Perception, 2004, volume 33, pages 1101 – 1127

DOI:10.1068/p5265

Psychophysics with junctions in real images

Josh McDermott
Perceptual Science Group, Massachusetts Institute of Technology, NE20-444, 3 Cambridge Center, Cambridge, MA 02139, USA; e-mail: jhm@mit.edu
Received 23 October 2003, in revised form 29 March 2004; published online 13 October 2004

Abstract. Junctions, formed at the intersection of image contours, are thought to play an important and early role in vision. The interest in junctions can be attributed in part to the notion that they are local image features that are easy to detect but that nonetheless provide valuable information about important events in the world, such as occlusion and transparency. Here I test the notion that there are locally defined junctions in real images that might be detected with simple, early visual mechanisms. Human observers were used as a tool to measure the visual information available in local regions of real images. One set of observers was made to label all the points in a set of real images where one edge occluded another. A second set of observers was presented with variable-size circular subregions of these images, and was asked to judge whether the regions were centered on an occlusion point. This task is easy if junctions are visible, but I found performance to be poor for small regions, not approaching ceiling levels until observers were given fairly large (~50 pixels in diameter) regions over which to make the judgment. Control experiments ruled out the possibility that the effects are just due to junctions at multiple scales. Experiments reported here suggest that, although some junctions in real images are locally defined and can be detected with simple mechanisms, a substantial fraction necessitate the use of more complex and global processes. This raises the possibility that junctions in such cases may not be detected prior to scene interpretation.†

1 Introduction

Vision can be characterized as the task of inferring scenes in the world from images. This inference process involves estimating a probability distribution of scenes given an observed image, and perhaps searching through the space of scenes to find the one that is in some sense best given the distribution. The problem, of course, is that the tasks of directly estimating this distribution and finding an optimum are intractable owing to the high dimensionality—a huge number of parameters is required to describe an entire scene. A more promising alternative may be to estimate many reduced versions of this distribution at different locations in an image, ie to estimate at each location the distribution over the local scene parameters (surface curvature, reflectance, etc) given the image region at that location. Such local distributions are presumably more tractable to estimate and use because of the smaller number of parameters involved. Locally the distribution of scene structure given the image is less peaked and unimodal because there is less information to constrain scene interpretation. However, the local distributions could narrow the possible interpretations and might then be combined to yield a global scene estimate. Methods have recently been suggested that accomplish this in some simple cases (eg Saund 1999; Freeman et al 2000).

Something akin to this general strategy seems likely to play a large role in vision; the big search involving the entire image is too hard, and yet we manage to see astoundingly well, which implies that shortcuts are available and are taken by biological visual systems. The hierarchical nature of mammalian visual systems may also be reflective of such a strategy. The main question, then, is just whether the local processing

† The results described here were first published in thesis form in 2000 (McDermott 2000) and then presented at the 2002 Vision Sciences Society Meeting (McDermott 2002).

will be intelligible, ie whether vision science will be able to produce a blueprint for what happens locally. It might be the case that the local processes are messy and difficult to make sense of, but the assumption implicit in much of vision research is that this is not the case. It may be for this reason that there is much interest in local cues such as junctions.

Junctions, formed at the intersections of image contours, are attractive candidates for elements of early vision because they are presumed to be local image features that nonetheless provide information about important events in the world. Junctions have been suggested to be involved in the recovery of surface occlusion geometry (T-junctions, Guzman 1968; L-junctions, Rubin 2001), shape (Malik 1987), transparency (X-junctions, Adelson and Anandan 1990; T-junctions, Anderson 1997), and reflectance, illumination, and atmospheric boundaries (Adelson 1993, 2000). Furthermore, junctions lend themselves to appealingly principled and rule-driven models of many aspects of mid-level vision (eg Sajda and Finkel 1995).

The issue addressed in this paper is whether there are in fact junction-like structures in real images that are locally detectable and could serve as building blocks for early vision. Although a growing body of research is consistent with an important role for local junction analysis in mid-level vision, virtually all the psychophysical and modeling work on junctions is conducted with synthetic images composed of discrete luminance values and step edges. It is unclear to what extent the junctions in such stimuli are representative of those in real images, which are noisier and more complex in many ways. Although it is easy to point out junctions in real images, when we do so we are making use of the entire image and our entire visual system. It is unclear whether most of these junctions can be locally detected, or instead become detectable only after a larger region of the image is interpreted.

One obvious approach to investigating this issue is to design a mechanism for detecting junctions, in that the success of a local mechanism at detecting junctions in real images would imply that they are indeed locally defined. However, the results of several previous such attempts have been discouraging (Beymer 1991; Freeman 1992; Heitger et al 1992; Perona 1992; Alvarez and Morales 1997; Lindeberg 1998; Parida et al 1998). Sensible computations involving oriented filtering and local templates result in apparently poor performance on real images. This is perhaps surprising, given that much of the interest in junctions is predicated on their apparent simplicity.

These results might be taken to indicate that the information in local regions of real images is insufficient to detect junctions. There is, however, an inherent difficulty in claiming on the basis of failures in machine vision that a task cannot be solved with purely local information. Even when a number of sensible approaches fail, it is hard to know whether the local mechanisms simply have not been optimized. The available options are to keep trying to develop better low-level mechanisms, or to resign oneself to designing a more complicated and global computation. This quandary is particularly evident in efforts to detect edges. Initial efforts focused on the design of a local operator (Marr and Hildreth 1980; Canny 1986), but disappointing results led to a variety of nonlocal post-processes to clean up edge maps (Canny 1986; Zucker et al 1988; Shashua and Ullman 1990). More recently, there have been further attempts to refine the local operators (Iverson and Zucker 1995). Although it is generally agreed that some degree of nonlocal post-processing is necessary to generate edge maps suitable for further analysis (eg inferring shape from contours), it is not clear how well a local operator should be expected to perform, in part because they are typically evaluated by looking at their outputs in the context of an entire image. It is thus difficult to ascertain the extent to which more sophisticated, nonlocal analysis is needed.

In this paper an alternative approach is adopted: psychophysics with real images as stimuli. Experiments with real images can complement machine vision and traditional human psychophysics by showing what human observers can and can not do given some fraction of the information available in real images. Simple manipulations (filtering, masking, etc) can be used to restrict the information available to an observer. An observer's competence at a task is proof that it can be done given the restrictions, whereas failure suggests that it cannot.

The goal of the experiments reported here was to examine the extent to which occlusion in real images produces simple, local, junction-like cues that might be easily detected by early visual mechanisms. The main idea is straightforward: if junctions are locally defined, then observers should be able to reliably detect them given only a small image region centered on the junction.

To test this idea, I first needed a large set of junction candidates to use as stimuli. Junctions are, by hypothesis, often generated by the occlusion of one edge by another, so this is where I looked for them. I restricted my study to junctions generated at points of occlusion, because this is by far the most common junction-generating event in real scenes; such points were plentiful enough to study in some detail. It was also fairly easy to describe to observers how they should label a point of occlusion. I thus began by having naïve observers identify points of occlusion in a set of real images.

2 Stimulus acquisition

Observers were asked to label the points in an image where "one edge occludes another", a task that tended to be intuitive and natural for naïve observers to do. Their instructions were as follows:

> "In each image you should label the 'points of occlusion'. A point of occlusion is defined as a point in an image where one edge is occluded by (ie goes behind) another. The occluding edge will be the border of a surface or object, but the occluded edge (ie the one that goes behind) can be any sort of edge (eg a surface marking such as paint, the borders of shadows, etc, in addition to the borders of surfaces and objects)."

Observers labeled points of occlusion by clicking with a mouse on the pixel that seemed to be closest to the perceived intersection of the edges. They could then adjust the cursor location by pressing keys on the keyboard in order to localize the occlusion point as precisely as possible. In some cases, points of occlusion were formed by lines rather than edges. When such lines were so thin as to make it impossible to distinctly label the two points of occlusion occurring on each side of the line, subjects were instructed to take the center of the line to be the point of occlusion.

Such locations were obtained for 20 real images taken from the online database of van Hateren and van der Schaaf (1998) consisting of images of various locations in the Netherlands taken with a digital camera. Images were logarithmically transformed and scaled for display purposes. I selected images which appeared to have lots of occlusion points. Some contained buildings and other man-made objects, whereas others consisted mainly of trees and plants. Most images were labeled by two observers, and labels tended to be consistent across observers. I then sifted through the sets of locations to eliminate the occasional erroneous label as well as to ensure that the locations were centered as well as was humanly possible. In the end approximately 1200 occlusion points were accumulated. An example image with labels is shown in figure 1.

These labeled points all were locations where occlusion could be seen upon inspection of the entire image at once. Given the interest in whether locally detectable junctions are produced at these points, one naïve approach would be to window each occlusion point with an aperture and ask a second set of subjects whether they see a junction. Unfortunately, junction detection is far from a well-defined task; we do not know what a junction is or how to ask an observer to look for one. Instead, subjects were asked to guess if the apertured region they were shown was centered on an occlusion point or not.

Figure 1. An image with labeled occlusion points.

This task at least has the advantage of having a right answer, albeit one defined relative to the first set of observers. Moreover, if observers can see a junction-like structure at the center of the image region, the task ought to be easy. This occlusion-related task thus seemed a reasonable way to test for the existence of junctions in real images.

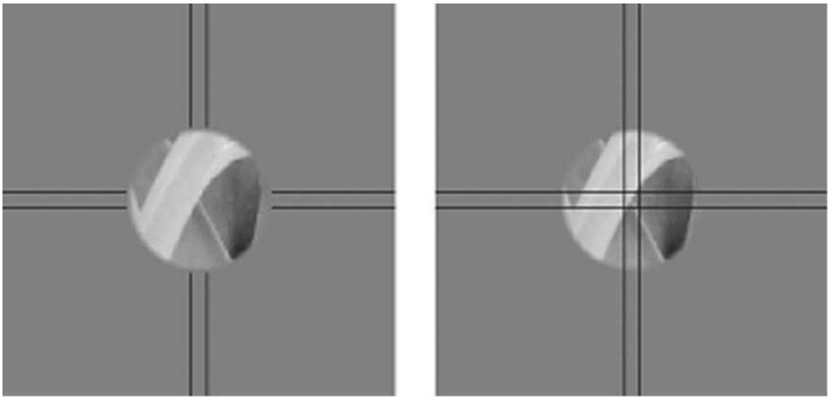
3 Experiment 1. Aperture size

In the first experiment, the effect of aperture size on subjects' ability to detect occlusion points was examined. Clearly they will do well on this task when the apertures are large, and the scene is readily interpreted; the question is what will happen when the apertures are small. If junctions are present at the occlusion points, performance should remain high.

Observers were presented with a succession of image patches of varying size. For each patch, they were asked to guess whether it was centered on an occlusion point. Some of the patches were centered on the previously labeled occlusion point locations, some were selected randomly from the same images, some were synthetic edges, and some were synthetic T-junctions and L-junctions. The synthetic junctions were included to ensure that, when crisp and well-defined junctions were present, subjects would indeed classify the image patch as an occlusion point.

3.1 *Methods*

An example of an experimental display is shown in figure 2a. Image patches were centered on a set of cross-hairs. Subjects were instructed to classify an image patch as being centered on an occlusion point if they thought there was an occlusion point within the cross-hairs of the display. The distance between the lines of the cross-hairs was 10 pixels, for all aperture sizes. This was a generous margin; I did not want uncertainty



(a) (b)

Figure 2. (a) An experimental display; patch is 97 pixels wide. (b) An experimental display with cross-hairs toggled in front.

about whether occlusion points were centred or not to cause classification errors. The cross-hairs could be moved in front of the image when localization was in doubt, as shown in figure 2b.

