Perception, 2004, volume 33, pages 1279 – 1290

DOI:10.1068/p5280

Impossible shadows and the shadow correspondence problem

Pascal Mamassian
Psychology Department, University of Glasgow, 58 Hillhead Street, Glasgow G12 8QB, Scotland, UK; e-mail: pascal@psy.gla.ac.uk
Received 1 February 2004, in revised form 15 March 2004; published online 10 November 2004

Abstract. Shadows cast by objects contain potentially useful information about the location of these objects in the scene as well as their surface reflectance. However, before the visual system can use this information, it has to solve the shadow correspondence problem, that is to match the objects with their respective shadows. In the first experiment, it is shown that the estimate of the light source position is affected by a gradual luminance ramp added to the image. In the second experiment, it is shown that observers process impossible shadow images as if they ignored the local features of the objects. All together, the results suggest that the visual system solves the shadow correspondence problem by relying on a coarse representation of the scene.

1 Introduction

There are no shadows without objects. Yet the problem of matching an object with its cast shadow is not trivial. Cast shadows provide potentially rich information about the visual scene (Casati 2000), in particular with respect to the location of objects in the scene (Mamassian et al 1998). However, in order to use that information, our visual system has first to segment regions in the image, decide that these regions are potential shadows rather than, say, ink blots, and then match these shadow candidates with objects in the scene. We call this last processing stage the "shadow correspondence problem". It is reminiscent of the correspondence problem in stereopsis or motion perception where one has to match features between the left and right images, or between consecutive frames of a movie (Marr 1982). If the shadow correspondence problem is not solved, the observer will be confused by spurious dark patches in the image and will be unable to use cast shadows for spatial layout. In contrast, if the shadow correspondence problem is solved, it will constrain the illumination conditions, in particular the location of light sources, and that in turn will help in inferring shape from shading (eg Koenderink et al 1996; Mamassian and Kersten 1996; Pollick et al 1996) and solving lightness constancy (Gilchrist et al 1999).

The difficulty in solving the shadow correspondence problem lies in its computational complexity. First, the problem is underconstrained, that is there is no guarantee that a search through the image to find the best matches between objects and shadows will result in a unique unambiguous solution. Some objects and shadows might also be partially occluded or even absent in the image, and this will make a point-by-point matching impossible. Second, the matching involves finding corresponding points on contours that do not necessarily lie in a plane (neither on the casting object nor on the surface receiving the shadow). While planar contours projected onto a plane preserve their shape up to an affine transformation, non-planar contours projecting to a plane or planar contours projecting to a curved surface can undergo very complex deformations. Finally, even if there is a unique solution, an exhaustive search to find that solution might take too much time for it to be of any practical use. For these reasons, it may be more beneficial to use a heuristic that will allow the visual system to find a quick solution to the shadow correspondence problem. One such heuristic is to use a coarse matching between objects and shadows. If fine details are discarded, the correspondence problem is necessarily simplified. The evidence for such a heuristic is here investigated in two separate experiments.

2 Experiment 1: Illumination direction

If observers use a coarse representation to process shadows, then they will be affected by global manipulations of the image (ie manipulations that will affect the low spatial-frequency content of the image). This prediction was tested by manipulating images of natural scenes in such a way that this manipulation would either confirm or contradict the illumination direction of the original image. Natural scenes were chosen that were clearly lit by a single light source located either on the left or on the right side of the scene. The manipulation consisted in combining the image with a luminance ramp, either a decrease or an increase of luminance from left to right. The motivation to introduce this luminance ramp was to simulate the effects of adding another extended light source in the scene, either on the left or on the right. The observers were then asked to report where they thought the light was coming from, left or right. If a coarse representation is used, observers should be impaired when the luminance ramp is in conflict with the original image, and should improve when it is consistent.

Figure 1. Examples of images used in experiment 1. All scenes were lit by a single natural light source and observers had to estimate the direction of illumination (left for these four images).

2.1 *Methods*

Five undergraduate students from the University of Glasgow participated in the experiment. They had normal or corrected-to-normal visual acuity, and were naïve as to the purpose of the experiment.

24 images of natural scenes were chosen from the Corel database, half of them landscapes, the other half objects such as buildings and animals (figure 1). The images were cropped to a square (each side subtended 3.7 deg or 256 pixels) and were converted to grey-scale.

The images were manipulated by combining each image with a vertically oriented luminance ramp (figure 2). The luminance ramp could be either consistent (from bright to dark from left to right if light was coming from the left), inconsistent with the original light direction (from dark to bright), or absent (no luminance ramp added). The combination consisted in adding a linear luminance ramp to the original image and renormalising the image such that its range of intensities matched that of the original image. In addition, each image was presented either in its original orientation or flipped about the vertical axis, so that on half of the trials light came from the left, and on the other half light came from the right. There were therefore two independent variables of interest: luminance ramp (3 levels) and image orientation (2 levels).

