-level mechanism for perceiving isoluminant chromatic motion[15,16]; similar arguments and counterarguments apply in both the color and the depth domains.

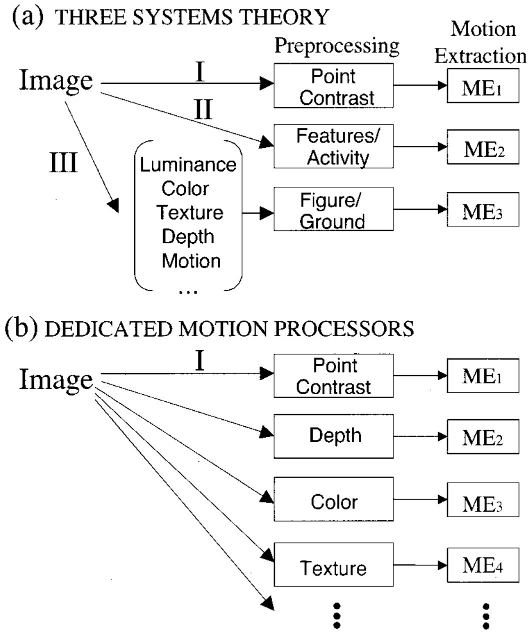

Fig. 1. Comparison of (a) the three-systems theory and (b) the dedicated-motion-processors theory of motion perception. In both theories, "Image" represents the visual input after it has been processed by light adaptation, spatial-frequency channels, and contrast gain control. ME represents a motion-energy detector (or, equivalently, a Reichardt detector); the subscripts indicate separate detectors.

2. *Perceiving Motion in Alternating-Feature Displays Requires a Third-Order Motion System*

Even if there were to be a low-level motion-energy system dedicated to stereomotion, there would still have to be a third-order motion system. This is because observers reliably detect the motion of alternating-feature displays in which the features that distinguish an object on even frames are completely unrelated to those that distinguish the object on odd frames.[11,17] For example, Lu and Sperling describe an apparent-motion display in which a stereo-defined grating on odd frames alternates with a texture-defined grating on even frames.[11] Neither a texture motion system nor a low-level stereo-depth motion system could detect motion in such an alternating-feature display. But a wide variety of alternating-feature displays produce good apparent motion.

It would be useless to have specialized motion detectors for every pair of features (e.g., depth and texture, depth and color, color and texture, depth and luminance, etc.) Not only is such a scheme of pairwise detectors extremely inefficient, but in natural scenes it would falsely detect motion between pairs of unrelated features.

Because stereo depth is a useful feature in alternating-feature displays, it means that the third-order motion system that detects alternating-feature motion can also detect stereomotion. Therefore, the issue raised here is not whether stereomotion is detected by the third-order motion system but whether, in addition to the detection of stereomotion by the third-order system, there is a dedicated low-level stereo-depth motion system.

B. Three "Old" Findings Alleged to Support a Low-Level Stereomotion System

In a minireview of stereomotion perception, Patterson[3] cited three lines of evidence to support his assertion of a low-level stereomotion system: (1) cross adaptation between stereomotion and luminance motion, (2) a difference between speed and position discrimination in stereo displays,[18] and (3) that failure of stereomotion to pass a pedestal test is not conclusive evidence against a specialized stereomotion mechanism. All these were considered in our review,[2] so we only summarize them briefly here.

1. *Cross Adaptation*

Cross adaptation between stereomotion and other forms of motion does not prove that they share a common motion-extraction mechanism. Motion signals that are extracted by different mechanisms share a subsequent common motion path; the common path is affected by the cross adaptation. Indeed, cross adaptation is the rule between first-order and second-order motion systems even though these two systems can be demonstrated to have independent extraction mechanisms.[19,20] With more elaborate paradigms that cancel the common motion component, it is possible to selectively adapt first- and second-order motion mechanisms.[19,20] Indeed, it is not the success but the failure of cross adaptation that establishes different mechanisms. So if one could selectively directionally adapt isoluminant chromatic motion and stereomotion, then that would suggest that they had dedicated motion-extraction mechanisms [a violation of the diagram represented in Fig. 1(a)]. Such selective adaptation would be evidence against a common third-order motion-extraction mechanism. So far, to our knowledge, this kind of selective adaptation has not been reported.

2. *Speed versus Position Discrimination*

Patterson[5] reports an experiment[18] in which subjects discriminate the speed of a stereo display but cannot discriminate the direction of translation when the first and last frames of the motion sequence are separated by a blank interval of 500 ms. Patterson *et al.*[18] take their observers' failure to discriminate the direction of translation as evidence that there were no features in the display that might have stimulated a feature-extraction or third-order mechanism. On the contrary, we take it as evidence that there was an abundance of features, and that an observer's memory for complex displays is remarkably poor over an interval of 500 ms (as first demonstrated by Phillips[21] and substantiated by a recent outpouring of "change blindness" experiments).