Subjects were given the following instructions:

> "The purpose of this experiment is to learn how the human visual system detects occlusion relationships in the world. We are particularly interested in whether the locations where one object occludes another can be detected using very local information in an image. In the first stage of this study, several observers labeled 'occlusion points' in a series of images of the world. In the present experiment, you will be presented with a series of image patches of varying size, taken from the images that were labeled in the first stage of the study. Some of these patches will be centered on the previously labeled occlusion points, while some will be centered on randomly selected locations. Of the latter, some will contain just a single edge, some will be regions of texture, some will be blank, and some will contain occlusion points that are not centered on the patch. You will be asked to guess whether a given patch is centered on an occlusion point. In some cases this will be obvious, and others less so."

The randomly selected patches were drawn randomly from the same images from which the occlusion points were obtained, subject to the constraint that the standard deviation of the intensities of the central 13 by 13 pixels exceeded 10 (intensities ranged from 0 to 255). This was to reduce the number of 'uninteresting' patches (eg those that were essentially uniform) that could be trivially distinguished from occlusion points. The synthetic edges and junctions were generated in MatLab. The different sectors of the junctions and edges were uniform in luminance and separated by linear boundaries. The intensities of the different sectors differed by at least 20 on the 0 – 255 scale, to ensure that the junctions and edges were well above threshold. The stimuli were generated at large sizes and blurred and subsampled to reduce aliasing. A small amount of noise was then added (normally distributed with a standard deviation of 5). Because they were so easy to see, the synthetic stimuli were only displayed at the three smallest aperture sizes. The proportions of stimuli in the experiments were as follows: 40% of the patches were centered on real point of occlusion, 45% were randomly selected patches of real images, and there were 5% each of synthetic edges, L-junctions, and T-junctions.

Patches were displayed at one of five sizes (13, 25, 49, 97, or 201 pixels in diameter) on a uniform background set to slightly above or below the mean luminance of the patch. This was achieved by multiplying each image region with a circular mask. At the borders, the mask faded gradually to zero over a 5 pixel radius increment.

(Note that as a result the effective size of each patch was roughly 4 pixels greater than the listed size.) Stimuli were generated and displayed in MatLab on a PC running Linux. Patches were displayed at full resolution on a Yama Vision Master Pro 450 monitor in ambient illumination. Subjects were instructed to sit at a comfortable viewing distance. This was usually a bit less than arm's-length, at which the smallest patch size subtended roughly 0.25 deg visual angle. On each trial the display was present until subjects entered their responses; viewing time was unlimited.

Subjects entered two responses on each trial. They were first asked to make their best guess whether the current patch was centered on an occlusion point or not, by pressing one of two buttons. They were then asked to make a confidence rating, pressing one of three buttons to indicate whether they were essentially guessing, somewhat sure, or quite sure that their previous judgment was correct. Subjects were instructed to envision a continuum of confidence and to assign thirds of that continuum to each button. Each response was confirmed with a second key press before which subjects had the option of changing the response. Subjects were debriefed after each experiment and consistently claimed to have made only a handful of such mistaken responses.

In the interest of helping subjects perform as well as possible, feedback was given following trials in which real-image patches were 'misclassified'. A tone was sounded if a labeled point of occlusion was not classified as such, or if one of the randomly selected image patches was classified as an occlusion point. The error signal was also given if one of the synthetic edges was classified as a point of occlusion. This was to discourage subjects from making occlusion classifications just on the basis of an edge running through the center of the patch. No feedback was given on trials with synthetic junctions, since these were intended as a barometer of a subject's strategy.

Subjects began their participation in the experiment by completing roughly an hour's worth of practice trials. Once fully comfortable with the judgments and response requirements, each subject completed approximately 700 experimental trials. The subjects who participated in the experiments were different from the subjects who labeled the images, and had never seen the images from which the patches were extracted. Data were collected from 5 subjects.

3.2 *Results*

The data from all five subjects showed the same trend, and were combined for analysis purposes. Figure 3a shows classification errors for the real-image patches as a function of the aperture size. At the smallest aperture size (13 pixels in diameter), subjects misclassify 27% of the image patches (chance performance is 50%). Errors drop gradually as the aperture size is increased, but do not reach a minimum until an aperture size of about 100 pixels. Figure 3b shows histograms of subjects' occlusion judgments for the real-image patches, divided into patches that actually were points of occlusion and those which were randomly selected. Most of the effect of aperture size is on misses: points of occlusion that are not classified as such; shown in figure 3b. At the smallest aperture size, just over half of the patches that are in fact centered on points of occlusion are correctly classified; this proportion increases to over 90% at the largest aperture size. Note that all of these image patches were clearly points of occlusion when the entire image was viewed at once; this was the criterion by which they were labeled. The fact that subjects misclassify a small proportion of the occlusion points at the largest aperture size tested implies that a degree of ambiguity exists even for this large size.

One might have expected to see a similar effect in the randomly selected locations, ie a decrease in false positives as the aperture size is increased. Perhaps surprisingly, however, false-positive levels (the proportions of randomly selected patches classified as point of occlusion) are roughly constant across condition (figure 3c). This can be

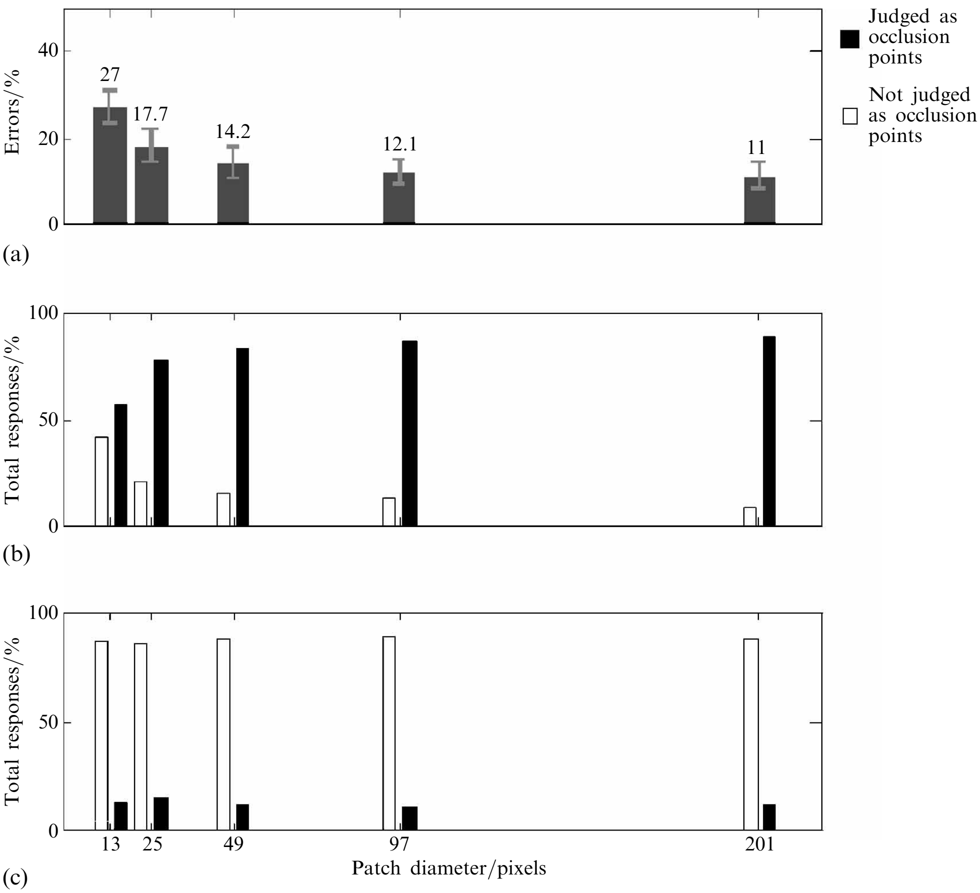


Figure 3. Results of experiment 1. Occlusion classification of real-image patches (five subjects). (a) Total errors; (b) judgments for points of occlusion; (c) judgments for random points.

attributed to at least two factors. First, one weakness of our paradigm is that most of the points that are supposed to *not* be points of occlusion are randomly selected. Occasionally, locations are randomly selected that happen to lie on points of occlusion. With purely random selection this would be a very improbable event, but, as described above, the patches were also constrained to have high variance in their pixel intensities, which predisposed the selection process towards edges and other regions of high contrast. Second, subjects were given intentionally lax guidelines with regard to the centering of the occlusion points: they were instructed to classify something as being centered on an occlusion point if there appeared to be such a point within the cross-hair boundaries, a 10 by 10 pixels region. This made it easier still for randomly selected patches to legitimately contain occlusion points. Thus many of the 'random' points that were classified as occlusion points were classified correctly, and are not true false-positive responses. False-positive rates are thus clearly overestimated; this is a weakness of the method.

The qualitative results are quite clear, however: even when forced to guess, human observers were unable to identify many points of occlusion just on the basis of local information. This is consistent with the notion that in many cases there is not much of a local cue such as a T-junction with which to detect the occlusion. Conversely, in many cases observers could reliably identify occlusion points with purely local information, providing an existence proof for local cues of some sort.

Figure 4 shows occlusion judgments for the synthetic junctions inserted as 'catch trials' in the experiments. In contrast to the real-image patches, the classification of the synthetic junctions depended very little on aperture size. Virtually all of the T-junctions were classified as occlusion points, whereas most of the L-junctions were not. Note that this does not show that observers' judgments on the real points of occlusion are mediated by junctions; they could very well be monitoring other things. It is, however, consistent with the notion that junctions are involved when they can be seen. Moreover, it suggests that if junctions had been present at most points of occlusion in real images, the observers would have made use of them to identify the occlusion points as such. Thus the fact that they fail to pick out many of the occlusion points suggests that junctions are not there.

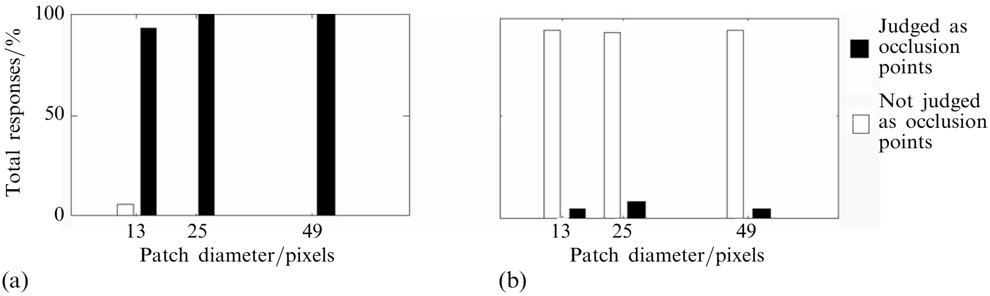


Figure 4. Results of experiment 1. Occlusion judgments for synthetic junctions. (a) T-junctions; (b) L-junctions.

