massachusetts institute of technology — artificial intelligence laboratory

Perceiving Illumination Inconsistencies in Scenes

Yuri Ostrovsky, Patrick Cavanagh and Pawan Sinha

AI Memo 2001-029 November 2001
CBCL Memo 209

© 2001 massachusetts institute of technology, cambridge, ma 02139 usa — www.ai.mit.edu

Abstract

The human visual system is adept at detecting and encoding statistical regularities in its spatio-temporal environment. Here we report an unexpected failure of this ability in the context of perceiving inconsistencies in illumination distributions across a scene. Contrary to predictions from previous studies [Enns and Rensink, 1990; Sun and Perona, 1996a, 1996b, 1997], we find that the visual system displays a remarkable lack of sensitivity to illumination inconsistencies, both in experimental stimuli and in images of real scenes. Our results allow us to draw inferences regarding how the visual system encodes illumination distributions across scenes. Specifically, they suggest that the visual system does not verify the global consistency of locally derived estimates of illumination direction.

The research reported in this paper was supported in part by funds from the Alfred P. Sloan Fellowship in Neuroscience to PS and an NIH Graduate Training Grant to YO.

INTRODUCTION

When special-effects cinematographers create composite scenes (figure 1a), they go to great lengths to ensure that all objects are consistently illuminated [Fielding, 1985; Brinkmann, 1999]. Differences in lighting directions across objects, they believe, would be immediately evident to the audience and reduce the realism of the scene. This intuition appears to be supported by formal experimental evidence. Several recent studies have demonstrated that in an array of identically lit three-dimensional objects (figure 1b), the visual system can rapidly (typically in about 100 milliseconds) and reliably spot an anomalously illuminated item [Enns and Rensink, 1990; Sun and Perona, 1996a, 1996b, 1997; Braun, 1993] (figure 1b).

a)

b)

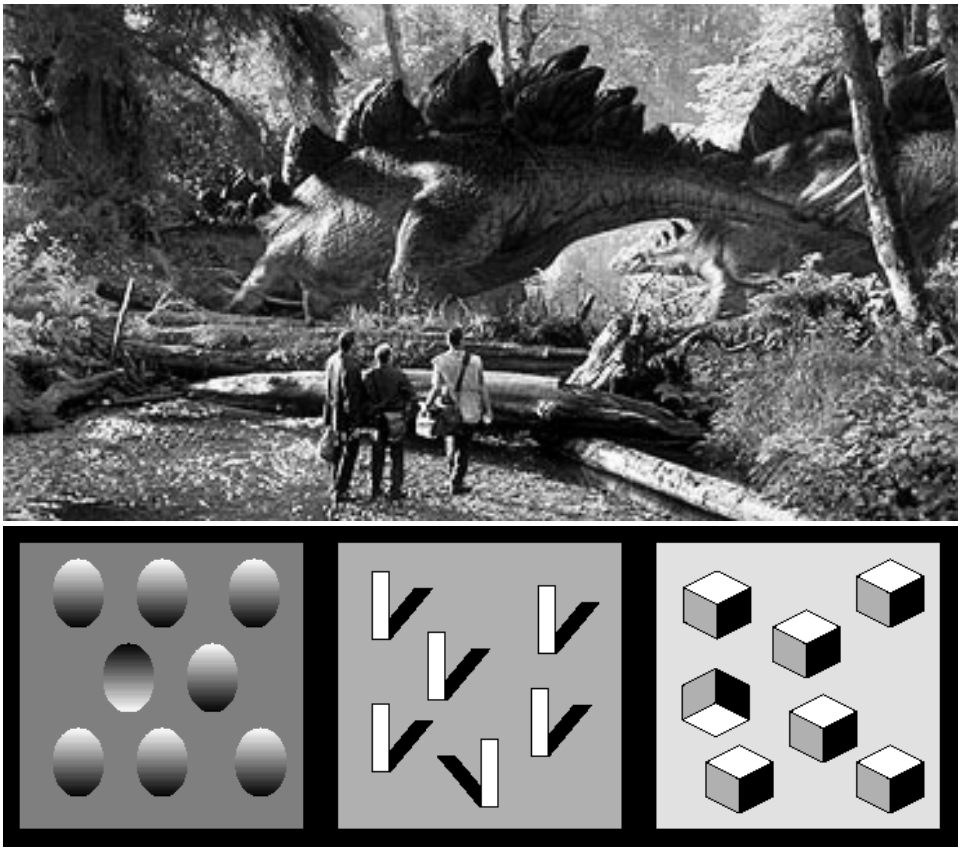


Figure 1. *Scenes for which anomalies in illumination direction across objects are either expected to be or, in fact are readily detectable. (a) A sample composite scene from the movie Jurassic Park-II. The dinosaurs and the humans are derived from separate scenes. Lighting directions for both are precisely equated. (b) Experimental displays which suggests that anomalies in lighting directions are perceptually very salient and 'pop-out' pre-attentively.*