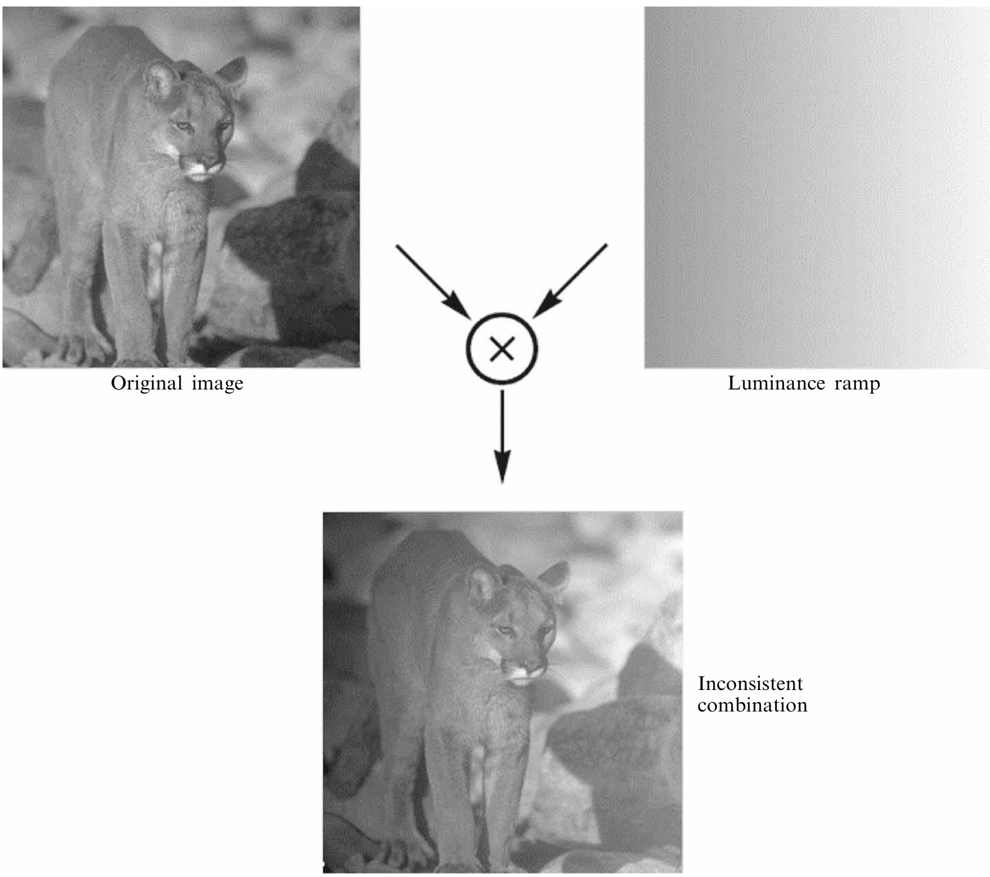


Figure 2. Luminance ramp manipulation. The manipulation of the images consisted in combining the original image with a luminance ramp that overlapped the whole image. The luminance ramp could be either consistent or inconsistent with the light direction of the original image. This figure illustrates the result of an inconsistent combination.

Stimuli were presented on a 17 inch Sony Trinitron monitor whose colour lookup table was linearised. Each image was shown for 200 ms and then followed by a mask (filtered noise image). Observers were asked to judge whether light was coming from the left or from the right. While this task might appear odd at first, previous studies have shown that observers have no difficulty to report the direction of illuminant (Todd and Mingolla 1983). A within-subject design was used, where each observer saw in random order every combination of the luminance ramp and image orientation in a block of trials. Each observer ran two such blocks, thereby providing 288 judgments of light direction (24 images $\times$ 3 luminance ramps $\times$ 2 image orientations $\times$ 2 blocks of trials).

2.2 *Results and discussion*
The proportion of times observers correctly identified the light direction was computed separately for each luminance ramp and image orientation conditions (figure 3). In this analysis, the correct light direction is taken to be the one that is consistent with the unmodified image. There was a main effect of the luminance ramp due to more accurate responses when the added luminance ramp was consistent rather than inconsistent with the original light direction ($F_{2,24} = 6.74$, $p < 0.005$). Observers were also more accurate when the image was oriented such that light came from the left rather than right ($F_{1,24} = 4.23$, $p = 0.05$). There was no interaction between luminance ramp and image orientation.

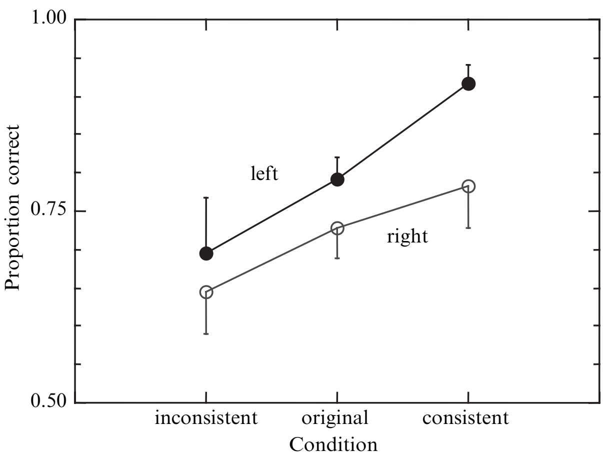


Figure 3. Results for experiment 1. Proportion correct is plotted against the experimental condition which determined whether the image was presented in its original form or combined with a luminance ramp that was either consistent or inconsistent with it. Images where light came from the left are distinguished from the ones where light came from the right. Error bars are standard errors across observers.