Figure 5 shows histograms of the observers' confidence ratings in each experimental condition. For display purposes, the ratings were combined with the occlusion judgments, such that the six bars in each histogram denote, from left to right, the proportion of (1) very confident, (2) somewhat confident, and (3) not confident "this is not an occlusion point" judgments; and (4) not confident, (5) somewhat confident, and (6) very confident "this is an occlusion point" judgments. For small aperture sizes, subjects were frequently guessing the category of the real-image patches. As the aperture size increased, more of the occlusion points were confidently categorized as such, and, by the largest aperture size, the majority of judgments are very confident and correct. There was a similar but weaker trend evident in the responses for the randomly selected patches. In contrast, the vast majority of the confidence ratings for the synthetic junctions were high, for both small and large apertures (data not shown). Evidently the occlusion evidence in real images is far more ambiguous.

3.3 *Illustrative examples*

A few examples of the stimuli help illustrate the range of situations encountered in real images. Figure 6 shows a number of real points of occlusion that are easily identifiable as such viewed through apertures 13 pixels in diameter. All of these points generate local image structures that could be comfortably termed T-junctions.

Figure 7 shows another group of occlusion points, again viewed through apertures 13 pixels in diameter. In contrast to the previous group, these image patches are not obviously the products of occlusion points, and as a result were incorrectly classified by the observers. None of the occlusion points in this group generates prototypical T-junctions, for reasons that vary from case to case. In some instances the occluded contour is very low contrast and hard to resolve; in others one or more of the contours is blurry. Still others have clearly defined contours but which are not configured in a manner suggestive of occlusion; they are difficult to interpret until seen in a larger context.

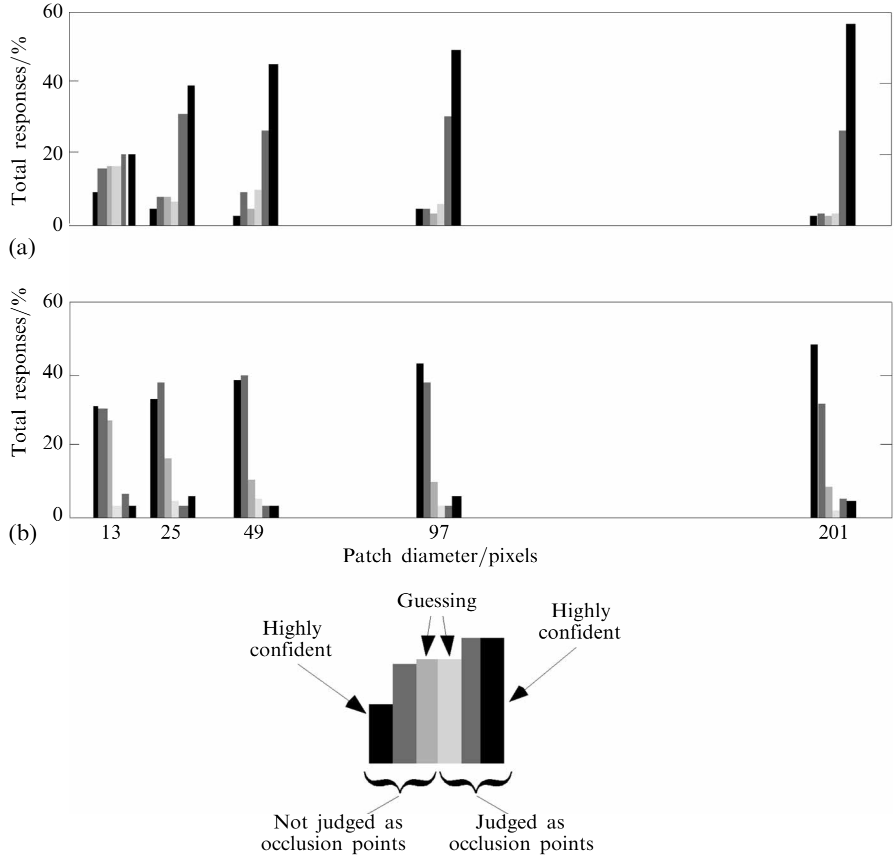


Figure 5. Confidence ratings for experiment 1. Occlusion classification of real-image patches (five subjects). (a) Judgments for points of occlusion; (b) judgments for random points.

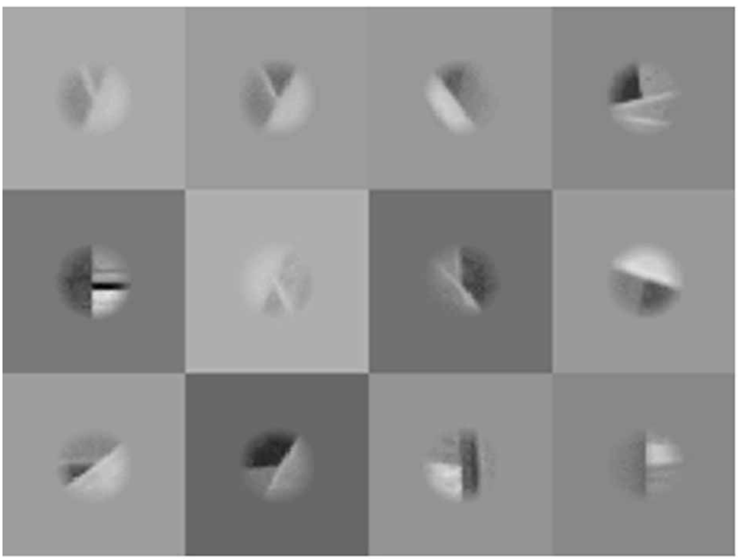


Figure 6. Examples of stimuli viewed through apertures 13 pixels in diameter. Even when only the local information visible through the apertures is available, all of these locations are good candidates for occlusion points, presumably because they generate T-junctions.

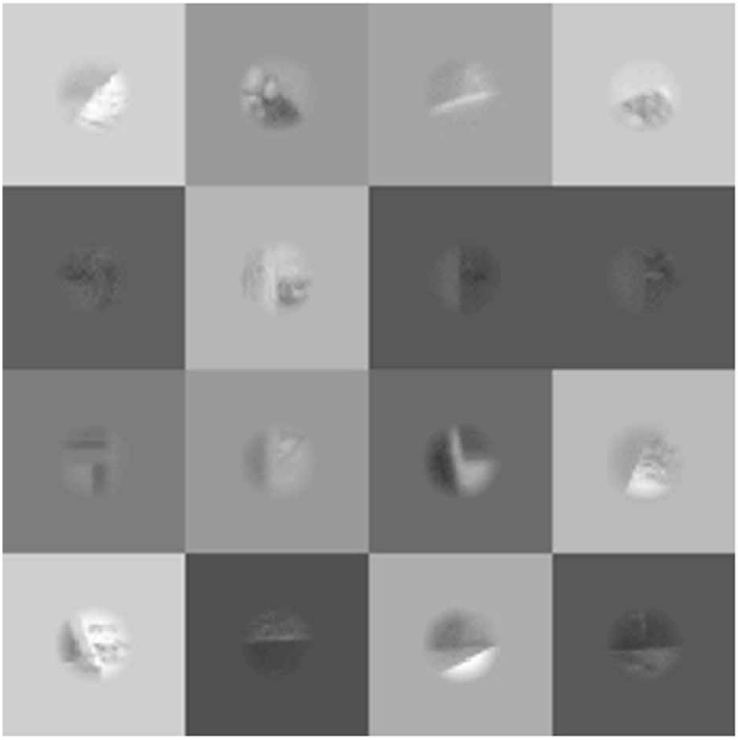


Figure 7. Examples of stimuli viewed through apertures 13 pixels in diameter. Just given the local evidence visible through these small apertures, these locations are relatively poor candidates for occlusion points.

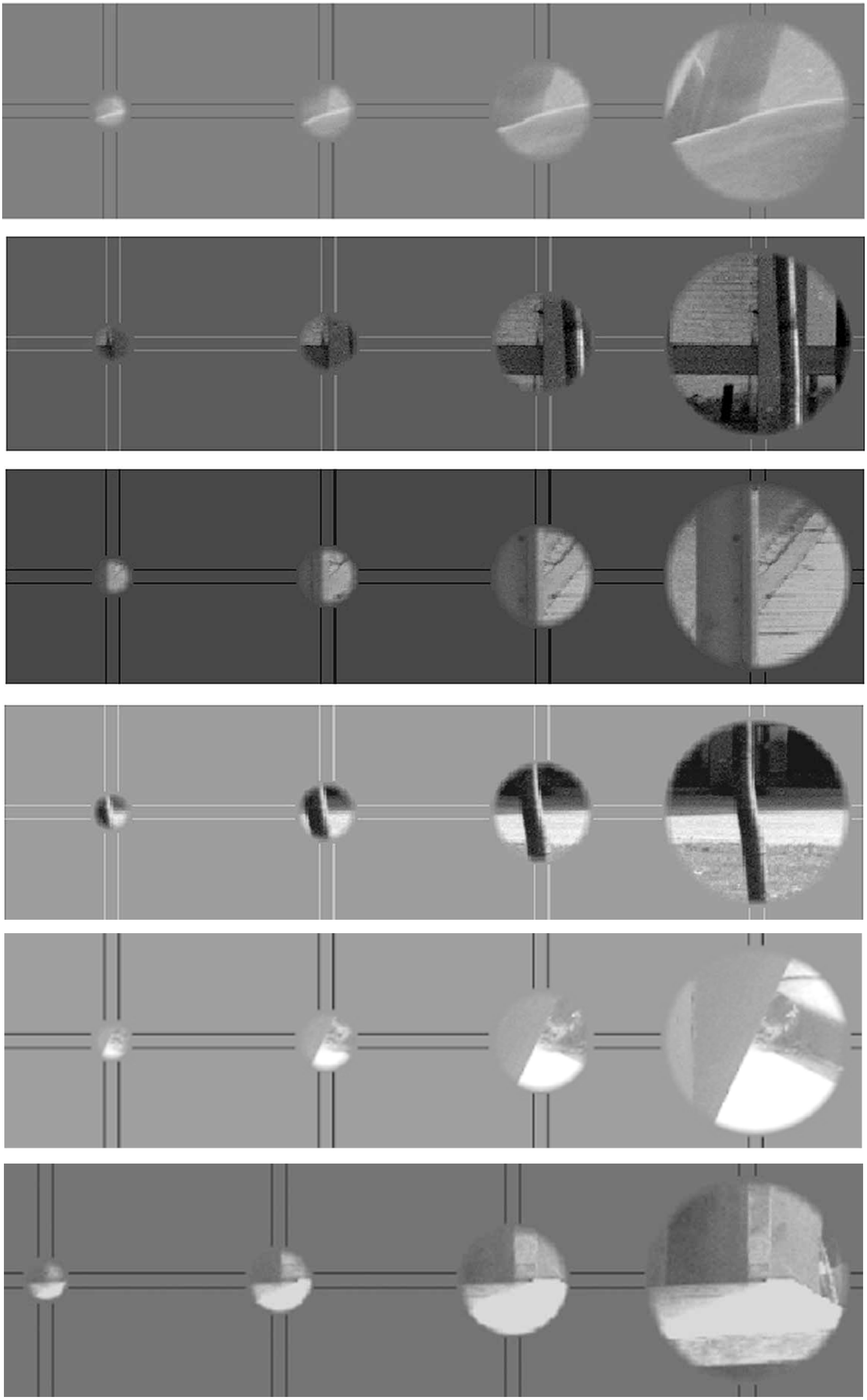
Figure 8 further illustrates this, showing some of these and other occlusion points over a range of aperture sizes (the four smallest sizes used in the experiments). Although not obviously occlusion points when viewed through the smallest apertures, as the aperture size increases it becomes easy to tell that the image patches are centered on points of occlusion. In most of these cases the local evidence seems weak; analysis over a larger region is necessary.

