In D. Knill and W. Richards (eds.), *Perception as Baysian Inference* (pp. 409-423). New York: Cambridge University Press (1996).

11

The perception of shading and reflectance

E. H. ADELSON

Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology

A. P. PENTLAND

Media Laboratory, Massachusetts Institute of Technology

11.1 Introduction

The luminance of a surface results from the combined effect of its reflectance (albedo) and its conditions of illumination. Luminance can be directly observed, but reflectance and illumination can only be derived by perceptual processes. Human observers are good at judging an object's reflectance in spite of large changes in illumination; this skill is known as "lightness constancy."

Most research on lightness constancy has used stimuli consisting of grey patches on a single flat plane. The models are typically based on the assumption that slow variations in luminance are due to illumination gradients, while sharp changes in luminance are due to reflectance edges. The retinex models for use with "Mondrian" stimuli are good examples (Land and McCann, 1971; Horn, 1974). But in three-dimensional scenes, sharp luminance changes can arise from either reflectance or from illumination, as illustrated in Figure 11.1(a). The edge marked (1) is due to a reflectance change, such as might result from a different shade of paint. The edge marked (2) results from a change in surface normal which leads to a change in the angle of incidence of the light -- an effect that we may simply refer to as "shading." As Gilchrist and his colleagues have emphasized (Gilchrist et al., 1983), three-dimensional scenes introduce large and important effects that are completely missed in the traditional approach to lightness perception.

11.2 Intrinsic image analysis

Using the terminology of Barrow and Tenenbaum (1978) we may cast the perceptual task as a problem of computing intrinsic images -- images that represent the underlying physical properties of a scene. To correctly interpret the scene of Figure11.1(a), one must derive a reflectance image, as

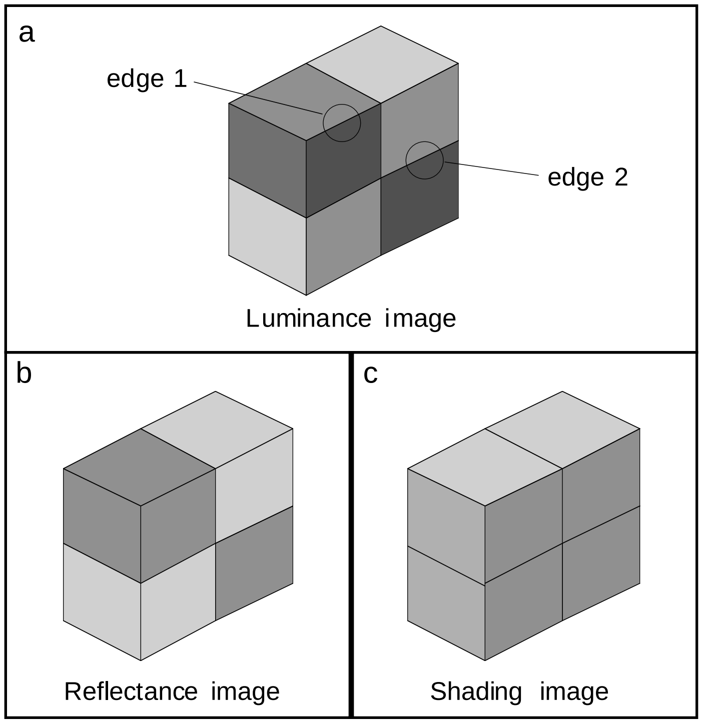

Fig. 11.1. (a) The two luminance edges marked 1 and 2 are exactly equivalent at a 2-D level, but are given different perceptual interpretations. Edge 1 is seen as a change in reflectance, while edge 2 is seen as a change in illumination due to a change in surface orientation. In an intrinsic image analysis, the image would be decomposed into the reflectance image (b) and the shading image (c).

shown in Figure 11.1(b), and a shading image, as shown in Figure 11.1(c). In addition one may derive images representing surface depth and orientation, which Marr called the 2 1/2 D sketch (Marr 1982).