The diversity of displays with which such results have been obtained suggests that the brain can detect illumination inconsistencies in arbitrary scenes, perhaps through the use of some general-purpose rules governing what constitutes a consistent illumination pattern. These rules could have been acquired through experience by encoding the regularities in illumination distributions. For instance, studies with humans and other animals suggest that they possess an innate bias towards assuming a single light source positioned above the terrestrial environment [Ramachandran, 1988; De Haan et al., 1995]. Given these data and the experimental results on illumination

anomaly spotting, it seems conceivable that statistical learning over the course of evolutionary or individual time scales might have endowed us with an ability to detect inconsistencies in illumination directions across a scene.

Although previous studies demonstrate that subjects possess an impressive ability to pick the "odd man out", it is unclear to what extent this ability is based on the perception of illumination inconsistencies per se. Homogenous fields of identical distractors, such as those used in these studies, may allow for the use of simple image-matching strategies for detecting the embedded oddity [Aks and Enns, 1992]. Furthermore, the experimental displays used so far fail to capture the characteristics of real-world scenes in at least one important way. They implicitly assume that all objects in the environment have the same three-dimensional pose. This is clearly unrepresentative of natural scenes where objects typically have different poses.

In order to overcome these potential problems in characterizing human sensitivity to illumination inconsistencies, we have designed displays that differ from those used thus far in a key respect. As in many previous studies, our displays comprise several identical three-dimensional (3D) objects with all distractors illuminated from one direction and the target from a different direction. However, instead of assigning the same orientations in space to these objects, we randomize them (see figure 2a). This makes illumination direction the only reliable differentiator between targets and distractors. By having the distractors not subscribe to an identical 2D pattern, this manipulation reduces the effectiveness of simple 2D pattern matching strategies. Furthermore, by not constraining all the objects in the display to assume identical poses, our stimuli better represent real-world conditions. We assess observers' ability to detect illumination inconsistencies in such scenes.

EXPERIMENTAL PROCEDURE AND RESULTS

We chose the cube, a simple 3D object with a history of use in this domain [Sun and Perona, 1996a, 1996b], as the stimulus item. Figure 2b shows a sample display. Half of all our displays were fully consistent (all objects illuminated from the same direction), while in the other half, one cube (the target) was illuminated from an orthogonal direction relative to the distractors. As in the previous studies, subjects were asked to report whether the display contained an anomalous lit target, *i.e.*, whether the scene had an illumination direction inconsistency. We report the results of two separate experiments, one with fixed presentation times and the other self-timed. The first experiment showed displays with set sizes of 4, 9, and 12 items for durations of 100ms, 500ms, or 1000ms. (Each trial was pseudo-randomly assigned a set size and presentation time.) The second experiment (taken after the first experiment by the same set of subjects) consisted of displays identical to those of the first experiment, but displays persisted on the screen until subjects made a response.

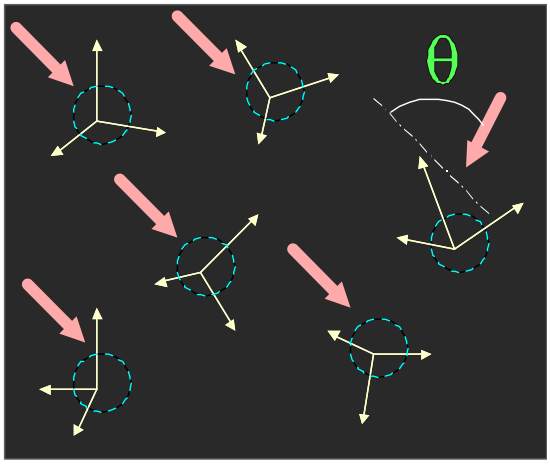
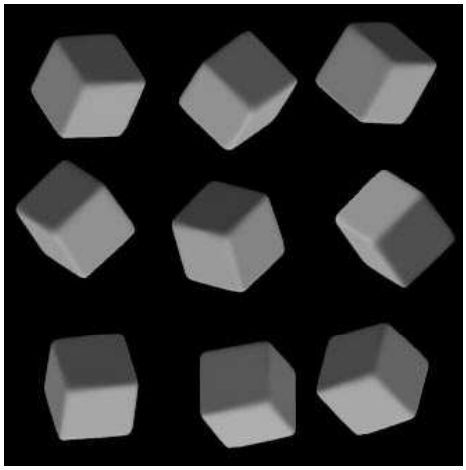
a)  b) 