The fact that a large number of pixels (eg 50 by 50) is needed to recognize many points of occlusion might indicate that a rather complicated mechanism is necessary for their detection (more complicated than mechanisms that could detect synthetic junctions, which can evidently be recognized with far less information). Before considering this possibility, however, I will first address several alternative accounts of the results.

4 Experiment 2. Controls with zoomed image patches

One relatively uninteresting explanation of the basic result of experiment 1 might involve lateral masking between the aperture border and aspects of the image patch. For small apertures, the borders are close to the center of the patch, and might thus interfere to a greater extent with subjects' ability to detect the image structure relevant to their occlusion judgments. It seemed unlikely that masking could be so strong as to generate the trend present in the results, but I nonetheless sought to rule it out as a contributing factor. I therefore repeated experiment 1 with the image patches 'blown up'; each image was upsampled and blurred to twice the original size.

Figure 8 (opposite). Examples of occlusion points viewed through apertures 13, 25, 49, and 97 pixels in diameter.



4.1 *Methods*
Image patches were blown up by a factor of two with the use of the upBlur function in Eero Simoncelli's MatLab Pyramid Toolbox (http://www.cns.nyu.edu/~eero/software.html). The patch sizes were 14, 26, 50, 98, and 194 pixels in diameter, each generated from a patch of the original image half that size. Thus the smallest patch size was generated from a patch 7 pixels in diameter, a size not included in the original experiment. The other four patch sizes were generated from the four smallest sizes of the original experiment.

4.2 *Results*
As shown in figure 9, the basic result of experiment 1 was replicated with the zoomed patches; discrimination is poor for small apertures and improves markedly as the aperture size is increased and a greater number of pixels become visible. Moreover, error rates were similar to those from the original experiment for patches generated from the same number of pixels.

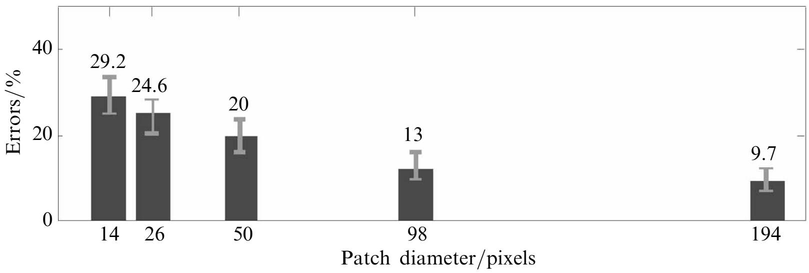


Figure 9. Error rates for different aperture sizes. Occlusion classification of zoomed real-image patches (four subjects). Error bars denote 95% confidence intervals. Compare to figure 3a.

I ran a second experiment on two subjects in which the small patches were blown up more than the large ones, such that the display size was the same for all the apertures. Again, discrimination was much worse for the displays generated from a small number of pixels, even though the patch borders were now at the same eccentricity in all conditions (data not shown). It thus appears that interference by the aperture borders is unlikely to account for the results.

The results of experiment 1 were also replicated with images reduced (blurred and sub-sampled) by a factor of two. I took the labels of the original images and weeded out points that were no longer visible in the reduced images (those that were formed by very thin contours in the original images, for instance), yielding a set of occlusion point locations for reduced versions of the original images. I ran two subjects in the aperture-size experiment with image patches generated from these locations in the reduced images, and found similar proportions of detectable occlusion points for each patch size. This controls for the possibility that the CCD used to take the pictures might have corrupted the high frequencies through aliasing.

In sum, it seems unlikely that the effects reported here are due to problems with the stimuli. What else might explain the results?

5 Experiment 3. Scale
Another obvious explanation of the results of experiment 1 involves the spatial scale of any putative signature of occlusion. Intensity changes in images due to object boundaries and other structures in the world occur over both large and small scales (Koenderink 1984), and one would expect junctions to exist at different scales as well. The detection of junctions at large scales would require analysis of a large image region even though they might require no more parameters for their description than do fine-scale, locally

defined junctions. As illustrated in figure 10, a large-scale junction lacking crisp edges is hard to detect when viewed through a small aperture—a large area is needed to detect the intensity gradients defining the junction. Note, though, that this example junction is a simple structure and could presumably be recognized with a fairly small number of image samples; they just need to be distributed over a large region of space. Similar examples can be found in real images; figure 11 shows one. Such image regions appear to contain relatively simple image features even though they require a large area to be detected. The dependence of occlusion detection on aperture size might thus be the result of simple image structures distributed across a variety of scales, and could still be consistent with the notion of junctions as simple and locally defined cues.

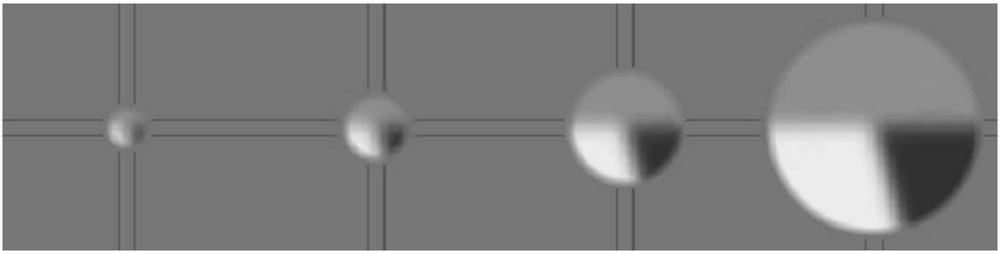


Figure 10. A blurred synthetic junction viewed through apertures 13, 25, 49, and 97 pixels in diameter. The junction is not visible as such when viewed through the smallest aperture.

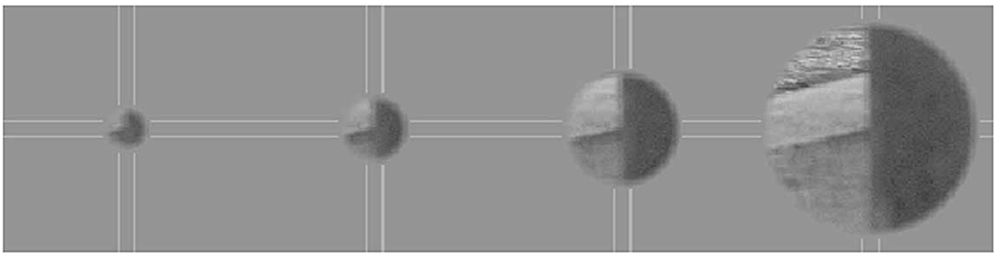


Figure 11. A real occlusion point which appears to have simple junction structure at a large scale owing to blurred edges.

How can we test the extent to which large-scale junctions could account for the results? Intuitively, we want to isolate the large-scale structure of image regions whose occlusion points are not visible through small apertures, and see if the occlusion points remain detectable when only the large-scale structure could be underlying performance. A standard technique for isolating structure at large scales is to blur and sub-sample an image (eg Burt and Adelson 1983). Blurring eliminates the fine-scale structure, and sub-sampling converts the remaining low frequencies into high frequencies defined over a small number of pixels. The point of this is twofold. First, simply blurring the image patches would be confounded by the fact that the resulting patches look blurry, which could conceivably affect subjects' judgments. Sub-sampling the blurred patches makes them look as sharp as the originals. Second, the sub-sampling step takes the large-scale structure that was formerly distributed over many pixels and compactly represents it over a small set of pixels. If large-scale junctions are mediating occlusion discrimination for the occlusion points that are visible only through large apertures, these points should be rendered visible through small apertures once the large-scale structure is isolated through blurring and sub-sampling. Figure 12 shows the effect of blurring and sub-sampling the large-scale synthetic junction of figure 10. Each row was generated by viewing a blurred and sub-sampled version of the image in the previous

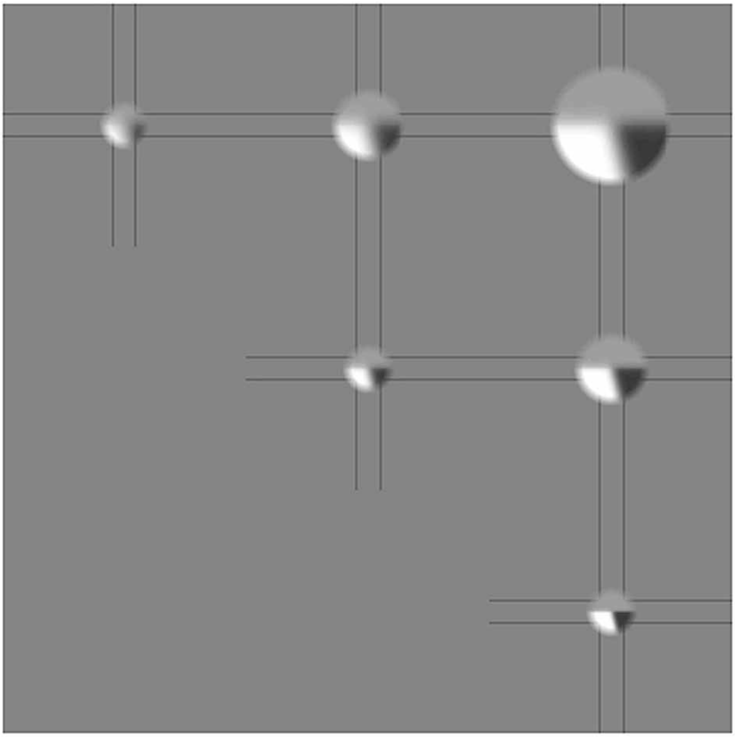


Figure 12. Reduced versions of a blurred synthetic junction viewed through apertures of different sizes. The image for each row is a reduced version of the image in the previous row. Note that the junction is visible in the reduced images viewed through the small apertures, because the large-scale structure defining the junction has been isolated.

row through apertures of different sizes. The junction is not visible in the original image when viewed through the small aperture, but once reduced it becomes crisp and easy to see. We thus have a test of the extent to which simple structures at multiple scales underlie the result reported here.

5.1 *Methods*

Each image location was shown in six conditions: through (i) small, (ii) medium, and (iii) large apertures at full resolution, (iv) blurred, sub-sampled, and viewed through a small aperture (ie the medium apertures reduced once), and (v) blurred, sub-sampled, and viewed through a medium aperture (ie the large apertures reduced once), and (vi) blurred and sub-sampled twice viewed through small apertures (ie the large apertures reduced twice, down to the size of the small apertures). The extent to which junctions at large scales might be mediating occlusion discrimination can be measured by examining whether occlusion points that are visible through a large aperture but not through a small one become visible through a small aperture once they have been reduced.

Unlike the previous experiments, in which each image location was included in just a single condition, in this experiment each location was used in all six conditions, because of the design of the experiment. This led to a concern that a subject's judgment for a particular location in one condition might influence his/her judgment for that location in another condition seen later in the experiment. For an occlusion point clearly visible when viewed through a large aperture, for instance, it seemed conceivable that when presented with the same image masked by a smaller aperture subjects might realize that they had seen it before, and base their judgment on having already seen it as an occlusion point regardless of the local evidence for occlusion. To minimize the chances of that sort of thing occurring, I randomly left–right reversed and rotated the image patches 90° clockwise or counter-clockwise in the different conditions. This made it less obvious that image patches were being shown multiple times. Subjects were not informed that this was the case, and none ever reported seeing the same thing more than once.

Subjects' instructions were the same as in the previous two experiments. All subjects were naïve as to the specific purpose of this experiment. As in the previous experiments, feedback was given to enable subjects to perform as well as possible.