In a scene consisting of Lambertian surfaces illuminated by a single distant light source, the observed luminance image I(x,y) is the product of the reflectance image, r(x,y), and the shading image (also termed the illuminance image), s(x,y),

$$I(x,y) = r(x,y)s(x,y) \tag{11.1}$$

where the variables (x,y) index the various points in these images. The shading image itself is the product of the luminous flux, l, and the cosine of the angle of incidence, i.e. the dot product (".") of the surface normal, N(x,y), and the illumination direction , L. Thus,

$$s(x,y) = \lambda N(x,y) \cdot L \tag{11.2}$$

Note that both the surface normal N(x,y) and the illumination direction L

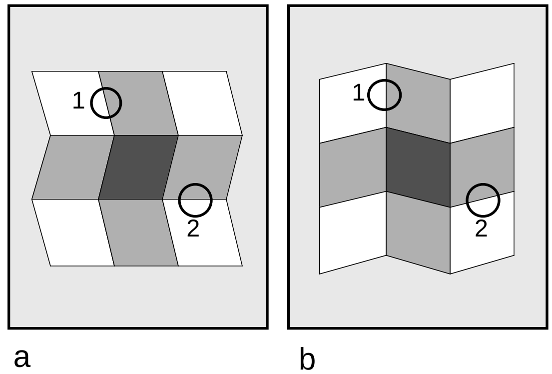

Fig. 11.2. (a) When the parallelograms are skewed horizontally, edge 1 is seen as a reflectance edge, while edge 2 is seen as a shading edge. (c) When the parallelograms are skewed vertically, edge 1 is seen as a shading edge, while edge 2 is seen as a reflectance edge. (c) Subjects adjusted the center-most patch to match the apparent reflectance of the patches above and below it. The matches were very different for the two images.

are three-dimensional vectors, but because they are defined as having unit length they have only two degrees of freedom.

A visual system must begin with the observed luminance image, $I(x,y)$, and infer the underlying shading and reflectance images, $s(x,y)$ and $r(x,y)$. Since there is no way to undo the multiplication by which the two images were combined, any mechanism for achieving the decomposition must make assumptions about regularities in the natural world.

The importance of three-dimensionality is further illustrated in Figure 11.2. Figure 11.2(a) and (b) each consist of 3x3 arrays of grey parallelograms, with the same shades of grey and the same adjacency relationships. The only difference between the images is the direction in which the parallelograms are skewed: horizontally in Figure 11.2(a) and vertically in Figure 11.2(b). If considered as mere 2-D arrangements of grey polygons, these images are quite similar. But our perceptual readings of these images are quite different.

In Figure 11.2(a), edge 1 is interpreted as a reflectance edge, while edge 2 is interpreted as a shading edge. On the other hand, in Figure 11.2(b), edge 1 is interpreted as a shading edge, while edge 2 is interpreted as a reflectance edge. The edges themselves, at the level of a local 2-D analysis, are equivalent. But the perturbations in the 2-D geometry lead to large changes in the 3-D interpretation, and these in turn lead to large changes in the way the edges are perceived.

The perceptual effects can be described in terms of intrinsic images. The Figure 11.3(a) the central vertical strip is seen as consisting of a single color in the 3-D object, while in Figure 11.3(b) the central vertical strip is seen as consisting of three distinct sections, the middle section being darker than the top or bottom.

It is worth nothing that the Retinex model will not correctly parse this image. It interprets sharp edges as belonging to reflectance boundaries,

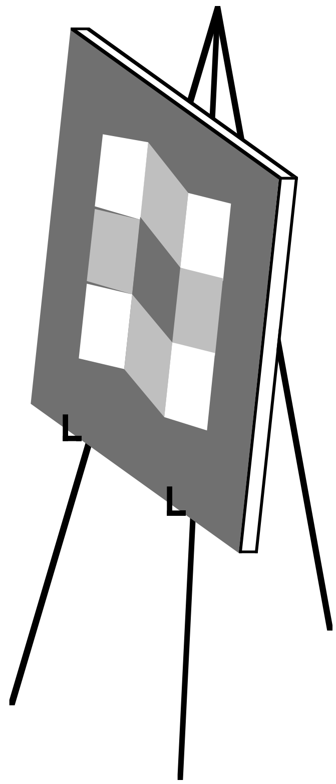

Fig. 11.3. The solution proposed by the painter. The scene consists of a flat surface, uniformly illuminated. All the image information is accounted for by variations in the grey tone (reflectance) of the paint.

and since all of the edges are sharp they will all be interpreted in terms of reflectance. In essence, the Retinex model interprets the image as a set of grey polygons on a flat surface, as it knows nothing about three dimensionality or the sharp illumination edges that can exist in 3-D scenes.