3. *Failure of Pedestal Test*

In the pedestal test, a moving sine grating is superimposed on a stationary sine grating of the same spatial frequency and twice the amplitude (the pedestal). Under appropriate conditions,[2] for first- and second-order motion stimuli, motion-direction thresholds are the same whether the pedestal is present or not. This remarkable property is predicted by Reichardt and motion-energy models.[2,8] For all third-order stimuli studied so far, motion direction is totally destroyed by the pedestal ("totally failing" the pedestal test). We agree with Patterson that while failing the pedestal test has heretofore been a reliable indicator of third-order motion, it is not by itself a proof of third-order motion.

Passing the pedestal test is more informative than failing, because failure can mean either that (i) the motion-extraction algorithm is not a motion-energy algorithm or (ii) the signal has been so distorted (e.g., by gain control and/or by coarse intensity coding) prior to the motion computation that the pedestal test is invalid. For example, high-contrast first-order stimuli fail or partially fail the pedestal test because of the distortion introduced by contrast gain control prior to the first-order motion computation. The pedestal test holds exactly only for undistorted—in this case, low-contrast—luminance sine-wave gratings and texture-contrast sine-wave gratings.

Isoluminant red–green gratings totally fail the pedestal test even at near-threshold color contrasts. This means that the third-order motion system is very different from the first-order or second-order motion systems. Total pedestal test failure could be due to a severely nonlinear transformation (e.g., a binary representation of figure versus ground) even for small, near-threshold modulations or to the third-order motion-extraction algorithm's not being a motion-energy algorithm. This issue is unresolved.

C. Detecting Movement of a Missing-Fundamental Depth Grating

The missing-fundamental grating (MFG) paradigm has been advocated for distinguishing motion energy versus feature tracking in first-order[22,23] and second-order[24] motion. As shown in Fig. 2, an MFG is constructed by subtracting the fundamental sine-wave component from a square-wave grating [subtracting the dashed curves from the solid curves in Fig. 2(a) results in the curves in Fig. 2(b)]. For an MFG moving 90 deg between successive frames, the features (peaks and valleys) move in the forward direction. Thus a feature-tracking motion algorithm is expected to report motion in the forward direction. However, a motion-energy system that decomposes the MFG waveform into various Fourier components might report forward, reversed, and/or transparent motion. This is because, in a 90-deg phase-stepping MFG, all the $4*k - 1$ harmonics move in the reversed direction and all the $4*k + 1$ harmonics move in the forward direction ($k = 1, 2, \ldots$). The remaining harmonics are neu-

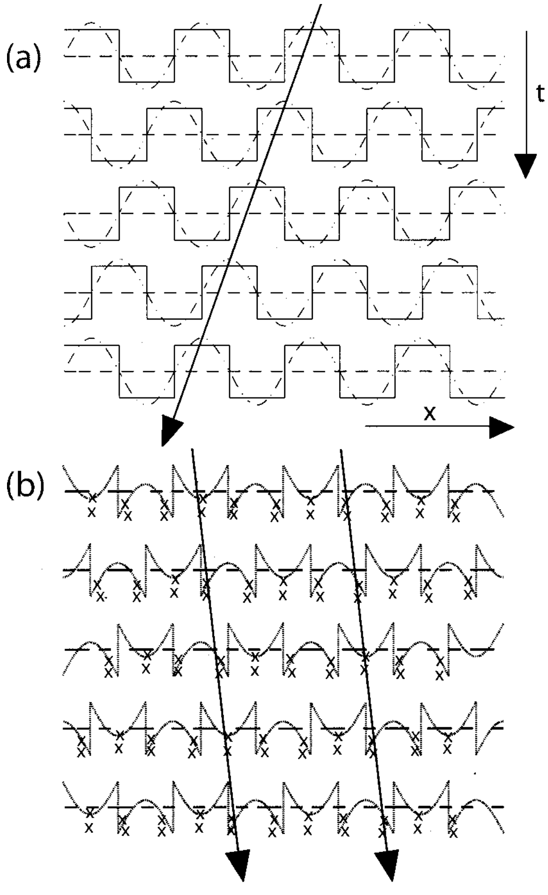

Fig. 2. Missing-fundamental stimuli and analysis. (a) Square wave (solid "curves") and the fundamental sine wave (dotted–dashed curves). Five consecutive frames are indicated, with a −90 deg (leftward) translation between frames. The slanting arrow indicates the direction of apparent motion. (b) The missing fundamental stimulus (continuous curves) produced by subtracting the fundamental sine wave from the square wave. xx indicates locations that are marked as being foreground (closer in depth) in the salience map. The arrow indicates the direction of motion based on the space–time modulation of salience (the xx's in consecutive frames). This is also the direction of motion of the third harmonic (not shown).

tral with respect to motion. The amplitude of each moving component is proportional to $1/(4^*k \pm 1)$. The strongest component in a MFG, the third harmonic, moves in the reversed direction. In the luminance domain and in the texture-contrast domain, 90-deg phase-stepping MFGs are normally perceived as moving in the reversed direction—the direction of the third harmonic. Occasionally, transparent forward and reversed motion is reported. Perceiving motion in the direction of the third harmonic has been taken as evidence for motion energy versus feature tracking as the algorithm for first- and second-order motion.