The effect of image orientation is surprising given that it was exactly the same images that were presented, except for a flip about the vertical axis. Because the images were presented in random order, and the order was different for different observers, this finding cannot be attributed to a sequential effect. In spite of its oddity, this better performance for left-lit images is consistent with our previous work that showed a preference for above-left illumination to interpret ambiguous images (Mamassian and Goutcher 2001; Mamassian et al 2003).

The analysis of the scenes that led to the best and worst performance is revealing. The best stimuli usually contained one dominant large shadow (eg the shadow cast by a building), whereas the worst stimuli contained multiple small shadows (eg shadows of sand dunes in a desert scene). This observation is the first hint that observers were using mainly the coarse scale of the image to determine the light direction. This strategy would explain why the luminance ramp that covered the whole image produced such a strong effect on accuracy. It is worth noting, though, that observers were always above chance performance even when the luminance ramp was inconsistent with the shadows.

This indicates that observers were still able to rely mostly on shadows in spite of the sometimes contradictory information from the luminance ramp.

A simple algorithm can be proposed to extract the light direction from the stimuli (figure 4). Imagine the image of a mountain peak lit from the left. At a coarse scale, the image will display a predominant vertical edge whose left side is brighter than its right side (a light – dark edge). This observation generalises to any image of convex objects. Collecting the presence of edges and their polarity will therefore be a crude way to estimate where the light is coming from. Detecting such edges is equivalent to convolving the image with an oriented filter (Marr and Hildreth 1980). The performance of such a model was overall fairly high in spite of its simplicity (up to 75% correct on the set of stimuli used here).

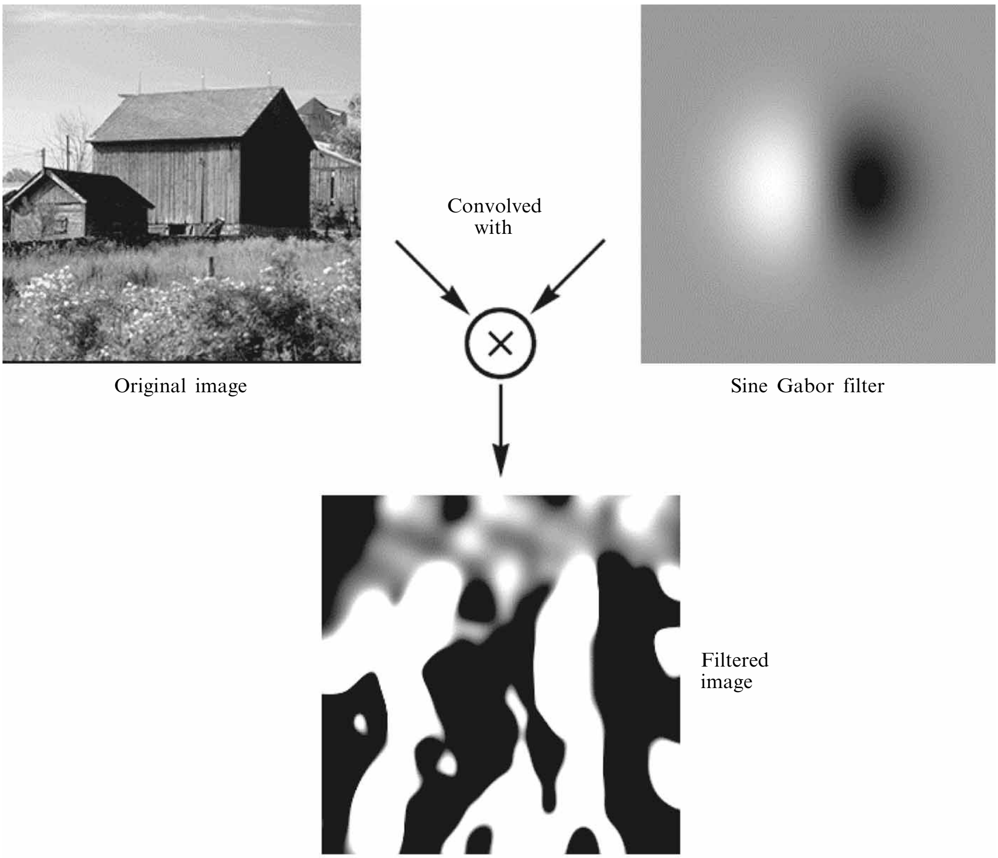


Figure 4. Simple edge-detection model to estimate the light direction. An image is filtered with an oriented sine Gabor filter to discriminate light – dark from dark – light edges. White areas indicate a positive response, black areas a negative one. The model decides that light comes from the left if the positive areas dominate the negative ones.