To deal with scenes in a three-dimensional world, one must employ three-dimensional constraints. We now discuss some approaches that can analyze polyhedral scenes of the sort shown in Figures 11.1 and 11.2. (See chapter 6 by Knill *et al*. for the use of geodesic constraint.)

11.3 The workshop metaphor

We begin by describing a "workshop" metaphor. Suppose that we are given the task of constructing a physical scene that will produce the image of Figure 11.2(b). We go to a workshop where a set of specialists build the scenery for the stage sets used in dramatic productions. One is a lighting designer; another is a painter; and a third is a sheet-metal worker. There is also a supervisor who can coordinate the actions of the individual specialists. We show them the desired image, and ask them to determine how to build a scene that will look the same. They are faced with a problem analogous to the one faced by the human visual system: given an image, try to figure out how it could have come about.

Let us imagine that the specialists charge according to a set of fixed prices. Simple and common operations are cheap, while more complex and unusual operations are more expensive. We can then cast the problem in terms of minimizing a cost function. The notion that a percept should correspond to the simplest or likeliest explanation of a scene has a long history in the perception literature (Helmholtz, 1962; Hochberg and McAlister, 1953; Attneave, 1959; Leeuwenberg, 1969; Restle, 1982;), and it has more recently been shown that formal concepts of simplicity (e.g. minimal length descriptions) and likelihood (e.g. maximum likelihood estimators) are fundamentally related (Pentland, 1989; Leclerc, 1989). These approaches can both be formalized as minimizing a cost function .

Consider the following fee structure:

Spray Painter Fees:

Paint rectang.patch:	$5 each.
Paint general polygon:	$5 / side.

Sheet metal Worker Fees:

Right angle cuts	$2 each.
Odd angle cuts	$5 each.
Right angle bends	$2 each.
Odd angle bends	$5 each.

Lighting Designer Fees:

Flood light	$5 each
Custom spot light	$30 each.

Supervisor Fees

Consultation	$30 / job.

Now there will be many different ways of constructing scenes that produce the same image. Indeed, each of the specialists can construct a model almost entirely without the help of the other specialists. For example, the painter could simply paint the appropriate arrangement of parallelograms on a flat sheet of metal and ask the lighting designer to illuminate it with a single flood; this solution is illustrated in Figure 11.3.