Smith and Scott-Samuels[6] applied the MFG paradigm to stereomotion. They found that the stronger percept is in the reversed direction, even though transparency was also sometimes reported. They concluded that stereomotion is served by a low-level motion-energy computation that independently computes motion at different spatial scales. The study is cited by Patterson[5] to support his assertion that the stereomotion is computed by a low-level motion-energy algorithm. Our reanalysis of the MFG stereomotion paradigm suggests an alternative interpretation of the results that is consistent with the theory that stereomotion is computed by the third-order motion system.

How does a third-order system compute motion from a 90-deg phase-stepping MFG? The same as in any other stimulus. It first computes a salience map and then computes motion from the salience map using a motion-energy algorithm. For a stereo MFG, the "near" regions are marked as figure (e.g., value 1); the "far" regions are regarded as ground and unmarked (e.g., value 0). In regions where there are rapid transitions from far to near, the marks are not exactly on the boundary but rather around the center of mass of the near regions.[25] As illustrated in Fig. 2(b), the salience maps generated by this algorithm indicate motion in the reversed direction. Smith and Scott-Samuels's results are completely consistent with a third-order computation.

The results of the demonstration in Fig. 2(b) can be taken more generally. Depending on how salience is assumed to be computed, the MFG paradigm may fail to distinguish between a third-order motion (or feature-tracking) computation and a motion-energy computation in other domains besides stereomotion.

D. Perceiving Forward and Reversed Motion in Moving Depth-Reversal Stimuli

1. *Reversed Phi*

The "reversed-phi" phenomenon was first reported by Reichardt[26] and named by Anstis.[27] In first-order motion, a bar (white or black) is flashed on a neutral background at location A and followed, after a brief interval, by another bar of *opposite* contrast polarity (black or white) at a nearby location B. In some circumstances, observers report perceiving motion in the *reversed* direction B to A instead of in the direction of the translation from A to B.[27]

2. *Two Mechanisms in First-Order Reversed Phi*

Chubb and Sperling[28] used a 90-deg-stepping, contrast-reversing grating instead of merely a contrast-reversing bar (as in the original demonstrations). The same stimuli that appeared to move in the forward direction (the direction of the translation) when viewed from near appeared to move in the reversed direction (opposite to the direction of the translation) from afar or in periphery vision—a finding that has been extensively corroborated in second-order perception.[29] Chubb and Sperling[28] proposed that there exist two motion mechanisms: a first-order mechanism that applies motion-energy analysis directly to the "raw" stimulus point contrast, and a second-order mechanism that applies motion-energy analysis to the full-wave-rectified stimulus point contrast.

3. *In 90-deg Reversed Phi, the Motion Energy Is Reversed*

Lu and Sperling[30] analyzed the mathematical properties of polarity-reversal moving gratings and demonstrated that for a periodic stimulus that steps 90 deg in successive frames, reversing the contrast of the even-numbered frames *reverses the direction of all the motion energy* in the stimulus. Therefore the mystery in such stimuli is not whether reversed-phi movement is perceived but how forward movement could be perceived. Perceiving forward motion in reversed-phi stimuli requires a higher-order motion computation.

Ito[7] studied motion perception of a moving depth-reversal stereo pattern made of random depth patches. He found that when the number of near and far patches was equal (the high-density condition) in each frame, observers perceived motion in the reversed direction more often with translations between 0.5 and 2 patch width and ambiguous motion with larger translations. When the number of near and far patches were very different (15:1; the low-density condition) in each frame, observers perceived motion in the forward direction. Following the logic of Chubb and Sperling[28] and Lu and Sperling,[30] Ito[7] concluded that two processes serve stereomotion: a depth polarity-independent, "feature-tracking" system in the low-density condition and a depth polarity-dependent, "passive" system. Patterson[5] cited Ito's results as counter evidence to our claim that stereomotion is processed only by the third-order system. We show here that Ito's results are consistent with a third-order motion computation.

4. *Low-Density Stereomotion Displays*

What does the third-order system predict in Ito's[7] low-density condition (1/15 patches differ in depth from the background). Small regions with unique properties are marked as figure in the salience map. The figure regions move forward in successive salience maps. This corresponds to Ito's depth polarity-independent process. Another important motion cue in the stimulus is the moving foreground rivalry regions. The observers verge on the ground (the region with larger area); the "unique" foreground regions produce rivalry patches. The rivalry patches are figure and move in the forward direction[4]—an instance of rivalry motion that we consider again in Subsection 3.E.

5. *High-Density Stereomotion Displays*

High-density stereomotion stimuli have equal numbers of near and far patches. Such stimuli are not periodic and obviously do not move exactly 90 deg in each step. The question of whether forward or reverse motion will be seen following a depth reversal of even frames is more complex. For example, the Fourier expansions of these stimuli typically involve a complex mixture of partially reversed and unreversed motion components.