The filters used in the model approximate well the receptive fields of simple cells in primates' primary visual cortex (area V1). Therefore, one could argue that determining the direction of light can be undertaken by a bottom – up mechanism starting as early as V1. However, V1 receptive fields are certainly too small to support the claim that shadows are computed at a coarse scale (see also next experiment). Also, if such a model bears any resemblance to the algorithm used by the visual system, one can wonder whether there is any bias in the distribution of cells that respond to bright – dark edges as opposed to dark – bright edges. More physiological work would be helpful to address these questions.

3 Experiment 2: Shadow correspondence problem

A coarse representation of the image preserves the general shape and location of objects but discards all fine details. Matching objects with their cast shadows is therefore much simplified at a coarse scale. This second study was designed to investigate the way observers solve this shadow correspondence problem when the shadow shape is sometimes inconsistent with the casting object. Such impossible shadow images will allow us to test whether observers use a more local or global strategy when they solve the shadow correspondence problem. Even though the distinction between local and global is different from the one between fine and coarse scales (in particular, fine scales can still be processed globally), it is unlikely that a mechanism that relies purely on a coarse representation can be precise enough to use local information. Therefore, if observers use a coarse scale, they would be expected to use a global strategy.

3.1 *Methods*

Ten undergraduate students from the University of Glasgow participated in the experiment. They had normal or corrected-to-normal visual acuity, and were naïve as to the purpose of the experiment.

The stimuli were images of simple wire-frame objects and their cast shadows (figure 5) that were presented on the same CRT monitor as the one used in the previous experiment. The objects consisted of two parts, a vertical stem and a cap that was parallel to the ground plane. The stem was in contact with the ground, such that there was no space between the shadow and the object. The shadow was always in front of the object, thereby simulating a light source behind the object. The object was not shaded so that light direction could only be estimated from the cast shadow. No cue about the ground plane was provided other than the presence of the shadow.

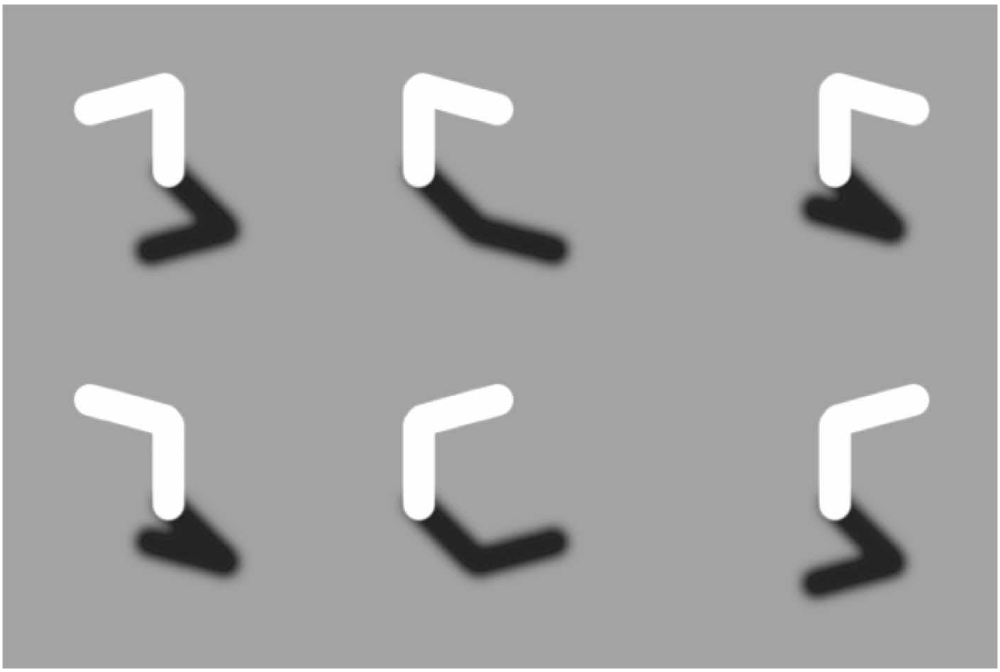


Figure 5. Examples of images used in experiment 2. The four images on the left present a shadow whose shape is consistent with the white object, whereas the two images on the right present shadows inconsistent with the objects (ie impossible-shadow images).

The independent variable was the length of the cap, and this length could take one of five values. Three other aspects of the stimulus were also manipulated so that the observer could not learn to recognise individual images. First, the light source was simulated to be either behind-left or behind-right of the object's base. Second, the cap of the object could take one of four orientations: front-right, back-right, back-left, or front-left. Third, the whole image (object and shadow) was displayed randomly in one of eight positions at an equal eccentricity from the fixation point.

On two-thirds of the trials, the shadow was consistent with the object, but in the remaining third of the trials, object and shadow were inconsistent. These impossible-shadow images were obtained by replacing the object's shadow with that of the object whose cap had the opposite orientation (see again figure 5 for some examples). One important point to note is that through this manipulation object and shadow were still piece-wise consistent but globally inconsistent; in particular, the shadow of the cap was still parallel to the cap of the object. Stimuli were displayed for 80 ms and then followed by a mask. At these brief presentation times, most observers do not even notice that some images contained impossible shadows. The experiment was run in a darkened room. The task of the observer was again to estimate whether the light was coming from the left or the right. Each observer judged 1440 stimuli presented in random order.