Figure 2. *Display design for our experiments.(a) Basic configuration of our displays. The objects are all identical, as in the previous studies, but their orientations in space have been randomized so that illumination direction is the only reliable differentiator between distractors and target. (b) A sample display.*

In order to replicate results from earlier studies and also to have a baseline condition, we first tested subjects on displays with all distractors oriented identically. Subjects achieved ceiling level performance (about 90% correct detection of anomalous displays) under these conditions even with just a 120 ms display time. Performance was invariant to distractor numerosity within a wide range (4 – 12), indicating a parallel search [Triesman, 1985].

For the next set of experiments, we controlled for the confounds introduced by distractor homogeneity by randomizing cube orientations. This control dramatically changed the results. Figure 3(a, b, c) shows results from three different fixed display-times and the self-timed conditions. The data demonstrate a remarkable inability on the part of the subjects to detect illumination inconsistencies even with long viewing durations. In contrast to the ceiling level performance with homogenous distractors, maximal performance with inhomogenous distractors averaged 65% even for small set-sizes. (Chance level performance is 50%.) Performance decreased with distractor numerosity and increased with display time, indicating a slow serial search strategy. Set size had a significant effect on performance (2-factor ANOVA: $p < 10^{-4}$), as did display time ($p < .03$ for the timed conditions, $p < 0.01$ for the self-timed condition). This pattern of results is clearly quite different from those obtained in previous studies.

Based on these results, we infer that subjects were quite insensitive to the illumination inconsistencies embedded in the experimental displays. The discrepancy between results from earlier experiments and our studies suggests that the use of a homogenous field of distractors may have rendered the task of spotting the illumination anomalies unnaturally easy. By the same token, however, our study may be criticized for making the task unnaturally hard – the heterogeneity of orientations may have a detrimental effect on visual search tasks in general. There are at least two responses to this concern. First, as mentioned in the introduction, non-homogeneity of object-poses in a real-world scene is the rule rather than the exception. By incorporating this characteristic of natural scenes, our displays are rendered more ecologically valid than those with entirely homogenous arrays of objects. The results are, therefore, more likely to be reflective of our perceptual abilities in the real-world. Second, it is not the case that visual search in general is compromised by the heterogeneity of item appearance in our displays. In separate studies, we have found that in such displays, observers' ability to rapidly detect

inconsistencies in a variety of other attributes such as spectral content and intensity of illumination, shape and stereo depth is not compromised by the heterogeneity.