We simulate predictions of the third-order system for moving normal and depth-reversal stereomotion stimuli made of random patches with various translations (translations were 0.5, 1.0, 1.5, and 2.0 patch width). The third-order system first computes the salience map for each input stereo frame by assigning a salience value 1.0 to near patches and 0.0 to far patches. It then submits the salience maps to two-point Reichardt detectors with various interpoint spatial separations ($\Delta x = 0.5, 1.0, 1.5$, and 2.0 patch width). These correspond to different scales at which attention operates and at which third-order motion is computed. (Different two-point separations characterize different classes of Reichardt detectors. It can be shown that properties of more complex Reichardt detectors are derivable from the properties of more elementary two-point Reichardt detectors.)[10,31] The outcomes of the motion simulations are consistent with Ito's experimental results: reversed motion in depth-reversal stereomotion.

Consider a two-frame stimulus with binary salience maps of $S_1(x, y)$ and $S_2(x, y)$ in the two frames. The motion energy output $\mathrm{ME}(x, y)$ of a simplified, two-point Reichardt detector[32] is

$$\mathrm{ME}(x, y) = S_1(x + \Delta x, y) S_2(x, y) - S_2(x + \Delta x, y) S_1(x, y), \tag{1}$$

where Δx is the spatial separation between the two subunits of a Reichardt detector. There is a Reichardt detector corresponding to each pixel location in each row and column. For each type of Reichardt detector (characterized by Δx) and for each spatial position (column x), the outputs of the Reichardt detectors in the four rows are summed. This yields the sets of graphs illustrated in Fig. 3.

Figure 3 illustrates the input stimuli (salience maps), and shows plots of the outputs of Reichardt detectors of various sizes (that is, the predictions of the third-order system). There is a separate set of graphs for each of the four different stimulus translations. Positive ME indicates leftward motion; negative ME indicates rightward motion. The stereomotion stimuli move to the left in the normal condition. In each translation condition, there is one Reichardt detector that has the optimal spacing Δx between its inputs to detect the motion. Reichardt detectors with nonoptimal spacing give very little useful directional information. A general observation from Fig. 3 is that classes of Reichardt detectors with dominantly positive ME in the normal condition give approximately equally dominantly negative ME in the depth-reversal condition. In other words, the "informative" Reichardt detectors signal forward direction in the normal condition and reversed direction in the depth-reversal condition. The third-order computation accounts nicely for Ito's results in both the low- and the high-density dot conditions.

E. Rivalry Motion

Consider a stereogram that is projected on a screen with polarizing projectors and viewed through polarizing filters so that the left eye sees only the left-eye image and the right eye sees only the right-eye image. Suppose that we view a dynamic random-dot stereogram that depicts a central rectangle, nearer in depth to the observer than the background. And suppose that the central rectangle drifts up (or down) in successive frames. This is how one might display ordinary stereomotion.

1. *Cyclopean Image*

We here define the cyclopean image as the algebraic sum of the image on the left retina plus the image on the right retina. Suppose each pixel in the left-eye image and the right-eye image has a value of 0.5. The cyclopean image has a value of 1.0. This is the same value that the cyclopean image would have if the left-eye pixel were zero and the right-eye pixel were 1.0. However, the two cases look quite different.

2. *Rivalry Image*

The rivalry between a black pixel in one eye and a white pixel at the corresponding point in the other eye produces quite a different appearance than does a gray pixel at corresponding points. Therefore we also must consider the rivalry image, which we define the absolute value of the difference between the left- and right-eyes' images or, equivalently, as the XOR image.[4] In many stereograms, conscious perception of the rivalry image is suppressed; the observer experiences little or no binocular rivalry. This means that the rivalry image is suppressed in the color–texture–shape system. Nevertheless, rivalry images can be demonstrated to strongly influence the third-order motion system.[4]

3. *Verging on Extreme Stimulus Elements*

Suppose that the observer verges his eyes on the moving rectangle. What are the cyclopean and rivalry images? We can produce the cyclopean image on the display screen by aligning the left- and right-eye images so that the central rectangle is in registry on the screen and then removing the polaroid filters. Now the two eyes have the same stimulus, the cyclopean image. What they see is depicted in Fig. 4A. The central rectangle is in registry; the background is not. When the rectangle moves, its motion is obvious, even when it is observed without the polaroid filters.[4] However obvious the movement in the cyclopean image may be, it is even more obvious in the rivalry image (Fig. 4A).

If the observer had verged on the background instead of the rectangle, the cyclopean image would have been that depicted in Fig. 4C. Again, when the background is in registry between the two eyes and the central rectangle is not, both the cyclopean and the rivalry image convey motion perfectly.

4. *Verging on the Depth Center of a Depth-Symmetric Display*

Only when the observer is verged precisely between the central rectangle and the background, and the central rectangle and the background are equally out of register, do the rivalry and cyclopean images convey no information about the shape and motion of the central rectangle. Because there is no actual stimulus within the stereogram of Fig. 4 to compel vergence on the in-between point (the display depicts only two depth planes), it would be impossible to actually verge on the in-between point. To maintain vergence at the in-between point requires very strong stimuli to vergence in addition to the random-dot stereogram itself. Even when vergence is perfectly controlled, only special classes of stimuli produce uninformative cyclopean and rivalry images.