Three simple models of the correspondence between the object and its shadow can be proposed. These models attempt to match either a specific feature of the object or a coarser description of the object. Model A matches the stem of the object with the shadow cast by the stem, model B matches the cap of the object with the shadow cast by the cap, and model C matches the centre of mass of the object with the centre of mass of the shadow (figure 6). When the shadow is consistent with the object, all three models indicate a same light source direction. For the impossible-shadow stimuli, however, the three models differ in their predicted light directions, especially as the cap length changes. More specifically, model A is not affected by the cap length because the correspondence is based purely on the stem; model B predicts a light direction that increasingly deviates from that of model A as cap length increases; finally, model C predicts a light tilt direction in between that from models A and B because both the stem and the cap are taken into consideration to compute the centre of mass (figure 7). The behaviour of the three models can be summarised by the predicted

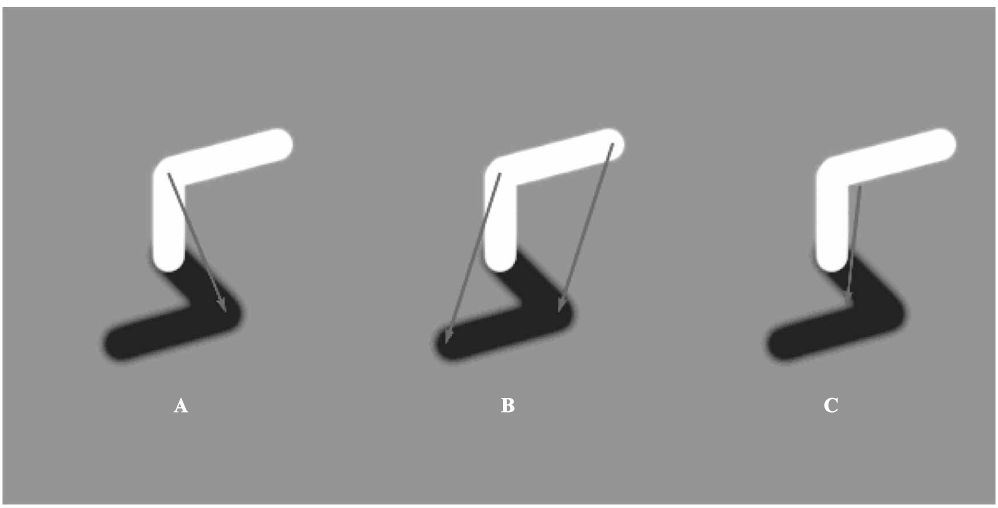


Figure 6. Shadow correspondence for impossible-shadow stimuli. Three predictions of shadow correspondence are based on matching the stem (A), the cap (B), or the centre of mass (C).

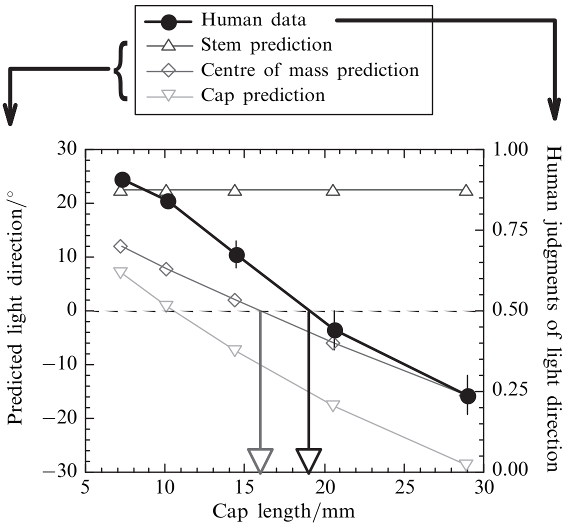


Figure 7. Results for the shadow correspondence for impossible-shadow stimuli. The three predictions for the light direction are plotted against the cap length (left scale). The proportion of human judgments of light direction (consistent with a stem correspondence) is also plotted against cap length (right scale). The human psychometric function intersects the horizontal midline close to the prediction based on the centre of mass. Error bars are standard errors across observers.

cap length that leads to an illumination directly overhead (the point where the curve intersects the 0° light direction in figure 7). For such a cap length, an algorithm designed to report whether light is coming from the left or right will perform at chance. These predicted cap lengths can therefore be compared to the point of subjective equality in this psychophysical task, that is the cap length for which the observers responded equally often that light was coming from the left and right.

3.2 *Results and discussion*

The analysis was restricted to the impossible-shadow images since the three models made different predictions only for these stimuli. For each object's cap length, the proportion of times that observers reported a light direction that was consistent with the shadow cast by the stem of the object was computed. These proportions were found to be significantly affected by the cap length, thereby discarding a strategy analogous to model A which is purely based on the stem. The point of subjective equality which corresponds to the cap length for which observers reported equally often light coming from the left or right was then computed. This point of subjective equality for the human observers was fairly close to the cap length leading to overhead illumination predicted from model C. In other words, participants followed a shadow correspondence strategy close to one based on matching the centre of mass of the object with the centre of mass of the shadow. Such a strategy suggests that the visual system is relying on a coarse representation and discards details of the shadow shape to match objects with their shadows.