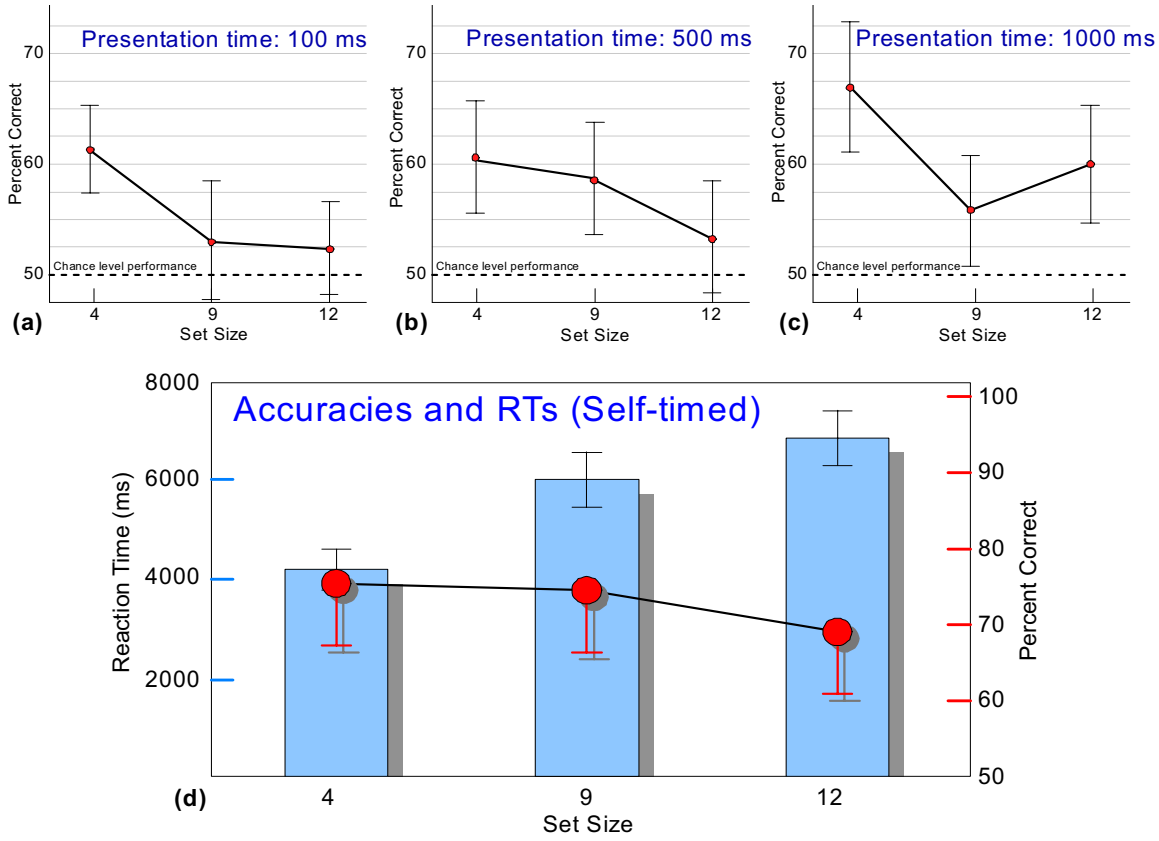


Figure 3. *Experimental results. Each graph shows performance as a function of set size, parameterized by presentation time. (a-c) Performance on timed conditions. Performance decreases with set size, and increases with display time, as expected for slow, serial search tasks. (d) Reaction time data (with performance) for self-timed condition. Reaction time increases with set size, whereas performance remains the same or even decreases with set size, despite the increase in reaction time. Note that, although performance is better than on the faster, timed trials of a-c, in absolute terms, it is still quite poor, illustrating that the task of detecting illumination anomalies may be fundamentally difficult for our perceptual system.*

In the light of these data, an important question that needs to be addressed is whether these results are specific to the experimental stimuli we used or if they apply to real-world scenes as well. To address this issue, we digitally modified images of real scenes to introduce illumination inconsistencies in them. The inconsistency between illumination directions averaged 90 degrees. Some examples of the resulting images are shown in figure 4. A cursory examination of these scenes suggests that their illumination inconsistencies are not immediately evident – consistent with the results we obtained with the experimental stimuli. To verify these informal observations, we designed an experiment wherein subjects were shown 23 pairs consisting of one modified and one unmodified real scene in random order. Subjects had to indicate which scene in each pair had illumination inconsistencies.