Of course, the cyclopean images depicted in Fig. 4 occur when the left- and right-eye stereo half-images are physically added into the stimulus. In a normal stereogram, the left eye sees only its image, and the right eye only its image. However, the most prevalent class of simple cells in cortical area V1, those that receive approximately equal input from each eye and have similar, similarly located receptive fields in each retina, see the cyclopean image. Less is known about how rivalry is represented in the brain. However, both the sum and the difference (XOR) of the left- and right-eye stereograms are represented in the brain.

5. *Consequences of Cyclopean and Rivalry Images*

The extent to which an observer is conscious of cyclopean and rivalry images is difficult to determine and is not relevant here. What is relevant for the present discussion is that by control of vergence and the choice of appropriate stimuli, the cyclopean and rivalry images in the brain can be manipulated just as well as the images on the screen depicted in Fig. 4. Thereby we[4] have been able to measure the effectiveness of cyclopean and rivalry images in controlling apparent motion independent of other factors (see also Ref. 33). We find that motion of a rivalry image can be an extremely effective cue for motion-direction discrimination, approximately as effective as stereo depth itself.

Rivalry images depend on the state of vergence. To prove that rivalry images do not play a role in a stereomotion experiment, it would not be sufficient to demonstrate that the screen images when added and viewed by one or two eyes are uninformative for motion. The summed screen image might well be that shown in Fig. 4B, which is indeed uninformative for motion. But when the observer is verged on some stimulus element, he or she would be seeing not the cyclopean and rivalry images

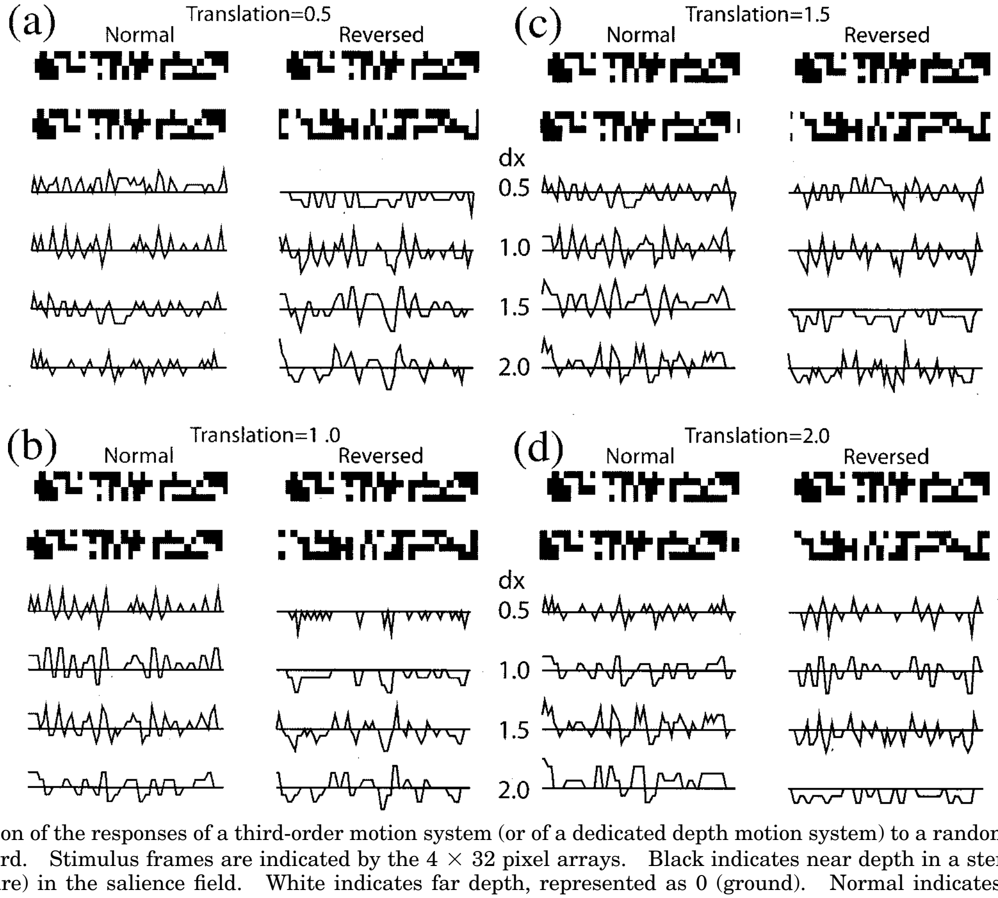

Fig. 3. Simulation of the responses of a third-order motion system (or of a dedicated depth motion system) to a random-dot pattern that translates leftward. Stimulus frames are indicated by the 4 × 32 pixel arrays. Black indicates near depth in a stereo display, represented as 1 (figure) in the salience field. White indicates far depth, represented as 0 (ground). Normal indicates that consecutive frames are identical except for a translation. Reversed indicates that the black–white relations are reversed between the first and second frames. The size of the translation (in pixels) is indicated for each quadrant. The jagged curves indicate the output of two-point Reichardt detectors with a separation of the two input points of exactly dx pixels (indicated). Reichardt detector outputs are computed separately for each of the four pixel rows of the stimulus and added to produce a summed output. Up represents leftward-motion output; down indicates rightward-motion output for detectors located at the indicated horizontal location. For each stimulus translation [panels (a)–(d)] there is a Reichardt detector of size dx that correctly detects the leftward motion of the normal translation and that reports the opposite direction for the reversed-phi translation.