One could, of course, argue that the results reported here are also consistent with a mixture of models A and B where the observer would attempt to match separately the stem and the cap with their respective shadow parts. This alternative strategy would then require the observer to reconcile the resulting conflicting information about the light source direction. It seems that the observers would have realised that some stimuli contained impossible shadows if they had followed this strategy. Instead, it was necessary to explain to the observers why some images displayed impossible shadows after they completed the experiment. Therefore, while the possibility that observers used a mixture of models based on local feature matching cannot be totally ruled out, the more parsimonious interpretation that they used a coarse mechanism to match objects and shadows is preferred.

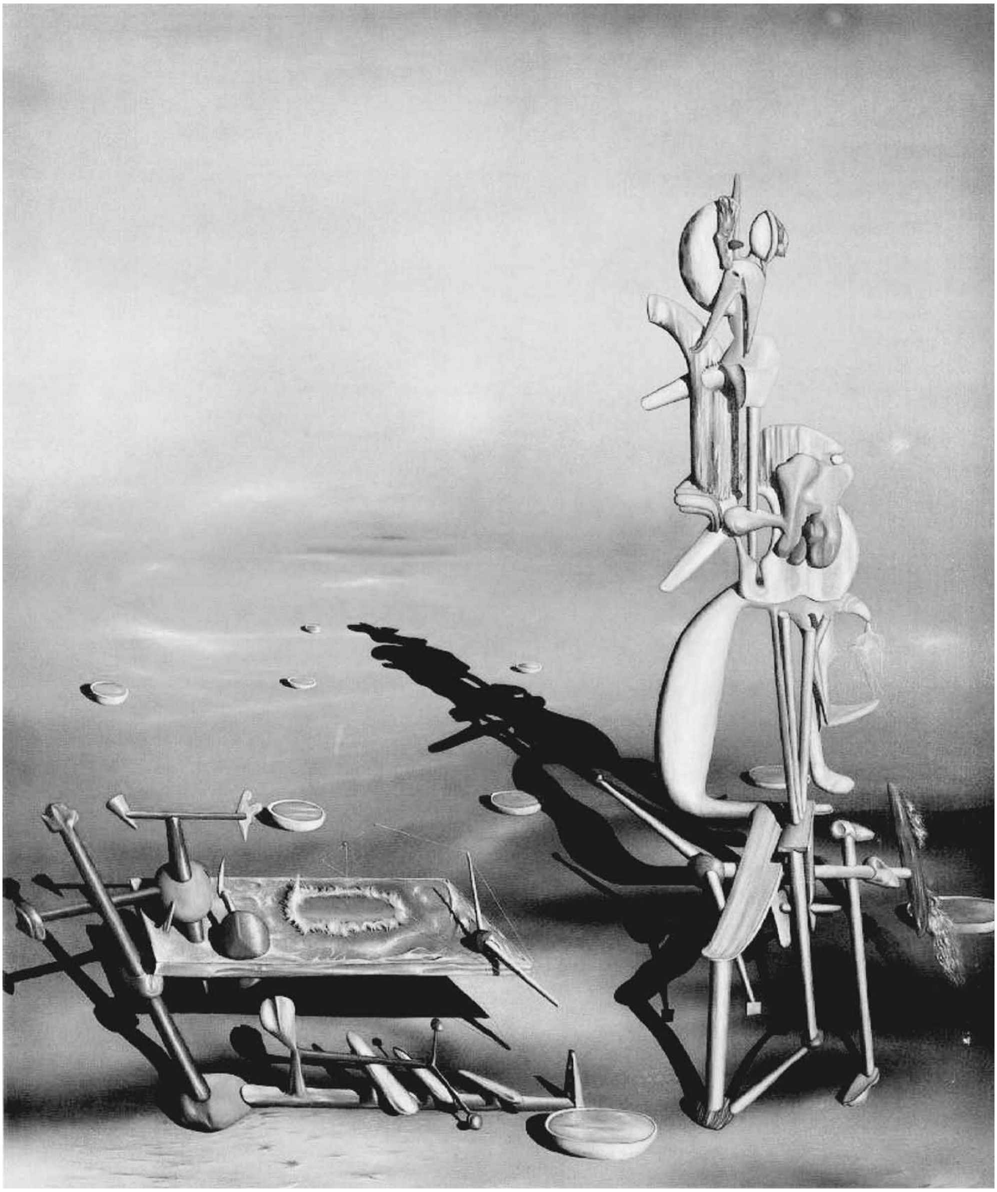
4 General discussion

In this paper, the shadow correspondence problem that the visual system has to solve to match each shadow in the image to an object has been introduced. While this problem is complex, it is proposed that a quick solution can be obtained if a coarse representation is used and if fine details are discarded. In the first experiment, it was shown that a gradual change of luminance across the image had a profound effect on the estimated light source position. In the second experiment, it was found that observers perceived impossible shadow images as if they were processed at a coarse scale. All together, the results suggest that the shadow correspondence problem is solved by a mechanism based on the coarse-scale rather than fine-scale information in the image. This mechanism would allow the observers to obtain a first quick interpretation of the scene that can then be re-evaluated subject to finer details in the scene and other high-level processes.