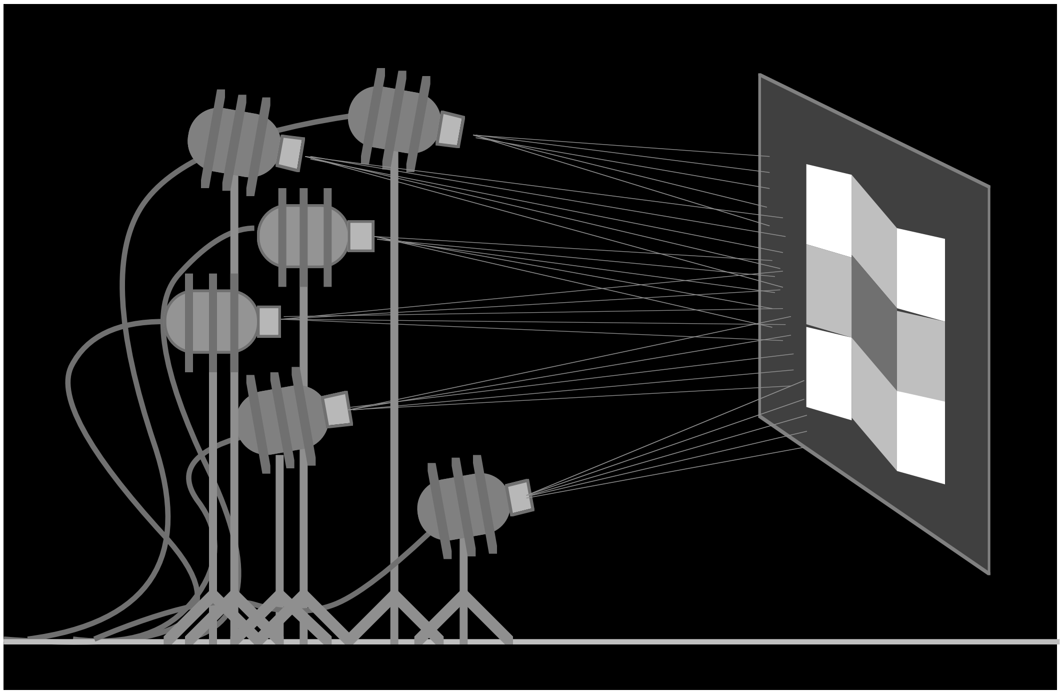

Fig. 11.4. The solution proposed by the lighting designer. The scene is assumed to be flat and of constant reflectance. All the image information is accounted for by variations in local illumination.

The lighting designer could start with a plain white sheet and project a set of nine custom spot lights onto it, having constructed a set of masks with just the right shapes, and projected at just the right positions and intensities so as to produce the desired image. Figure 4 shows this solution. It is also possible for the sheet metal worker to bend some metal sheets into very special shapes so that, when illuminated and viewed from precisely the correct angle, they will give rise to the desired image. This solution is shown in fig. 5. Finally, of course, the image could be produced by painting a square of metal with strips of two different shades of grey, and then bending the square into a zig-zag shape; this is the solution that leaps immediately to mind for a human observer. This last solution depends on the cooperation of the various specialists.

The prices for these solutions will be as follows:

Painter's solution:

Paint 9 general polygons:	$180.	
Setup 1 flood light	$5.	
Cut 1 rectangle		$8
Total		$193.

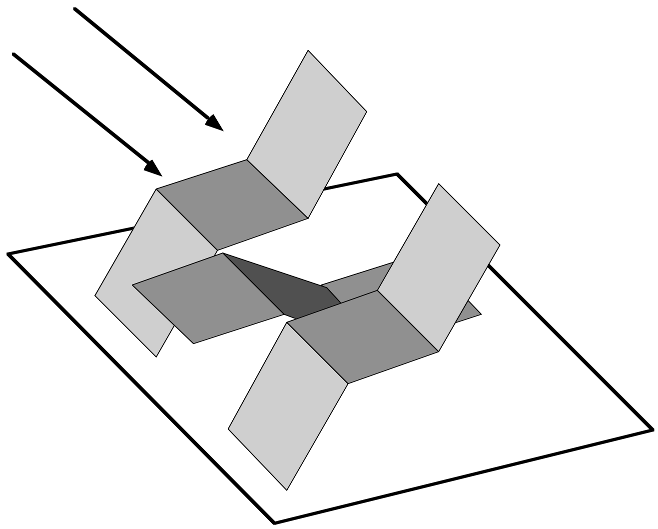

Fig. 11.5. The solution proposed by the sheet-metal worker. The scene is assumed to be of constant reflectance, illuminated by a single distant light source. All the image information is accounted for by the shading that results from the different surface normals. This scene can only be viewed from a single position, in order for the surfaces to line up properly.

Lighting Designer's solution:

Cut 1 Rectangle	$8.
Set up 9 Custom spots	$270.
Total	$278.

Sheet metal worker's solution:

Cut 24 odd angles	$120
Bend 6 odd angles	$30
Set up 1 flood light	$5
Total	$155.

Supervisor's solution:

Cut 1 rectangle	$8
Paint 3 rectangles	$5
Bend 2 right angles	$4
Supervisor's fee	$30
Total	$47.

It is clear that when each specialist tries to generate a solution on his own, the result works but is expensive. The lowest cost solution is the one suggested by the supervisor, which involves the efficient combination of all the skills of the specialists. Although the supervisor charges a fee of his own, it is more than offset by the savings that his cooperative solution allows.

The workshop metaphor is an example of a system that evaluates the cost of solutions and seeks the minimum. We do not intend it as a serious model of visual perception, but we feel it does highlight some important issues in vision.

One interesting point is the multiplicity of available solutions. This is related to the problem is inverse optics: there are many scenes that could produce a given image. And there are some solutions that are trivial to produce. Once we allow a painter into our stable of specialists, there is nothing to prevent him from explaining everything with paint. After all, any image we see might merely be a skillful piece of trompe de l'oeil. For that matter, it might be the result of a pattern of light cast by a slide projector on a white screen, or it might be the pattern of self-luminous dots on a CRT. Since these are legal solutions that are easy to construct they must be discouraged by other means. In terms of cost functions, we must make them expensive, indicating our preconception that they are less likely or less useful than other solutions. At the same time we must not make paint or light so generally expensive that they are never used; we need to find a balance in which they are used appropriately.