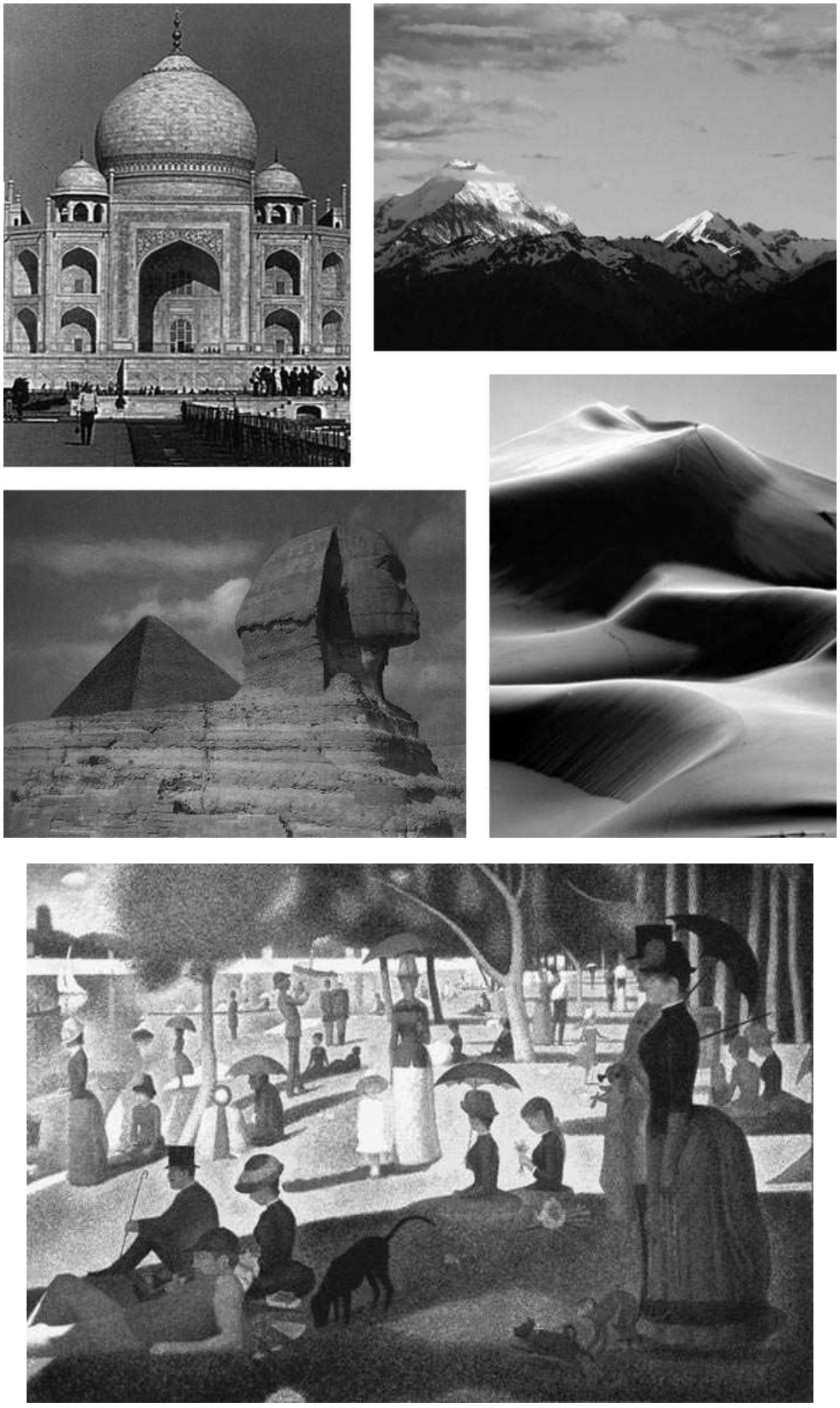


Figure 4. *A few examples of scenes with digitally introduced illumination anomalies. Just as with the experimental displays shown in figure 2, the inconsistencies in these scenes are not perceptually salient.*

Figure 5 shows subjects' performance as a function of presentation time. Just as with our previous experimental displays, subjects performed poorly even with extended presentation times. Notwithstanding the explicit instructions to look for illumination direction inconsistencies, subjects were not significantly above chance at presentation times of one second. Their performance improved to 70% when presentation time was increased to 5 seconds. These results indicate that the illumination inconsistencies do not 'pop-out', but require a relatively slow scan of the scene. In fact, it is conceivable that illumination inconsistencies in real scenes may be even less evident than is suggested by our results. In our stimuli, though we were careful to avoid them as best as we could, there may have been some local image artifacts (such as edges and chromatic differences) arising out of the image doctoring operation. These artifacts may allow subjects to distinguish between modified and unmodified images. Furthermore, subjects were explicitly told before the start of the experiment to look for illumination inconsistencies. Unprimed subjects can be expected to be less sensitive to the inconsistencies in the scenes. Indeed, in preliminary tests with subjects who were asked to pick out 'doctored' from 'undoctored' images without explicitly being asked to look for illumination inconsistencies, we found performance to be at chance even at the longest (5 sec.) presentation times.