While it has been argued that shadows are processed at a coarse scale, it has not been claimed that they are totally ignored. Most studies in object recogniton have indeed found that changing the illumination conditions (Tarr et al 1998) or the shape of the shadow (Castiello 2001) slows down recognition (although see Braje et al 2000). Similarly, studies in visual search have generally found that objects with anomalous shadows are easy to detect (Enns and Rensink 1990; Sun and Perona 1996; although see Ostrovsky et al 2001). These studies clearly indicate that shadows are processed and participate to a full 3-D reconstruction of the objects in the scene (Cavanagh and Leclerc 1989; Norman et al 2000).

There is a similarity between the impossible-shadow stimuli used here and the popular impossible objects such as the Reutersvärd – Penrose impossible triangle (Penrose and Penrose 1958; Draper 1978; Gregory 1998). In both cases, the images are locally consistent but globally impossible (unless they represent objects seen from a very accidental viewpoint). Also, in both cases, the impossibility is not noticed at first, and a careful scrutiny of the image seems to be required for the subject to be convinced (see Huffman 1971 for an algorithm to label impossible objects).

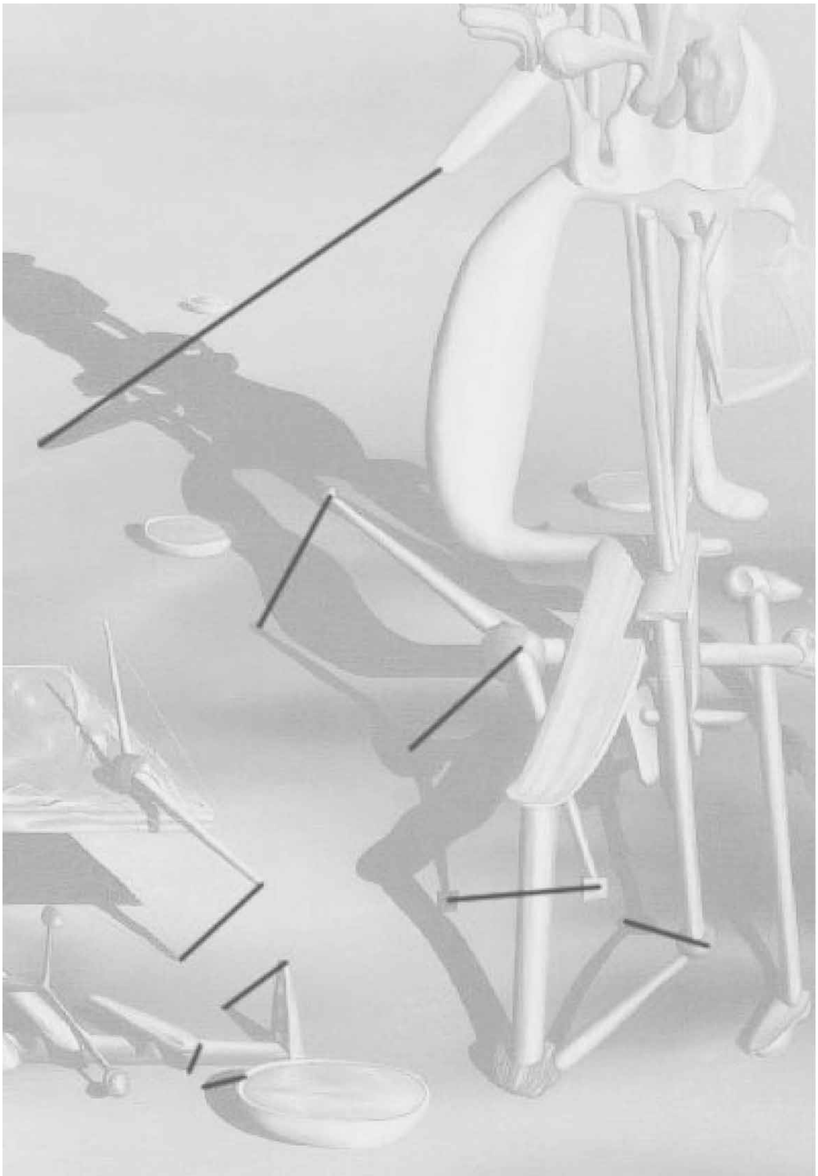
If our visual system is indeed using a quick mechanism to solve the correspondence problem, then it should be easily fooled by impossible shadows. Such images actually abound in Western art paintings, even in those from artists who mastered the art of shadows, such as Salvador Dali, Giorgio de Chirico, or Yves Tanguy. Figure 8a shows the exquisite *Indefinite Divisibility* by Yves Tanguy, a typical example of one of his surrealist desert landscapes where nonsense objects cast long shadows. The shadows have high contrast and their contours seem to follow faithfully the intrinsic details of the casting objects. However, the positions of these shadows are inconsistent with a unified light source. Figure 8b shows the solution of the shadow correspondence problem for a few selected details. If a single light source was used, all the lines representing light rays between objects and shadows should be parallel to each other, and they are not. One may argue that Tanguy was aware of these inconsistencies, and was using them explicitly to emphasise the surrealist impression, in the same way that de Chirico used irrational perspective. Alternatively, it could simply be that Tanguy did not care so much about the location of his shadows because he knew that the viewer would not notice the inconsistencies. This latter interpretation is consistent with our hypothesis that observers use only a coarse representation to solve the shadow correspondence problem.