Another issue is how to assign the costs. In the example above the costs were simply chosen to make the story come out right. In a real system assigning the costs would have to be done more carefully. There are several ways that one might proceed. First, one could try out various cost schedules, tweaking them experimentally to see which ones led to the "right" answers, meaning the answers that humans see when they look at the same images. Second, one could do psychophysical experiments on humans, attempting to determine the costs that they assign to various aspects of the solutions. Third, one could empirically or theoretically determine the conditional probabilities that relate images to objects, and thereby estimate the proper costs that should be assigned so as to encourage likely interpretations and discourage unlikely ones [Dickenson, Pentland, and Rosenfeld 1992]. These all represent interesting avenues of exploration, but no one has yet undertaken them for the painted polyhedral objects we are using here.

Another problem becomes evident from the workshop metaphor. We have a well-defined cost function, which allows us to evaluate possible solutions, but we have no way of generating promising candidate solutions to be evaluated. When each specialist operates alone it is fairly easy to find a candidate solution but these solutions are usually poor. The good solutions -- the cooperative ones -- are much more difficult to find. In our story above the correct solution was simply announced by the supervisor, who unfortunately did not tell us how he found it. The problem can be described as one of searching the solution space and finding the point with minimum cost. But the space is enormous and there is no hope of simply searching it.

We have devised an algorithm that can correctly interpret images like those of figure 2. It begins with a description of the image in terms of a set of 2-D grey polygons, and generates a description of the 3-D shape, along with intrinsic images of the reflectance and the shading. The algorithm attempts to construct an internal model that accounts for the image data with minimum cost, where cost is defined so as to capture some of the intuitive notions of "simplest," or "most likely."

Our algorithm is based on a set of specialists, each of which is a subprocess utilizing knowledge about some particular aspect of visual scenes. For the problem at hand we employ a shape specialist, a lighting specialist, and a reflectance specialist. Each specialist seeks to explain what it can within its particular domain of expertise, and the three converge on a single solution.

The system that we will describe here is not actually cooperative across specialists. The shape process goes to work first and generates its best guess about the shape, seeking the 3-D configuration that explains the 2-D shapes with minimal cost. Then the lighting specialist seeks to explain as many grey-level edges as possible by adjusting the light source direction. Finally the reflectance specialist is allowed to explain whatever is left over. This particular hierarchy gives good solutions to many simple polyhedral images.

(1) *The shape specialist*: We assume the image was created by orthographic projection. The shape process is constrained by the observed (x,y) coordinates of edges and vertices, but it is free to vary the z coordinates, since these are not observed. The operation of the shape process may be understood by reference to the example in Figure 11.6. The input stimulus is shown in Figure 11.6(a); it is a 2-D image consisting of two parallelograms. Observers typically interpret this figure three-dimensionally, seeing it as folded in space. How can the 3-D percept be generated from the 2-D image? Our shape specialist uses a representation like that shown in Fig. 11.6(b). The vertices are like beads sliding on rods, and the lines between them are like infinitely elastic strings (cf. Arnheim, 1954; Barrow and Tenenbaum, 1981; Ullman 1984). The beads are constrained to maintain the observed (x,y) coordinates, but are free to move along the z-coordinate. The shape specialist can explore these configurations at will, since they all project orthographically to the same 2-D image.

We have experimented with various simplicity measures. In the case of quadrilaterals, such as the grey patches of Figure 11.2, one plausible notion of simplicity is that angles tend to be 90° angles, since squares and rectangles are the simplest quadrilaterals. A penalty (cost) is assigned to non-right angles, and the shape process seeks the 3-D configuration that minimizes this cost. This mechanism leads to the correct behavior for Figure 11.2, but it leads to incorrect configurations for other figures.

