Science, 286:2231, 1999.

Filtering Reveals Form in Temporally Structured Displays

In a recent report, Lee and Blake (1) asked whether the visual system could use temporal microstructure to bind image regions into unified objects, as has been proposed in some neural models (2). They presented two regions of dynamic texture. The elements of the target region changed in synchrony according to a random sequence, while the elements of the background region changed at independent times. The stimulus was designed in an attempt to remove all classical form-giving cues such as luminance, contrast, or motion, so that timing itself would provide the only cue. Subjects were readily able to distinguish the shape of the target region. Lee and Blake posited the existence of new visual mechanisms "exquisitely sensitive to the rich temporal structure contained in these high-order stochastic events." The results have already generated much excitement (3).

We believe the effects can be explained with well-known mechanisms. The filtering properties of early vision can convert the task into a simple static or dynamic texture discrimination problem. A sustained cell (temporal lowpass) will emphasize static texture through the mechanisms of visual persistence; a transient cell (temporal bandpass) will emphasize texture that is flickering or moving.

We simulated a lowpass mechanism to see what would emerge. Lee and Blake's stimuli were composed of randomly oriented Gabor elements, where the Gabor phase shifted forward or backward on each frame according to a coin-flip. We downloaded one such movie from their web site and ran it through a temporal lowpass filter (4). An input frame is shown in Figure 1(a). A filtered output frame is shown in Figure 1(b). At the particular moment shown, the target region has a lower effective contrast than the background, providing a strong form cue. At other moments the target's contrast may be above or below the background's contrast, due to statistical fluctuations in the reversal sequences. If a single Gabor element happens to have a run of multiple shifts in one direction, its effective contrast is low due to the temporal averaging. Conversely, if it has a run of alternating forward and backward shifts (thus "jittering" in place), its contrast remains fairly high. Within the unsynchronized background, the local contrasts fluctuate randomly, but within the synchronized target region they all rise and fall in unison, revealing a distinct rectangular form.

In a second experiment Lee and Blake synchronized both the target and background region, each to its own random sequence. The target was even more clearly visible. This result is predicted by our hypothesis. Since both background and target are synchronized, they will both yield uniform texture contrasts after temporal filtering. There will be moments when, by chance, one region's contrast is high while the other's is low, and the target will become especially clear. Figure 1(c) shows one such moment, again the result of filtering a movie from the website with the lowpass filter. Figure 1(d) shows a moment when the relative contrasts are reversed. We also ran the movies through a temporal bandpass filter (5) with a biphasic impulse response, to simulate a transient mechanism. Again, the target was clearly revealed.

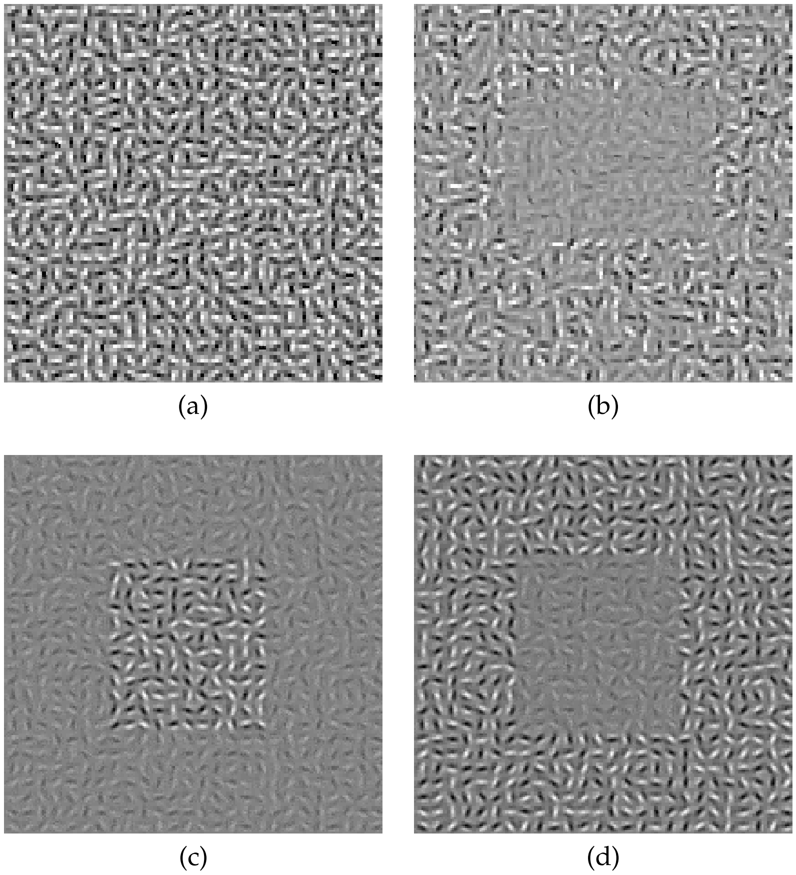
With either filter type, our hypothesis also predicts Lee and Blake's finding that discrimination will be best when the reversal sequences have high entropy, i.e., when the coin-flip is unbiased. The contrast cue is best when the target "jitters" in place while the background has a run in a single direction (or vice versa). This condition happens most frequently at high entropy.

Lee and Blake's stimuli are cleverly designed to remove form cues from single frames and from frame pairs. However, when one considers the full sequence, strong contrast cues can emerge due to the spatio-temporal filtering present in early vision. These cues probably suffice to explain the perception of form in the experiments. We do not see the need to posit special mechanisms other than those already known to exist.

Edward H. Adelson
Hany Farid
Department of Brain and Cognitive Sciences
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
{adelson,farid}@persci.mit.edu

REFERENCES AND NOTES

1. S. Lee and R. Blake, Science, 284, 1165-1168 (1999)
2. W. Singer and C. Gray, Ann. Rev. Neurosci., 18, 555-586 (1995)
3. News, Science, 8, 1098-1099 (1999)
4. The lowpass impulse response was of the form $h(t) = (t/\tau)^2 e^{-t/\tau}$, with $\tau = 0.01$. The "integration time" was roughly 40 msec.
5. The bandpass impulse response was of the form $h(t) = (kt/\tau)^n e^{-kt/\tau}[1/n! - (kt/\tau)^2/(n+2)!]$, with with $\tau = 0.01$, $k = 2$ and $n = 4$. The peak response was at 5 Hz.



(a)

(b)

(c)

(d)

Figure 1: Shown are: (a) one input frame; (b) result of temporal integration with synchronized target and unsynchronized background; (c-d) results of temporal integration when the target and background are each synchronized.