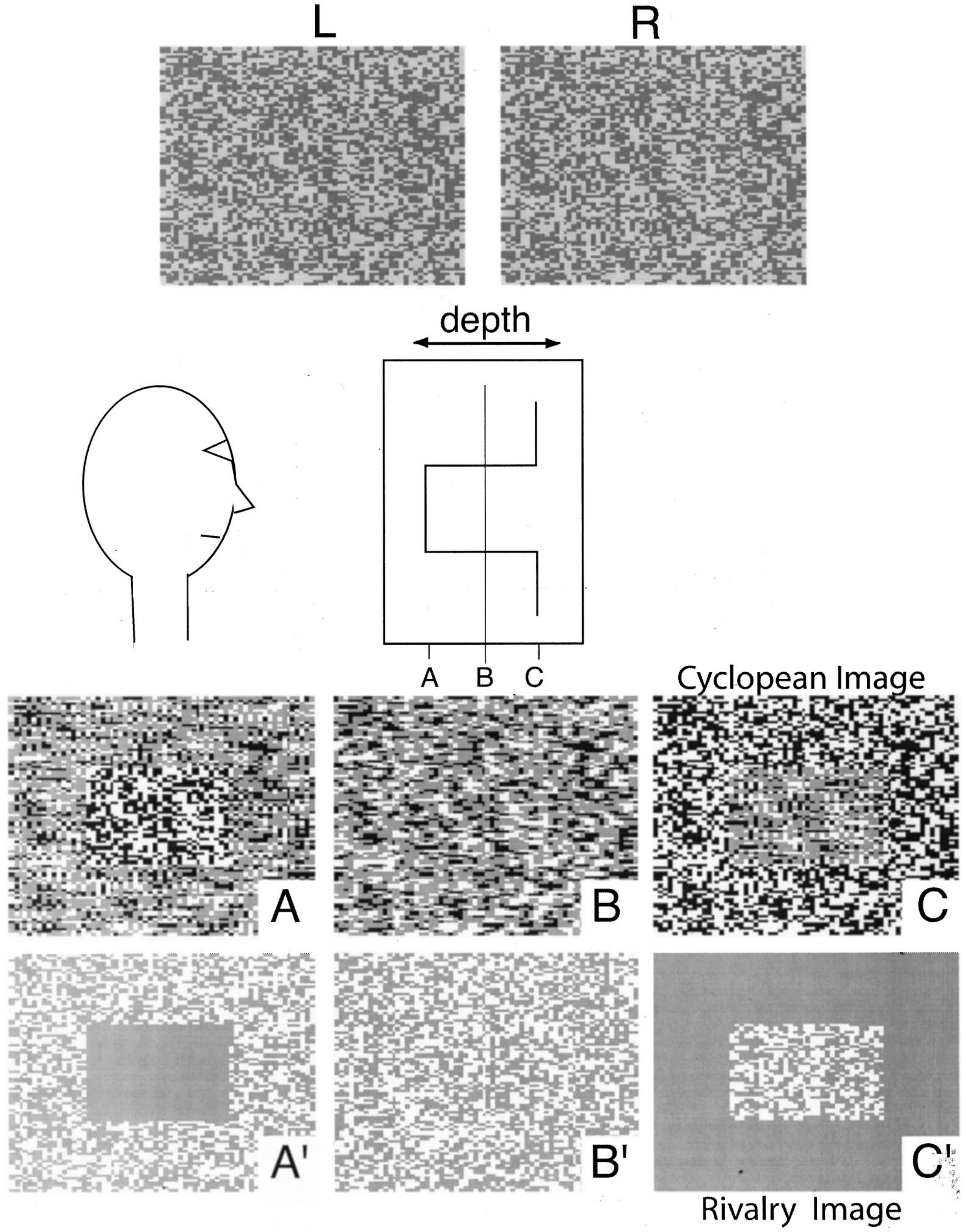

Fig. 4. Vergence, cyclopean images, and rivalry images. L and R represent two halves of a random-dot stereogram that defines a central rectangle that appears to be nearer to the observer than the background, as shown schematically in the middle section. A, B, and C represent planes drawn, respectively, through the front, middle, and back of the stereoscopic image. The cyclopean images A, B, and C represent the sum of L and R inputs falling on corresponding retinal points (cyclopean images) when the observer fixates on planes A, B, and C, respectively. The rivalry images A′, B′, and C′ represent the absolute value of the difference between L and R images. To produce a pure stereoscopic depth display with no object cues in the cyclopean or rivalry images requires fixation to be perfectly in between the front and back planes (B).

of Figs. 4B and 4B′ but the cyclopean and rivalry images of Figs. 4A and 4A′ or Figs. 4C and 4C′, which are highly informative for motion. For the stimuli described by Patterson,[5] it is difficult or impossible to avoid rivalry images. Therefore rivalry motion is a probable confound in these experiments. To draw conclusions that are unconfounded by rivalry images requires better experiments.