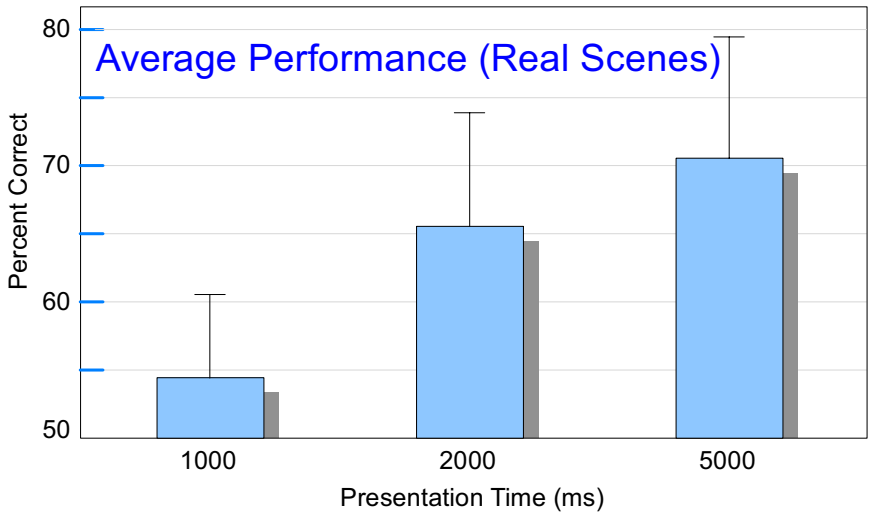


Figure 5. *Summary of our results with real images. Despite being primed specifically to look for illumination inconsistencies in the images, subjects' performance was quite poor even with long presentation times. Chance level performance is 50%.*

CONCLUSION

In summary, our results suggest that humans are quite insensitive to illumination direction inconsistencies in the experimental displays we used and, more generally, in many real world scenes as well. Artists have often exploited this insensitivity by choosing to depict illumination patterns in their paintings based on compositional aesthetics and social norms rather than constraining the patterns by physical laws [Gombrich, 1995].

What might account for the visual system's insensitivity to illumination direction inconsistencies? It is unlikely to be due simply to the visual system ignoring shading and shadows altogether.

Studies such as [Johnston et al., 1992; Hietanen et al., 1992; Braje et al., 1998, 2000; Tarr et al., 1998] have shown that shading patterns are indeed encoded by the brain. In fact, even very young infants have been shown to be sensitive to information provided by shadows in pictures [Yonas et al., 1978; 1979; Cameron and Gallup, 1988] Also, previous studies have shown that the visual system can determine illumination direction for local image regions [Hagen, 1976; Todd and Mingolla, 1983; Pentland, 1982]. The contribution of our experiments lies in allowing us to address the issue of how these local estimates are combined across a scene. The results suggest that the visual system does not attempt to verify the global consistency of the local estimates. A corollary of this finding is that it is unlikely that the visual system encodes global illumination distributions [Langer and Zucker, 1997].

There are two potential ecological roots of this 'deficiency'. First, there appears to be little adaptive advantage to be gained from having the ability to perform global illumination consistency verification. Local analysis typically suffices for key tasks like shape recovery [Erens et al., 1993; Weinshall, 1994]. Second, in a single-source world, local analysis suffices for global illuminant direction estimation. According to this idea, our indifference to verifying global consistency of light directions may, curiously enough, derive from the fact that our evolutionary history took place in an environment where a unitary light source (the sun) automatically enforced global consistency of local illumination patterns. Notice that this idea is distinct from the hypothesis that our insensitivity to illumination inconsistencies is due to our willingness to tolerate multiple light sources. While this hypothesis does account for the data, arguments against it include the accumulated body of work that indicates the visual system's bias towards assuming a single light-source [Ramachandran, 1988; Kleffner and Ramachandran, 1992], the lack of sensitivity observed even when the source is likely to be unitary (large-scale sunlit scenes) and instances for which even the assumption of multiple light sources does not provide an adequate explanation (for instance, the inconsistency between the directions of shading gradient and shadow of the woman's skirt in Seurat's painting (figure 4, bottom panel)).