Two versions of the experiment were run, each at a different set of scales: one (experiment 3a) in which the aperture sizes were 7, 13, and 25 pixels in diameter and the images were expanded by a factor of 2 as in experiment 2, to minimize any possible masking effects from the smallest apertures; and one (experiment 3b) in which the aperture sizes were 13, 25, and 49 pixels in diameter and the images were shown at full resolution. Images were reduced by a factor of 2 or 4 in each dimension depending on the condition, and the aperture sizes were chosen so that image patches visible through the larger apertures would reduce down to the size of the smaller apertures. Images were reduced by blurring with a 5-tap binomial filter and then sub-sampling by a factor of two. Half the stimuli were generated from the set of labeled occlusion point locations, and half were generated from random locations, as described in experiment 1.

5.2 *Results*

To test for the presence of large-scale junctions, I examined all the points of occlusion which were visible through large apertures but not through small apertures at normal resolution. Each of the two experiments had three aperture sizes, so there were three such sets of points to analyze for each experiment. If the failure of subjects to detect these occlusion points for the small apertures is due to the presence of a large-scale but fundamentally simple signature such as a junction, these points should be visible through the small apertures after blurring and sub-sampling.

Figure 13 shows the proportion of such occlusion points that remained visible after blurring and sub-sampling, for each of the three analysis conditions in each of the two experiments. The results of the two experiments are fairly similar: between 60% and 80% of the occlusion points that become visible only when viewed through large apertures remain visible after blurring and sub-sampling, even though they are represented with far fewer pixels (1/4 or 1/16 as many, depending on the analysis condition). This suggests that in many cases in which occlusion points require a large region of analysis to be detected, there is nonetheless a fairly simple signature of their presence. This could conceivably provide a simple explanation why many occlusion points cannot be detected when viewed through small apertures.

5.3 *Scale examples*

Figure 14 shows some examples of occlusion points for which there is clearly a simple but coarse-scale junction-like structure. As before, each row consists of the same image viewed through apertures of different sizes. The images for each successive row are blurred and sub-sampled versions of the images from the previous row. The aperture sizes are arranged such that the image patches in each column cover the same region of the original image at different scales. Occlusion is hard to see through the small apertures at high resolution, but is much clearer after blurring and sub-sampling when the large-scale structure is isolated, mainly because T-junctions become evident.

However, upon inspection, it becomes clear that in some cases what appears to be evidence for occlusion after blurring and sub-sampling is in some sense artifactual.

Figure 15 shows one such example. The contour being occluded at the center of the patch is the edge of a tree trunk which happens to be fairly low-contrast and visible in part owing to a texture boundary. It is not visible until a large region is available for analysis. When the image is reduced, the vertical edge of the labeled occlusion

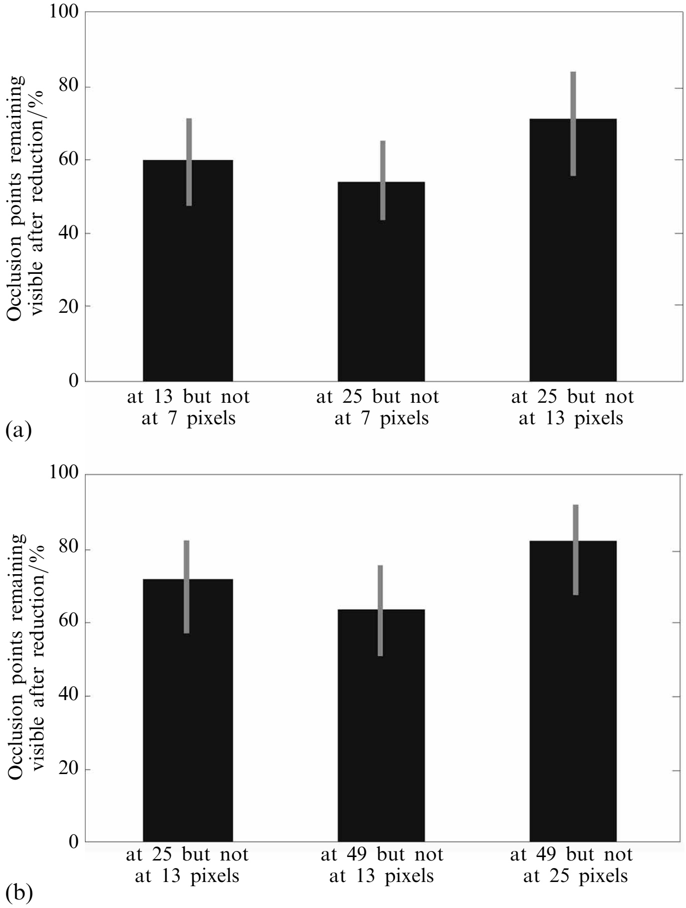


Figure 13. Scale analysis of points of occlusion. Results of experiment 3a (four subjects) and experiment 3b (five subjects). Error bars denote 95% confidence intervals.

point becomes invisible, but the shading on the tree trunk produces a high-contrast edge that is shifted towards the center of the patch by the reduction process. This image was classified as an occlusion point, but the T-junction responsible for the judgment is not the product of the occlusion point under consideration, and is not even produced by something that was labeled as an occlusion point in the original image. The fact that subjects make the correct response for a stimulus such as this is thus an artifact of the reduction process. The response is artifactual in the sense that the apparent large-scale signature of occlusion that seems to be responsible for the judgment is (a) not generated by the occlusion point in the world that was originally labeled, either being due to another nearby occlusion point brought closer to the center of the image by the reduction process, or an accidental large-scale image structure, and (b) presumably not responsible for the (correct) percept of occlusion in the full-scale patches.

5.4 *Scale artifacts*

We are thus likely to be overestimating the extent to which occlusion points are often indicated by large-scale occlusion cues; some of what appear to be occlusion cues are artifacts. Fortunately, one can estimate the frequency with which these artifactual occlusion signatures are produced by examining the subjects' occlusion judgments on the randomly selected patches of experiment 3. I looked at those randomly selected patches

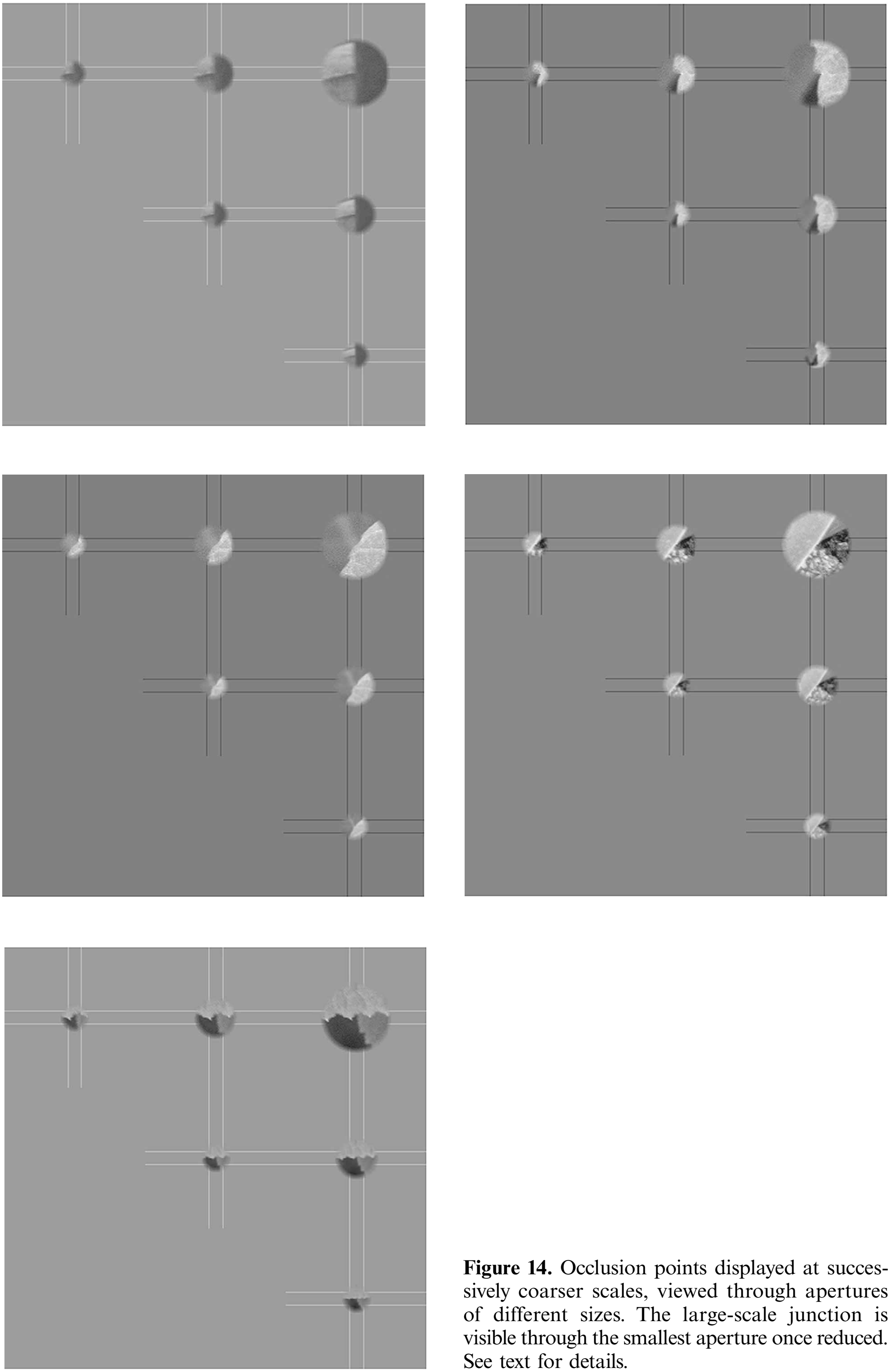


Figure 14. Occlusion points displayed at successively coarser scales, viewed through apertures of different sizes. The large-scale junction is visible through the smallest aperture once reduced. See text for details.

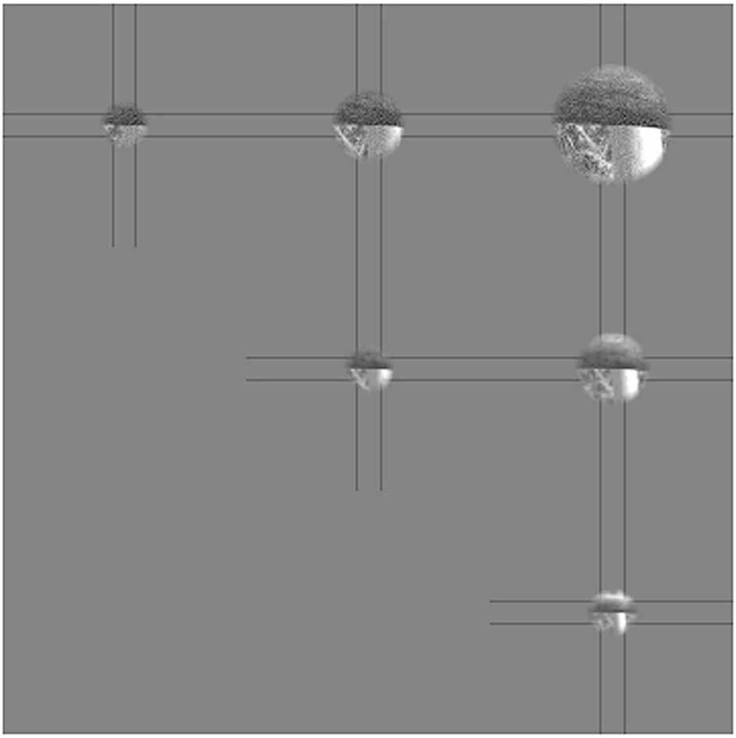


Figure 15. An occlusion point that is artifactually visible given local information at a large scale. See text for details.

which were correctly rejected at two aperture sizes, and computed the proportion which generated false-positive responses (ie that were incorrectly labeled as occlusion points) when they were reduced from the size of the large aperture to that of the small one. Since these points were originally selected at random and did not appear to be occlusion points at either aperture size at full resolution, one can be reasonably sure that the appearance of occlusion in the reduced image is an artifact of the reduction. Moreover, the appearance of occlusion in the reduced image clearly is not the determinant of occlusion judgments for the full resolution images, because I have selected patches that at full resolution are judged not to be occlusion points. One might thus say that the large-scale structure at such points is both physically artifactual (in that they are not really due to an occlusion point at the center of the patch) and psychologically artifactual (in that they are not responsible for occlusion percepts at full resolution). Figure 16 shows several examples of image patches from a randomly selected set that artifactually look like occlusion points at large scales.