4. PHASE PARADIGM FOR DISCRIMINATING BETWEEN THIRD-ORDER AND DEDICATED (LOW-LEVEL) STEREOMOTION PROCESSING

Figure 1 illustrates the preprocessing for a third-order motion system and for a low-level, dedicated stereomotion-energy system. The essential difference is the preprocessing, not the motion-extraction algorithm. By considering only the two systems' responses to simple stimuli, it would be difficult to discriminate between a third-order motion system based on a salience map and a dedicated motion-energy computation based on a depth map. For example, suppose that salience varies from 0 for ground to 1 for figure. Suppose that stereo depth varies from -1 for background to 1 for foreground, with the horopter at 0. It is easy to construct plausible computations for figure–ground motion that perfectly mirror computations for stereo depth motion.

A. Phase Dependence of Motion Signals Traveling in the Same Direction

The essential ingredient of the third-order motion theory is that the many components of figure–ground processing combine *before* the motion computation. This aspect of processing is something for which there is a powerful test, the phase test.[1] Suppose that two signals, each at its own threshold level, each traveling at the same speed and with the same wavelength, arrive at a motion detector simultaneously. If the two signals are directed to different motion detectors, then motion processing is independent of the phase between the two signals. And because the two signals are in the same direction, motion strength should always be greater with the two signals together than for either signal individually. When the two signals arrive at the same motion detector, they can either enhance each other if they are in phase or cancel each other if they are out of phase. The ability of two motion signals that are traveling together in the same direction to cancel is the signature of their interaction prior to or within a motion computation.

B. Results of Phase-Dependence Tests

The phase test was used to establish that first- and second-order motion signals traveling in the same direction have virtually zero phase dependence and therefore are processed by independent detectors.[1] Experiments to test the phase dependence of different third-order motion stimuli are currently underway. Preliminary data with phase tests of concurrent stereo-depth motion and rivalry motion indicate strong phase dependence, indicating that these motions are processed by the same detector—the motion detector of the third-order motion system. Preliminary data obtained by C. Tseng and the second author indicate that isoluminant chromatic motion and stereo depth motion have a very strong phase dependence. By transitivity, this means that isoluminant chromatic motion, stereo depth motion, and rivalry motion are all processed by the same motion system. The phase tests already carried out and those underway will enable us to create a taxonomy of motion systems based not on conjecture and argument but on a paradigmatic principle.

5. SUMMARY AND CONCLUSION

A more careful analysis of the experiments cited by Patterson to refute third-order (figure–ground) motion processing shows that they are in fact completely consistent with the third-order motion algorithm. The fact that stereomotion follows the typical third-order temporal tuning function and fails the pedestal test[1] strongly suggests the third-order motion system. That stereo depth gratings in odd frames and texture gratings in even frames of a motion sequence can combine to produce motion indicates that stereo depth could be processed in the third-order motion system (no other system combines such disparate stimuli). That it is easy to produce motion standstill in stereomotion[34] suggests there is only one system for detecting stereomotion, because it seems somewhat unlikely that the same condition would simultaneously silence two motion systems while leaving stereo depth intact. Rivalry motion and stereomotion are of approximately equal strength, and phase tests show that they are processed by the same (presumably third-order) motion system.

Taken together, the data considered here and the lack of evidence to the contrary, suggest that as of this writing (2002), it is reasonable to conclude that stereomotion is perceived by, and only by, the third-order motion system.

ACKNOWLEDGMENT

This research was supported by the U.S. Air Force Office of Scientific Research, Life Science Directorate, Visual Information Processing Program. For reprints, please write Zhong-Lin Lu, Laboratory of Brain Processes, Department of Psychology, SGM 501, University of Southern California, Los Angeles, California 90089-1061; or George Sperling, Department of Cognitive Sciences, University of California, Irvine, California 92697-5100. Phone, fax, and e-mail are as follows: Z.-L. Lu: 213-740-2282, 213-746-9082, zhonglin@rcf.usc.edu. G. Sperling: 949-824-6879, 949-824-2517, sperling@uci.edu.