How can we reconcile these experimental results with our subjective experience of noticing illumination anomalies in some old movies or poorly composited images? Three factors may increase the perceptual salience of the anomalies in these cases. First, inconsistencies in other aspects of illumination besides just direction (for instance, intensity, spectral content and light-source numerosity) may make them more easily detectable. Second, the existence of compositing-related artifacts, such as luminance, texture or color edges, may signal the presence of inconsistencies. Third, familiarity with a scene may reduce the task of spotting illumination inconsistencies to one of novelty detection in images. The familiarity-based hypothesis would predict that inconsistencies in highly familiar scenes or objects would be perceptually salient. While we await a thorough test of this hypothesis, preliminary evidence does lend support to it. We experimented with images of human faces – a highly familiar class for observers. Consistent with the hypothesis, anomalies in illumination of the kind shown in figure 6 are readily perceived (mean reaction time to distinguish between anomalous and non-anomalous facial illumination was 300 ms). This result does not, however, distinguish between two possibilities. It may be the case that all lighting anomalies are perceptually salient so long as they are within one object, irrespective of whether the object is familiar or not. Alternatively, in order for the anomalies to be perceptually salient, the scene (showing a single object or multiple objects) may need to be familiar. Our preliminary results support the latter possibility. We measured reaction times with vertically inverted versions of our face stimuli. Given the lower level of familiarity observers have with inverted faces, the second possibility, but not the first, would predict that detection of illumination inconsistencies in inverted face images would be more difficult relative to upright faces. This is indeed what we find. Reaction times with inverted faces are nearly twice as long as

with upright faces. The reader may verify the reduction in anomaly salience by turning figure 6 upside down.

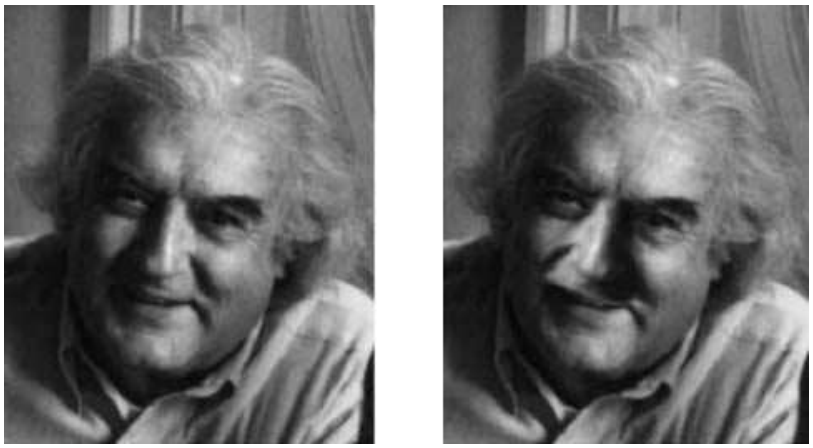


***Figure 6**. Illumination inconsistencies in faces. Results from upright-inverted experiments.*

In summary, our results show that observers are often remarkably insensitive to illumination direction inconsistencies in experimental and natural scenes. They lead us to conclude that the visual system does not attempt to verify global consistency of the local illumination direction estimates. These results bring up additional interesting questions such as the roles of motion, albedo changes and object relatability in facilitating the detection of illumination inconsistencies. Most of these issues can be investigated using the experimental paradigm we have presented in this paper.

METHODS

Experiments 1 and 2 (Identifying inconsistent cube displays, timed and self-timed conditions)
The display was located approximately 50-70 cm from the subject, with simulus items spanning approximately 2 degrees, located at a maximum of about 10 degrees from the center. In experiment 1 (the timed condition), after an image was shown for the allotted time (100, 500 or 1000ms), a gray screen appeared and remained until the subject made a response. Times and set-sizes were pseudorandomly assigned at each trial. In experiment 2 (the self-timed condition), the image remained on the display until the subject made a response. Subjects indicated whether the display was consistent or inconsistent in illumination direction by pressing one of two keys on a computer keyboard.

Experiment 3 (Identifying inconsistent natural scenes)
The display was located approximately 50-70 cm from the subject, with pictures of natural scenes spanning approximately 12-20 degrees. In blocks 1, 2 and 3, each image was shown for 1000, 2000 and 5000ms respectively. Each block showed 23 image pairs, each of which contained one image with inconsistent illumination on a black screen. A pair was presented as follows: (1) The first image in the pair was shown for the allotted time. (2) A gray screen was shown for 200ms. (3) The second image was presented for the same amount of time. (4) A gray screen was shown until the subject indicated which of the two images was inconsistent by pressing one of two keys on the keyboard.