Figure 17 shows the proportion of these artifactual occlusion judgments for each of the three analysis conditions in the two parts of experiment 3. In experiment 3a the aperture sizes were 7, 13, and 25 pixels; and in experiment 3b they were 13, 25, and 49 pixels in diameter. The plots in figure 17 show the proportion of false-positive responses for correctly identified random patches when they were reduced and viewed through small apertures. On average, blurring and sub-sampling introduces artifactual occlusion signatures in around 20% of the randomly selected patches. Assuming that the frequency of these artifacts is the same for image regions centered on actual points of occlusion and those randomly selected, we can estimate that on the order of 20% of the occlusion points visible after blurring and sub-sampling are artifactual: the large-scale structure, although apparently consistent with occlusion, is neither

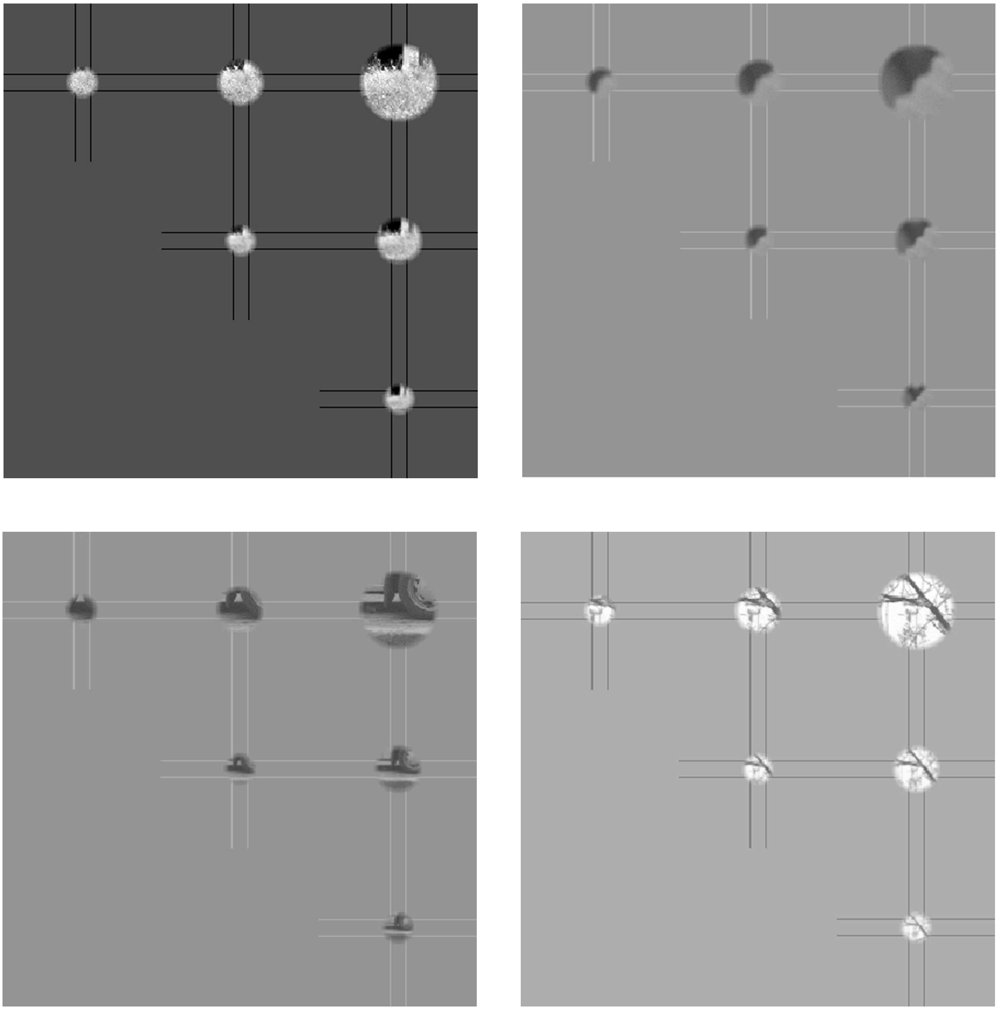


Figure 16. Examples of randomly selected patches which at large scales generate what looks like a junction close to its center. In the bottom two figures there appears to be a T-junction at the largest scale (bottom right). Note that there is a proper occlusion point where the shadow passes behind the tire, but that this is not what produces the T-junction in the bottom right display.

physically due to occlusion nor psychologically responsible for occlusion judgments at full resolution. The proportion of occlusion points which are not visible locally but which produce a large-scale signature that could be responsible for our subjects' occlusion judgments is thus probably closer to 50% (compared to our original figure of 60% to 80%). This means that there is a substantial fraction of occlusion points whose visibility cannot be attributed to a simple (low-dimensional) signature at any given scale. This can be estimated to be about 25% of all occlusion points. About half of all our occlusion points were visible locally (figure 3), and then half of the remainder produce a large-scale signature, leaving about 25%. The effects of aperture size found in the first two experiments thus cannot merely be due to junctions at multiple scales.

5.5 *Nonlocally defined occlusion points*

Figure 18 shows several examples which fall in this category: these occlusion points do not become visible until viewed through large apertures, and blurring and sub-sampling them does not preserve this visibility. This is for a variety of reasons. In some cases one or

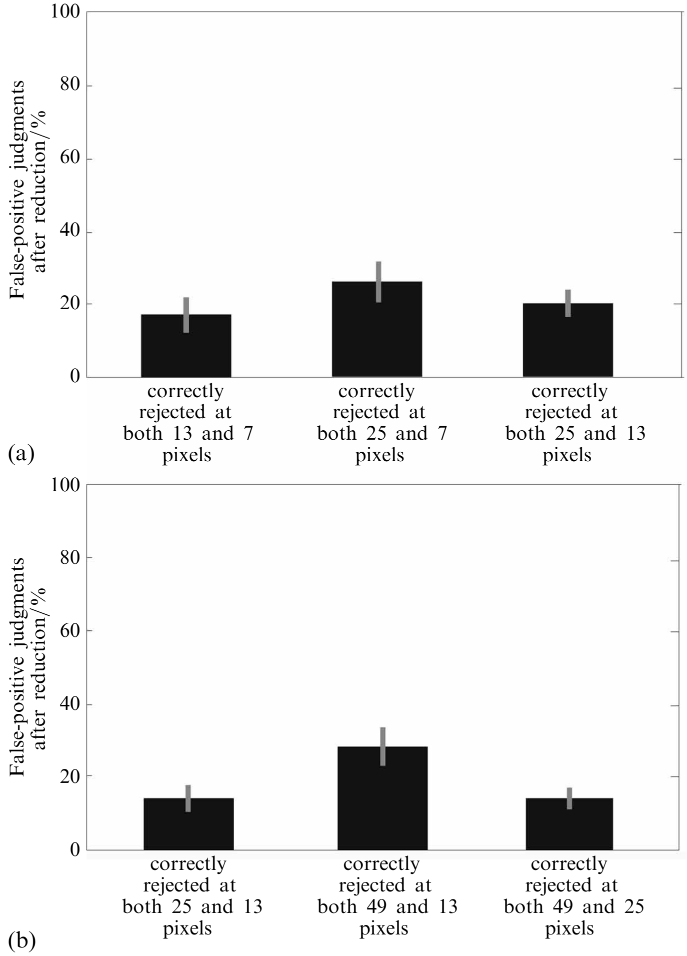
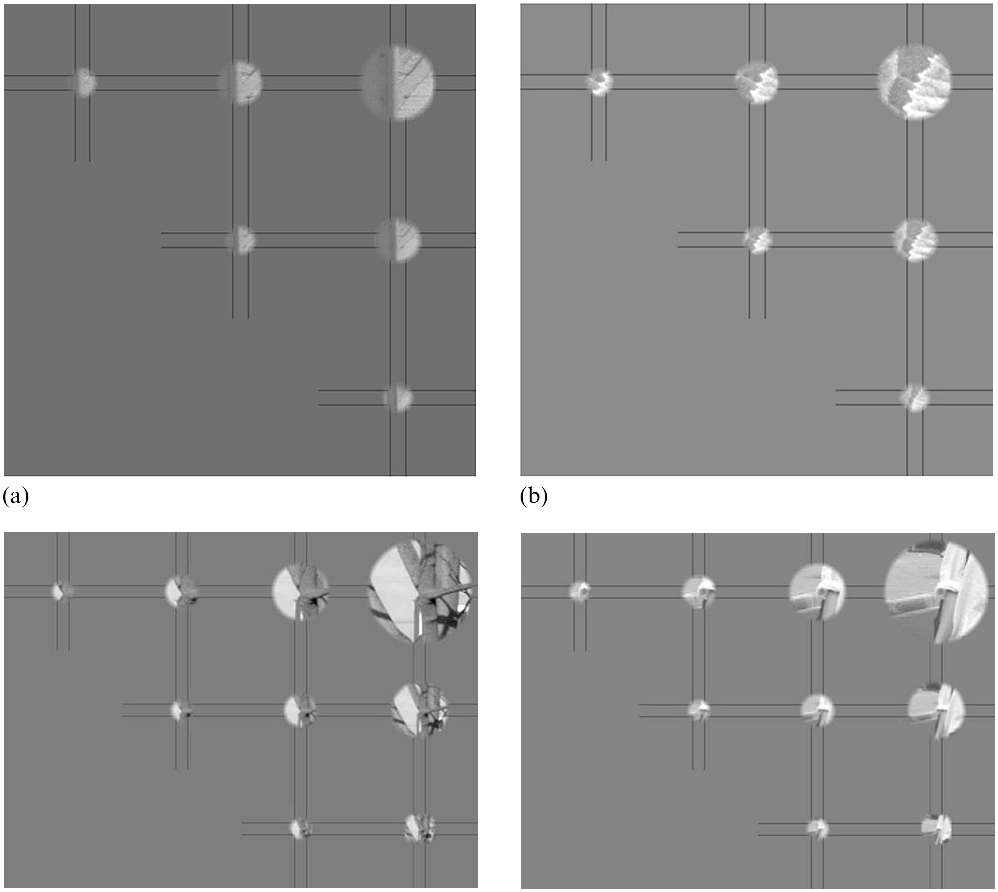


Figure 17. Artifactual large-scale occlusion signatures. Results of experiment 3a (four subjects) and experiment 3b (five subjects).