REFERENCES

1. Z.-L. Lu and G. Sperling, "The functional architecture of human visual motion perception," Vision Res. **35**, 2697–2722 (1995).
2. Z.-L. Lu and G. Sperling, "Three-systems theory of human visual motion perception: review and update," J. Opt. Soc. Am. A **18**, 2331–2370 (2001).
3. R. Patterson, "Stereoscopic (cyclopean) motion sensing," Vision Res. **39**, 3329–3345 (1999).
4. H. J. Kim, Z.-L. Lu, and G. Sperling, "Rivalry motion versus depth motion," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **42**, 3947 (2001).
5. R. Patterson, "Three-systems theory of human visual motion perception: review and update: comment," J. Opt. Soc. Am. A **19**, 2142–2143 (2002).
6. A. T. Smith and N. E. Scott-Samuel, "Stereoscopic and contrast-defined motion in human vision," Proc. R. Soc. London Ser. B **265**, 1573–1581 (1998).
7. H. Ito, "Two processes in stereoscopic apparent motion," Vision Res. **39**, 2739–2748 (1999).
8. J. P. H. van Santen and G. Sperling, "Temporal covariance model of human motion perception," J. Opt. Soc. Am. A **1**, 451–473 (1984).
9. E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of apparent motion," J. Opt. Soc. Am. A **2**, 284–299 (1985).
10. C. Chubb and G. Sperling, "Drift-balanced random stimuli: a general basis for studying non-Fourier motion perception," J. Opt. Soc. Am. A **5**, 1986–2006 (1988).

11. Z.-L. Lu and G. Sperling, "Attention-generated apparent motion," Nature **377**, 237–239 (1995).
12. E. Blaser, G. Sperling, and Z.-L. Lu, "Measuring the amplification of attention," Proc. Natl. Acad. Sci. USA **96**, 11681–11686 (1999).
13. P. Cavanagh, "Attention-based motion perception," Science **257**, 1563–1565 (1992).
14. Z.-L. Lu, L. A. Lesmes, and G. Sperling, "Perceptual motion standstill in rapidly moving chromatic displays," Proc. Natl. Acad. Sci. USA **96**, 15374–15379 (1999).
15. P. Cavanagh, M.-A. Henaff, F. Michel, T. Landis, T. Troscianki, and J. Intriligator, "Complete sparing of high-contrast color input to motion perception in cortical color blindness," Nat. Neurosci. **1**, 242–247 (1998).
16. Q. Zaidi and S. J. DeBonet, "Motion energy versus position tracking: spatial, temporal and chromatic parameters," Vision Res. **40**, 3613–3635 (2000).
17. P. Cavanagh, M. Arguin, and M. von Gruenau, "Interattribute apparent motion," Vision Res. **29**, 1197–1204 (1989).
18. R. Patterson, M. Donnelly, R. E. Phinney, M. Nawrot, A. Whiting, and T. Eyle, "Speed discrimination of stereoscopic (cyclopean) motion," Vision Res. **37**, 871–878 (1997).
19. S. Nishida, T. Ledgeway, and M. Edwards, "Dual multiple-scale processing for motion in the human visual system," Vision Res. **37**, 2685–2698 (1997).
20. Z.-L. Lu, G. Sperling, and J. Beck, "Selective adaptation of three motion systems," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **38**, 237 (1997).
21. W. A. Phillips, "On the distinction between sensory storage and short-term visual memory," Percept. Psychophys. **16**, 283–290 (1974).
22. E. H. Adelson, "Some new motion illusions, and some old ones, analysed in terms of their Fourier components," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **34**, 144 (1982).
23. M. A. Georgeson and T. M. Shackleton, "Monocular motion sensing, binocular motion perception," Vision Res. **29**, 1511–1523 (1989).
24. A. T. Smith, "Correspondence-based and energy-based detection of second-order motion in human vision," J. Opt. Soc. Am. A **11**, 1940–1948 (1994).
25. P. Burt and B. Julesz, "A disparity gradient limit for binocular fusion," Science **208**, 615–617 (1980).
26. W. Reichardt, "Autocorrelation, a principle for the evaluation of sensory information by the central nervous system," in *Sensory Communication*, W. A. Rosenblith, ed. (Wiley, New York, 1961), pp. 303–317.
27. S. M. Anstis, "Phi movement as a subtraction process," Vision Res. **10**, 1411–1430 (1970).
28. C. Chubb and G. Sperling, "Two motion perception mechanisms revealed through distance-driven reversal of apparent motion," Proc. Natl. Acad. Sci. USA **86**, 2985–2989 (1989).
29. T. V. Papathomas, A. Gorea, and C. Chubb, "Precise assessment of the mean effective luminance of texture patches: an approach based on reverse-phi motion," Vision Res. **36**, 3775–3784 (1996).
30. Z.-L. Lu and G. Sperling, "Second-order reversed phi," Percept. Psychophys. **61**, 1075–1088 (1999).
31. C. Chubb and G. Sperling, "Texture quilts: basic tools for studying motion-from-texture," J. Math. Psychol. **35**, 411–442 (1991).
32. J. P. H. van Santen and G. Sperling, "Elaborated Reichardt detectors," J. Opt. Soc. Am. A **2**, 300–321 (1985).
33. R. P. O'Shea and R. Blake, "Depth without disparity in random-dot stereograms," Percept. Psychophys. **42**, 205–214 (1987).
34. C. H. Tseng, H. Kim, J. L. Gobell, Z.-L. Lu, and G. Sperling, "Motion standstill in rapidly moving stereoptic depth displays," Invest. Ophthalmol. Visual Sci. ARVO Suppl. **42**, S504 (2001).