Barrow and Tenenbaum (1981) and Marill (1991) have proposed to inter-

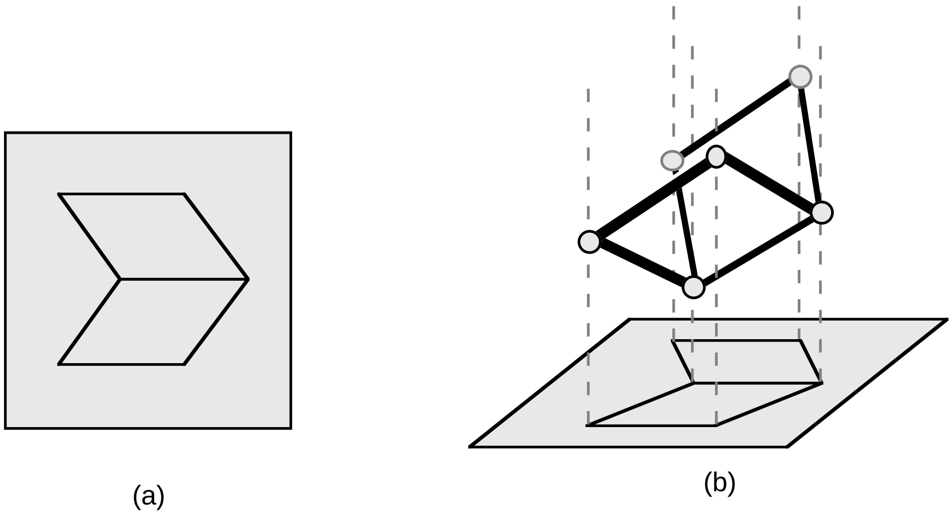

Fig. 11.6. (a) This 2-D image could be the result of projecting many possible 3-D shapes into the plane. In orthographic projection, the x and y coordinates are completely constrained by the image, but the z coordinate is completely unconstrained. (b) One may imagine a set of beads sliding on pegs, where the pegs constrain the beads to remain in the correct (x,y) position, but allow them to assume arbitrary z positions. All such configurations are legal interpretations of the image, but some configurations are simpler or more likely than others.

pret wire-frame drawings by minimizing a cost function that is proportional to the variance of the vertex angles. In the special case of quadralaterals it tends to prefer rectangles, and more generally it favors regular interpretations. But this approach leads to unsatisfactory configurations for many wire-frame figures. The main problem seems to be that humans prefer interpretations with planar faces, whereas the angle-variance cost function takes no account of whether the faces are planar or not.

We have found that by combining the angle-variance constraint with a planarity constraint, the behavior of the model is much improved. It is also advantageous to add a compactness constraint, so the the algorithm does not select configurations that are elongated along the line of sight (Sinha and Adelson, 1992, 1993). A similar algorithm has been described by Fischler and Leclerc (1992).

(2) *The lighting specialist*: The lighting specialist knows about the interaction of light with reflectance and surface orientation, as embodied in equation (1). It is given a single distance light source and is permitted to move it around so as to illuminate the object from various directions. (See chapter 9 by Freeman.)

The lighting specialist also knows about the shape specialist's current estimate of 3-D shape, and it uses that estimate to calculate the effects of different lighting directions. The optimal lighting direction is the one that explains as much of the luminance variation as possible in terms of shading, thereby minimizing the need for reflectance edges.

The lighting specialist starts by assuming the current estimate of surface reflectance r(x,y), surface normal N(x,y), and illuminant intensity l. Then for each image edge the specialist produces two equations, one for each side of the edge:

$$I_1 = r_1 \lambda N_1 \cdot L \tag{11.3}$$
$$I_2 = r_2 \lambda N_2 \cdot L \tag{11.4}$$

Each variable in these two equations is known except the two components that make up the light direction L so that for each edge we have two linear equations in two unknowns and so may directly solve for L. By combining the equations from all the image edges into a single linear regression we can, therefore, determine the best overall estimate of L.

(3) *The reflectance specialist*: This process assigns a reflectance to each region of the image. It must take care of any image data that is not explained by the shape process and the lighting process. The cost associated with reflectance edges is not explicitly evaluated. It is the responsibility of the lighting specialist to minimize the need for paint; it is the responsibility of the reflectance specialist to take care of any luminance variation not explained by the lighting specialist.