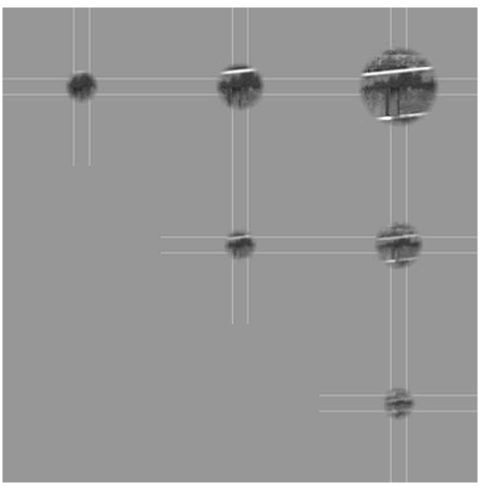
more of the contours is not visible at the larger scales (because lines, unlike step edges, are not scale-invariant); in other cases spurious large-scale structures are introduced which seem inconsistent with occlusion. The detection of these occlusion points evidently requires a mechanism that can make use of all of the pixels in a large region.

5.6 *Scale selection*

Although many occlusion points seem to necessitate analysis of fine-scale structure over a large region, a substantial portion produce fairly simple signatures that are visible at fine scale, and the remainder seem to produce fairly simple signatures only at a coarse scale. If there was a simple way to select the scale of analysis, a large proportion of occlusion points could thus be detected by means of fairly simple mechanisms. The selection of the correct scale is critical, though. The existence of artifactual junctions at large scales discussed earlier implies that one cannot simply take a junction at a large scale as evidence for occlusion; large-scale structures are frequently spurious. Moreover, spurious large-scale structure generally does not 'fool' the human visual system, implying that scale selection mechanisms are at work in human vision. What might they involve?



(a) (b) (c) (d)



(e)

Figure 18. Examples of junctions whose visibility requires a large number of pixels regardless of the scale of analysis. (a) The stem of the junction is a thin contour that is undefined in the immediate vicinity of the junction center but is also lost at large scales. (b) Another example with three aperture sizes. (c) Four aperture sizes are used; the branches are difficult to individuate for all but the largest apertures. (d) Four apertures sizes; the structure of the woman's shirt sleeve near the occlusion point is confusing unless one has access to a large number of pixels at high resolution. (e) Three aperture sizes are used; occlusion point can only be seen over a large region at full resolution.

Many theories of scale selection rely on the magnitude of image gradients at different scales to choose the scale of analysis. Elder and Zucker (1998), for instance, have proposed an elegant framework that chooses the minimum reliable scale for edge detection by comparing gradient magnitudes at different scales with the values that would be produced by sensor noise alone (see also Morrone et al 1995; Lindeberg 1998). Edge detection at a particular location is conducted at the minimum scale where the gradient exceeds these noise levels. This prevents spurious responses to noise (to which fine-scale operators are more vulnerable) while minimizing the interference of neighboring image structure (which is a problem at large scales, as we have seen). One could envision a similar scheme for junction detection at different scales; detection would occur at the minimum scale at which the gradient magnitudes exceeded the levels expected just from noise.

For some junctions, such a strategy seems reasonable. Consider, for instance, the blurry synthetic junction of figure 12 that was used as a motivating example: there is very little high-frequency structure. When viewed through a small aperture at high resolution, little is visible other than a few blurry image gradients, if that. Derivative magnitudes computed for this junction will be small for fine scales, and a large scale will correctly be selected for analysis.

In many other cases, though, there is plenty of high-frequency structure; it is just not the sort that is obviously associated with junctions or occlusion. A good example of this is shown in figure 19. The image is of a plank of a fence which occludes a round support pole. The plank casts a shadow on the pole that has an unusual shape due to the pole's curvature. When viewed through a small aperture, the shadow is hard to interpret as such and prevents most observers from correctly categorizing the image as an occlusion point. When observers have access to a larger region of the image, the shape of the pole becomes evident and the shadow starts to look like a shadow, and it becomes obvious that there is a point of occlusion at the center of the aperture.

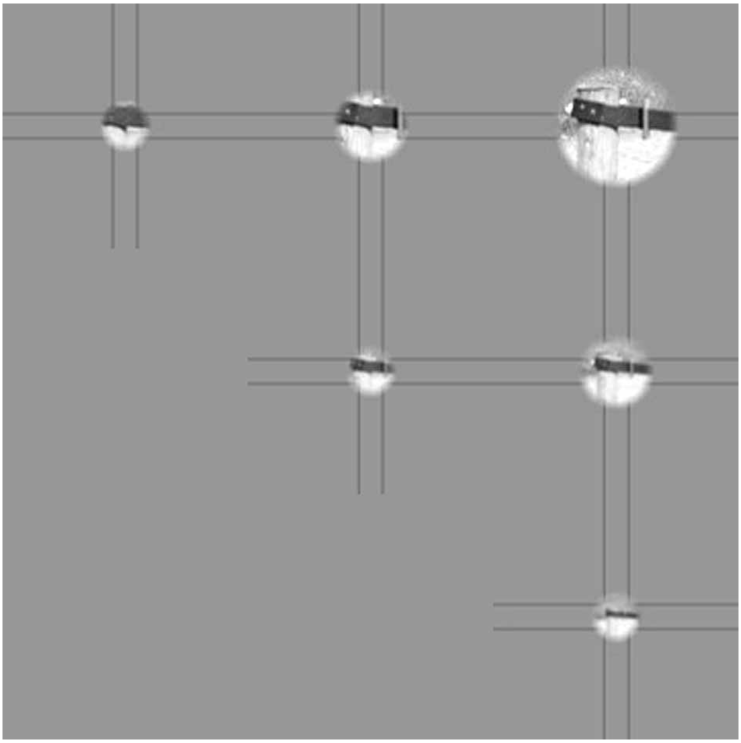


Figure 19. Example of an occlusion point where a cast shadow drastically alters the local intensity signature. Although there is a junction at coarse scales, how should one choose to ignore the fine-scale shadow unless it has already been recognized as such?

Notably, most observers continue to classify this image as an occlusion point after it is reduced. This is because the reduction erases the small shadow, leaving what appears to be a fairly crisp T-junction. However, the high-frequency structure in the image is salient and inconsistent with the large-scale T-junction. The scale selection methods outlined above would clearly fail to isolate the large scale structure in images such as this—there is abundant high contrast, high-frequency structure, and no obvious bottom–up method with which to select the large-scale for junction analysis. We cannot directly estimate the frequency of this situation from our data, but inspection of our stimuli suggests that it is quite common. It is worth noting that all examples in figure 14, although containing a low-frequency junction, also contain substantial high-frequency information. In particular, it is common for one of the edges forming the occlusion point to be blurred while the other is sharp. This is because edges at occlusion points are often at substantially different depths, and one will be out of the focal plane of the imaging device.

Thus, even though points of occlusion in many cases generate large-scale junctions, the utility of these junctions as simple, easy-to-detect occlusion cues depends critically on a method for selecting the proper scale of analysis. If scale is not properly selected, false-positive responses to spurious large-scale structure will result. In many cases, it is not obvious that there is a bottom–up method for scale selection (eg in the shadow example of figure 19), in which case the presence of large-scale junctions does not simplify the analysis of the many-pixel arrays needed to detect many of the points of occlusion. In the shadow example, the shadow may need to be recognized as a shadow before occlusion can be identified and the junction labeled as such.

The shadow example of figure 19 also highlights a shortcoming of the paradigm used here, namely that it is difficult to say anything specific about why so many pixels are needed to detect certain points of occlusion. The fact that observers require a large number of pixels is consistent with both bottom–up (low-contrast contours must be integrated across large numbers of pixels to be detectable; textures require large numbers of pixels to be segregated; etc) and top–down (shadows must be correctly interpreted) accounts of how occlusion points are being detected. For individual examples, we can speculate about what is going on, but it is not obvious how to assess this issue experimentally across an ensemble of occlusion points.

6 General discussion

I have introduced a psychophysical method for studying local occlusion cues in real images. The results reported here offer some insights into junction detection that would be difficult to obtain with traditional psychophysics experiments with synthetic stimuli that lack the ambiguities of real images. Although it is easy to go through an image and point out junction after junction, when observers are restricted to purely local information, many so-called junctions are no longer visible, suggesting that junctions in such situations are not detected with local operators.

Many of the issues discussed here in the context of junction detection also apply to edge detection. Given that junctions are just the intersections of edges, one might wonder why I did not simply do these experiments on edges instead of junctions. The main difficulty that discouraged me from trying was the fact that edges are so common in images. Particularly because edges are two-dimensional structures, the labeling process seemed as though it would be much more tedious. Points of occlusion are one-dimensional and can therefore be labeled with a single click of a mouse. Moreover, their relative infrequency in images makes it reasonable to label most of them in a moderate number of images, which would be harder to do with edges. There are also additional computational issues raised by the multiple orientations present in junctions (Freeman 1992), so junctions are worth studying in their own right.

However, it would nonetheless be possible to label a small set of points on each edge in an image, and to employ the present methodology to ask similar questions of edge detection. Since I conducted these experiments several investigators (Geisler et al 2001; Elder and Goldberg 2002) have successfully employed human subjects to label edges in images, demonstrating the feasibility of the labeling process.

However, until recently, psychophysicists have avoided using real images as stimuli, in part because the stimuli are poorly controlled and the judgments ill-defined in comparison to traditional psychophysics with synthetic stimuli. I think that in many instances these difficulties are not insuperable, in which case approaches such as mine may be a useful addition to standard psychophysical methodology. Nonetheless, there are a number of potential concerns with my paradigm that I wish to address before discussing my conclusions.

First, the success of my method depends critically on the ability of observers to perform the task effectively. One might argue that people might not know what characterizes occlusion points in real images, since they are not used to seeing them in isolation. Perhaps as a result of this they might be impaired at the task irrespective of the available local information. This is a possibility one can never fully rule out, but I did take a number of steps to help ensure that subjects performed as well as was possible. First, stimulus duration was unlimited, and subjects were encouraged to take as long as they needed on each trial. Second, the responses to the synthetic junctions inserted as catch trials helped to insure that subjects were doing something reasonable. Inspection of their responses to various stimuli further confirms this. Third, the subjects received feedback on their responses, were instructed to use the feedback to maximize their performance, and were initially given extensive practice prior to taking part in the actual experiments. In addition, the task the subjects were required to perform was described by many of them as quite natural and easy to learn. Given all this, it seems likely that the subjects were fairly effective at picking out local occlusion cues when they were present.

A second area of concern is the choice of randomly selected image patches as negative examples of points of occlusion. The negative examples are critical, since otherwise subjects could simply label everything a point of occlusion without even looking at the stimuli. Given that I am interested in how occlusion points might be detected in real images, patches of real images not centered on occlusion points seemed the logical thing to use as negative examples. As I have discussed, many randomly selected patches of real images are fairly uniform at the center, and are thus trivially easy to distinguish from a likely point of occlusion. To force subjects to make a more specific discrimination, I screened out those randomly selected patches whose intensity variance was below a threshold and added synthetic edges as additional negative examples, but these were essentially arbitrary steps. The negative examples could conceivably be chosen more carefully or generated artificially so as to force subjects to make an even more specific judgment, and only label points of occlusion as such if they are signified by very well-defined junctions. However, it is not obvious that such a stimulus set would be a more principled choice than the one I employed.