ACKNOWLEDGEMENTS

The authors would like to thank Antonio Torralba, Daniel Kersten, Ted Adelson and Heinrich Buelthoff for useful discussion regarding this work.

REFERENCES

Aks, D. J. & Enns, J. T. (1992). Visual search for direction of shading is influenced by apparent depth. *Perception & Psychophysics*, **52**, 63-74.

Braje, W.L., Legge, G.E., & Kersten, D. (2000) Invariant recognition of natural objects in the presence of shadows, *Perception*, **29**(4), 383-398.

Braje, W.L., Kersten, D., Tarr, M.J. and Troje, N.F. Illumination effects in face recognition. (1998) *Psychobiology*. **26**(4), 371-380.

Braun, J. (1993). Shape-from-shading is independent of visual attention and may be a 'texton'. *Spatial Vision*, **7**, 311-322.

Brinkmann, R. (1999). *The art and science of digital compositing*. Morgan Kaufmann Publishers

Cameron, P. A. & Gallup, G. G. (1988). Shadow recognition in human infants. *Infant Behavior & Development*, **11**, 465-471.

De Haan, E., Erens, R.G.F., & Noest, A. J. (1995). Shape from shaded random surfaces. *Vision Research*, **35**, 2985-3001.

Enns, J. T. & Rensink, R. A. (1990). Influence of scene-based properties on visual search. *Science*, **247**, 721-723.

Erens, R. G. F., Kappers, A. M. L. & Koenderink, J. J. (1993). Perception of local shape from shading. *Perception & Psychophysics*, *54*, 145-157.

Fielding, R. (1985). *The technique of special effects cinematography*. Focal Press, London.

Gombrich, E. H. (1995). *Shadows: The Depiction of Cast Shadows in Western Art*. National Gallery Publications, London.

Hagen, M. A. (1976). The development of sensitivity to cast and attached shadows in pictures as information for the direction of the source of illumination. *Perception & Psychophysics*, **20**, 25-28.

Hietanen, J. K., Perrett, D. I., Oram, M. W., Benson, P. J. & Dittrich, W. H. (1992). The effects of lighting conditions on responses of cells selective for face views in the macaque temporal cortex. *Experimental Brain Research*, **89**, 157-171.

Johnston, A., Hill, H. & Carman, N. (1992). Recognising faces: Effects of lighting direction, inversion, and brightness reversal. *Perception*, **21**, 365-375.

Kleffner, D. A. & Ramachandran, V. S. (1992). On the perception of shape from shading. *Perception & Psychophysics*, **52**, 18-36.

Langer, M. S. & Zucker, S. W. (1997). Casting light on illumination: A computational model and dimensional analysis of sources. *Computer Vision and Image Understanding*, **65**, 322-335.

Pentland, A. P. (1982). Finding the illuminant direction. *Journal of the Optical Society of America*, **72**, 448-455.

Ramachandran, V. S. (1988). Perception of shape from shading. *Nature*, **331**, 163-166.

Sun, J. & Perona, P. (1996a). Early computation of shape and reflectance in the visual system. *Nature*, **379**, 165-168.

Sun, J. & Perona, P. (1996b). Preattentive perception of elementary three-dimensional shapes. *Vision Research*, **36**, 2515-2529.

Sun, J. & Perona, P. (1997). Shading and stereo in early perception of shape and reflectance. *Perception*, **26**, 519-529.

Tarr, M. J., Kersten, D., & Bülthoff, H. H. (1998). Why the visual system might encode the effects of illumination. *Vision Research*, **38**, 2259-2275.

Todd, J. T. and Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. Journal of Experimental Psychology: Human Perception and Performance, **9**, 583-595.

Triesman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics and Image Processing*, **31**:156-177.

Weinshall, D. (1994). Local shape approximation from shading. *Journal of Mathematical Imaging and Vision*, **4**, 119-138.

Yonas, A., Goldsmith, L. T. & Hallstrom, J. L. (1978). Development of sensitivity to information provided by cast shadows in pictures. *Perception*, **7**, 333-341.

Yonas, A., Kuskowski, M. & Sternfels, S. (1979). The role of frames of reference in the development of responsiveness to shading information. *Child Development*, **50**, 493-500.