(a)

Figure 8. (a) *Indefinite Divisibility* by Yves Tanguy (1942), collection of the Albright-Knox Art Gallery, Buffalo, NY. (b) An analysis of the shadows in *Indefinite Divisibility* reveals that they are not consistent with a single light direction.

Acknowledgments. This manuscript is a summary of work presented at various conferences, in particular experiment 1 was presented at ECVP 1999 (Trieste, Italy), experiment 2 at ARVO 1999 (Fort Lauderdale, USA), and the analysis of surrealist paintings at the symposium on "Shadows: from Art to Neurosciences" in November 2003 (Rovereto, Italy). I appreciated the financial supports of the Royal Society of London, the Programme Alliance between UK and France, a Cognitique grant from the French Ministry of Education, and an invitation from Umberto Castiello to the symposium in Rovereto. I would like to thank Roberto Casati and Jeanny Hérault for stimulating discussions, as well as the reviewers for their comments.



(b)

Figure 8 continued.

References

Braje W L, Legge G E, Kersten D, 2000 "Invariant recognition of natural objects in the presence of shadows" *Perception* **29** 383–398

Casati R, 2000 *La Scoperta dell'Ombra* (Milan: Mondadori)

Castiello U, 2001 "Implicit processing of shadows" *Vision Research* **41** 2305–2309

Cavanagh P, Leclerc Y G, 1989 "Shape from shadows" *Journal of Experimental Psychology: Human Perception and Performance* **15** 3–27

Draper S, 1978 "The Penrose triangle and a family of related figures" *Perception* **7** 283–296

Enns J T, Rensink R A, 1990 "Influence of scene-based properties on visual search" *Science* **247** 721–723

Gilchrist A, Kossyfidis C, Bonato F, Agostini T, Cataliotti J, 1999 "An anchoring theory of lightness perception" *Psychological Review* **106** 795–834

Gregory R L, 1998 *Eye and Brain: The Psychology of Seeing* fifth edition (Oxford: Oxford University Press)
Huffman D, 1971 "Impossible objects as nonsense sentences", in *Machine Intelligence* Eds B Meltzer, D Michie (Edinburgh: Edinburgh University Press) pp 295 – 323
Koenderink J J, Kappers A M L, Todd J T, Norman J F, Phillips F, 1996 "Surface range and attitude probing in stereoscopically presented dynamic scenes" *Journal of Experimental Psychology: Human Perception and Performance* **22** 869 – 878
Mamassian P, Goutcher R, 2001 "Prior knowledge on the illumination position" *Cognition* **81** B1 – B9
Mamassian P, Jentzsch I, Bacon B A, Schweinberger S, 2003 "Neural correlates of shape from shading" *NeuroReport* **14** 971 – 975
Mamassian P, Kersten D, 1996 "Illumination, shading and the perception of local orientation" *Vision Research* **36** 2351 – 2367
Mamassian P, Knill D C, Kersten D, 1998 "The perception of cast shadows" *Trends in Cognitive Sciences* **2** 288 – 295
Marr D, 1982 *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (New York: W H Freeman)
Marr D, Hildreth E C, 1980 "Theory of edge detection" *Proceedings of the Royal Society of London, Series B* **207** 187 – 217
Norman J F, Dawson T E, Raines S R, 2000 "The perception and recognition of natural object shape from deforming and static shadows" *Perception* **29** 135 – 148
Ostrovsky Y, Cavanagh P, Sinha P, 2001 "Perceiving illumination inconsistencies in scenes" MIT Artificial Intelligence Laboratory Memo No 2001-029, Massachusetts Institute of Technology, Cambridge, MA, USA
Penrose L S, Penrose R, 1958 "Impossible objects: A special type of illusion" *British Journal of Psychology* **49** 31 – 33
Pollick F E, Watanabe H, Kawato M, 1996 "Perception of local orientation from shaded images" *Perception & Psychophysics* **58** 762 – 780
Sun J, Perona P, 1996 "Preattentive perception of elementary three-dimensional shapes" *Vision Research* **36** 2515 – 2529
Tarr M J, Kersten D, Bülthoff H H, 1998 "Why the visual recognition system might encode the effects of illumination?" *Vision Research* **38** 2259 – 2275
Todd J T, Mingolla E, 1983 "Perception of surface curvature and direction of illuminant from patterns of shading" *Journal of Experimental Psychology: Human Perception and Performance* **9** 583 – 595

ISSN 0301-0066 (print) ISSN 1468-4233 (electronic)

PERCEPTION

VOLUME 33 2004

www.perceptionweb.com