The choice of images used was also fairly arbitrary. The van Hateren and van der Schaaf database was chosen mainly because it was one of the few publicly available sources of natural images. The images that were used in the study were selected by me because they contained many unambiguous occlusion events. My set of occlusion points came from 20 different images. Some images were of man-made scenes with buildings and street lights; others were of forests and other natural environments. The issues documented experimentally, involving aperture size and junction visibility, can be found to varying extents in all of these images, as can be seen from inspection of the example stimuli. I have also noted informally that many other images from outside the van Hateren and van der Schaaf database show similar effects when viewed through apertures. I thus think it unlikely that my results are specific to the images and particular methodology used here.

My results suggest that, although junctions as traditionally conceived do indeed occur at a significant number of occlusion points in real images, many other occlusion points lack a simple, local signature. I found that about half of my sample of occlusion points could not be detected when viewed through small apertures. Of those that are not visible through small apertures, up to half may generate a large-scale junction that could be detected by a simple mechanism if there were some way to select that scale of analysis. I thus estimate that in the neighborhood of 25% of all occlusion points require quite a 'large' number of pixels in order to be detected, and this proportion could be substantially higher if bottom – up scale-selection mechanisms are not available. Such occlusion points do not become visible until the image region containing them is sufficiently large that many other things are visible besides just the junction: other edges, other junctions, patches of texture, etc. Any mechanisms for detecting such junctions are performing a nontrivial analysis of a high-dimensional input space. Indeed, it is questionable whether it makes sense to use the term 'junction' to refer to such complicated image patches. There may be a sense in which junctions are present in the mental representation of such image regions, in that when we look at a large image region surrounding one of the occlusion points, we think we see a local feature even though it is not present in the local image data. But to the extent that the term is meant to denote an image feature, its use may be inappropriate. The recovery of occlusion geometry may thus involve something considerably more complicated than simply detecting a local feature, even in cases in which such a local feature appears subjectively to be present.

Note that locality is defined here in terms of pixels rather than degrees of visual angle. This is in part because the number of pixels determines the amount of information in an image region, and in part because pixels are what determines performance in the task given to the subjects. One of the control experiments (described in the section on experiment 2) tested this explicitly, repeating experiment 1 with images reduced by a factor of two. Similar error rates for each aperture size (in pixels) were found, even though the apertures in this experiment covered twice as much (in linear dimensions) of the image. This is unsurprising given the scale-invariant nature of images (Ruderman and Bialek 1994), but it confirms that pixels are an appropriate unit of measurement. For a given occlusion point, increasing the resolution within a given aperture would no doubt improve performance in some cases, less so in others. However, all the occlusion points were readily identified as such when the image was viewed in full at the prescribed resolution, so the fact that they are frequently not visible when viewed through apertures indicates that information beyond the confines of the aperture is needed to detect the junction.

Local detectors, such as might be implemented with receptive fields in early visual cortex, would therefore seem ill-suited for the analysis of the large and complicated image regions over which the signatures of many occlusion points are defined. This may provide some a posteriori justification for the failure to find neurons tuned to junctions in mammalian visual cortex. Despite numerous anecdotal reports of attempts to find such neurons, there have been no published reports of neurons selective for junctions in the nearly 40 years since the first Hubel and Wiesel papers. The results reported here also serve as a cautionary note to those, including me (McDermott et al 2001; McDermott and Adelson 2004a, 2004b), who use synthetic stimuli to study junctions, in showing that synthetic stimuli typically used in junction-related psychophysics are somewhat unrepresentative of stimuli encountered in real life.

The results reported here nonetheless show that a significant fraction (half of the sample used) of junctions is locally detectable. This means that local mechanisms, operating on image regions on the order of 10–15 pixels in diameter, ought to be able to detect this portion of the things that people label as junctions when viewing the whole image. The method used here thus serves as a more appropriate criterion with which to evaluate applications in machine vision. When the results of a local operator are evaluated by eye in the context of an entire image, the optimality of the operator is confounded with the locality of the task. The performance of a human observer given restricted information can serve as a more useful benchmark for machine-vision applications, in that the competence of a human observer at a particular task under particular conditions is an existence proof that it can be done. Note also that this is not something that can be done effectively by informally inspecting an image. Many of the occlusion points in the present sample which are not locally defined do not appear so in the context of the entire image; in many cases the human visual system seems to fill in the missing information. Presenting image patches through an aperture circumvents this problem. Recently Malik and colleagues have adopted this approach, comparing the performance of human observers viewing edges through apertures with edge detection algorithms given the same input (Martin et al 2003).

Although some subsets of junctions do appear to be locally detectable, a significant fraction are not. To the extent that vision involves the detection and propagation of local cues (eg Saund 1999; Freeman et al 2000), the results reported here suggest that junctions as traditionally conceived are not always the local cues that are used. In many cases, at least, junctions are not locally defined, and so there is no junction-like cue to detect and propagate. The experiments described here confirm that local image data are both highly ambiguous and highly varied, so much so that they may be difficult to characterize with a simple catalog of junctions, edges, and other categorical features. The appealing, rule-based conceptions of vision suggested by traditional notions of junctions may bear little resemblance to what happens when we view real images. If junctions are in some sense detected, nonlocal and more complicated processes appear to be implicated, perhaps involving top–down feedback (eg Lee and Mumford 2003). Alternatively, it is possible that junctions are not detected at all, but merely become apparent in retrospect after a scene has been interpreted.

Acknowledgments. I thank Geoff Hinton for hosting and advising me during my time in England, Bernard Buxton and Mark Georgeson for their helpful suggestions on an earlier version of the manuscript, and Ted Adelson, Peter Dayan, Hany Farid, Bill Freeman, Geoff Hinton, Denis Pelli, and Eero Simoncelli for helpful conversations and suggestions. This research was conducted at the Gatsby Computational Neuroscience Unit and was supported by a Marshall Scholarship.

References

Adelson E H, 1993 "Perceptual organization and the judgment of brightness" *Science* **262** 2042–2044

Adelson E H, 2000 "Lightness perception and lightness illusions", in *The New Cognitive Neurosciences* 2nd edition, Ed. M Gazzaniga (Cambridge, MA: MIT Press) pp 339–351

Adelson E H, Anandan P, 1990 "Ordinal characteristics of transparency", in *AAAI Workshop on Qualitative Vision* (Boston, MA: AAAI) pp 77–81

Alvarez L, Morales F, 1997 "Affine morphological multiscale analysis of corners and multiple junctions" *International Journal of Computer Vision* **25** 95–107

Anderson B L, 1997 "A theory of illusory lightness and transparency in monocular and binocular images: the role of contour junctions" *Perception* **26** 419–454

Beymer D, 1991 "Finding junctions using the image gradient", in *Proceedings of the IEEE: Computer Vision and Pattern Recognition* (New York: Institute of Electrical and Electronic Engineers) pp 720–721

Burt P J, Adelson E H, 1983 "The Laplacian pyramid as a compact image code" *IEEE Transactions on Communication* **31** 532–540

Canny J, 1986 "A computational approach to edge detection" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **8** 679 – 698

Elder J, Goldberg R M, 2002 "Ecological statistics of Gestalt laws for the perceptual organization of contours" *Journal of Vision* **2** 324 – 353

Elder J H, Zucker S W, 1998 "Local scale control for edge detection and blur estimation" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** 699 – 716

Freeman W T, 1992 *Steerable Filters and Local Analysis of Image Structure* PhD thesis, Massachusetts Institute of Technology, Cambridge, MA, USA

Freeman W T, Pasztor E C, Carmichael O T, 2000 "Learning low-level vision" *International Journal of Computer Vision* **40** 25 – 47

Geisler W S, Perry J S, Super B J, Gallogly D P, 2001 "Edge co-occurrence in natural images predicts contour grouping performance" *Vision Research* **41** 711 – 724

Guzman A, 1968 "Decomposition of a visual scene into three-dimensional bodies", in *AFIPS Fall Joint Conference Proceedings* volume 33 (Arlington, VA: AFIPS Press) pp 291 – 304

Hateren J H van, Schaaf A van der, 1998 "Independent component filters of natural images compared with simple cells in primary visual cortex" *Proceedings of the Royal Society of London, Series B* **265** 359 – 366

Heitger F, Rosenthaler L, Heydt R von der, Peterhans E, Kubler O, 1992 "Simulation of neural contour mechanisms: from simple to end-stopped cells" *Vision Research* **32** 963 – 981

Iverson L, Zucker S W, 1995 "Logical/linear operators for image curves" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **17** 982 – 996

Koenderink J J, 1984 "The structure of images" *Biological Cybernetics* **50** 363 – 370

Lee T S, Mumford D, 2003 "Hierarchical Bayesian inference in the visual cortex" *Journal of the Optical Society of America A* **20** 1434

Lindeberg T, 1998 "Feature detection with automatic scale selection" *International Journal of Computer Vision* **30** 79 – 116

McDermott J, 2000 *Experiments on Junctions in Real Images* MPhil thesis, University College London, London, UK

McDermott J, 2002 "Psychophysics with junctions in real images" *Journal of Vision* **2** 131a

McDermott J, Adelson E H, 2004a "Junctions and cost functions in motion interpretation" *Journal of Vision* **4** 552 – 563, http://www.journalofvision.org/4/7/3

McDermott J, Adelson E H, 2004b "The geometry of the occluding contour and its effect on motion interpretation" *Journal of Vision* **4** in press

McDermott J, Weiss Y, Adelson E H, 2001 "Beyond junctions: nonlocal form constraints on motion interpretation" *Perception* **30** 905 – 923

Malik J, 1987 "Interpreting line drawings of curved objects" *International Journal of Computer Vision* **1** 73 – 103

Marr D, Hildreth E, 1980 "Theory of edge detection" *Proceedings of the Royal Society of London, Series B* **207** 187 – 217

Martin D, Fowlkes C, Walker L, Malik J, 2003 "Local boundary detection in natural images: Matching human and machine performance" *Perception* **32** Supplement, 55 (abstract)

Morrone M C, Navangione A, Burr D, 1995 "An adaptive approach to scale selection for line and edge detection" *Pattern Recognition Letters* **16** 667 – 677

Parida L, Geiger D, Hummel R, 1998 "Junctions: detection, classification and reconstruction" *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20** 687 – 698

Perona P, 1992 "Steerable-scalable kernels for edge detection and junction analysis" *Image and Vision Computing* **10** 663 – 672

Rubin N, 2001 "The role of junctions in surface completion and contour matching" *Perception* **30** 339 – 366

Ruderman D, Bialek W, 1994 "Statistics of natural images—scaling in the woods" *Physical Review Letters* **73** 814 – 817

Sajda P, Finkel L H, 1995 "Intermediate-level visual representations and the construction of surface perception" *Journal of Cognitive Neuroscience* **7** 267 – 291

Saund E, 1999 "Perceptual organization of occluding contours of opaque surfaces", in *Computer Vision and Image Understanding* Special Issue on Perceptual Organization, Fall, 1999, pp 70 – 82

Shashua A, Ullman S, 1990 "Grouping contours by iterated pairing network" *Advances in Neural Information Processing Systems* **3** 335 – 341

Zucker S W, David C, Dobbins A, Iverson L, 1988 "The organization of curve detection: Coarse tangent fields and fine spline coverings", in *Second International Conference on Computer Vision* (New York: IEEE Computer Society) pp 568 – 577

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

PERCEPTION
VOLUME 33 2004

www.perceptionweb.com