11.4 An example

We ran the algorithm on the zig-zag shape shown in Figure 11.7(a). The starting interpretation is shown in Figure 11.7 (b-e). The 3-D shape, shown in an oblique view in Figure 11.7(b), is initially assumed to be flat, as if the object were merely a 2-D painting lying on a table. (The image in Figure 11.7(b) may appear to be slightly folded, but it is actually flat, consisting of a set of adjoining parallelograms). The light source is initially assumed to be head-on, from the direction of the eye, as indicated in the spherical plot of Figure 11.7(c). In these conditions there is no opportunity for shading to produce luminance variation, and so the shading image is completely uniform, as in Figure 11.7(d). The reflectance specialist is therefore initially responsible for explaining all of the image luminance information, which in this case means that it replicates the original image, as shown in Figure 11.7(e). In summary, when the algorithm begins it assumes that the image is just a painting that is flat and uniformly lit.

After the algorithm is run, the interpretation settles into the configuration shown in Figure 11.7(f-i). The shape specialist finds a 3-D shape , shown in side-view in Figure 11.7(f), in which the panels are square in shape, and the folds are at right angles, and which therefore has minimal cost. The lighting

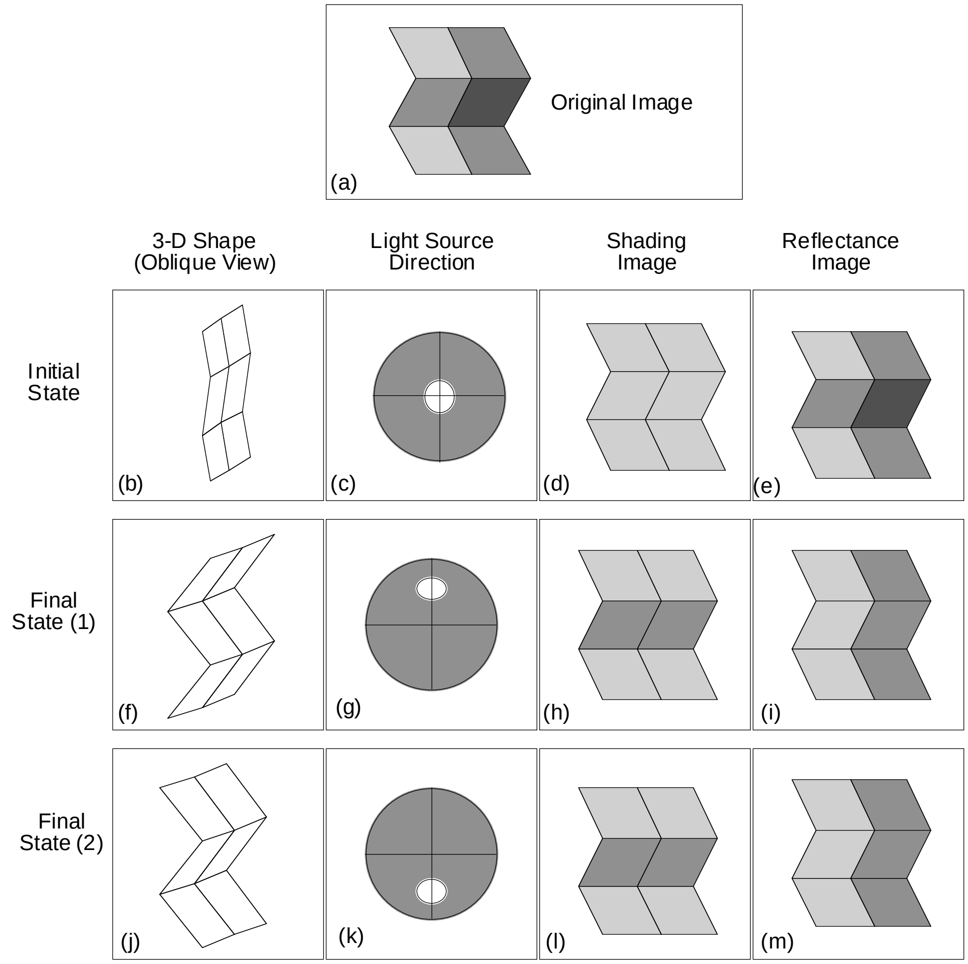

Fig. 11.7. (a) The original image that is to be analyzed. (b) The algorithm initially represents the object as being flat; the observed parallelograms are assumed to be actual parallelograms which are literally the shape of the panels in the image. The shape is shown here from an oblique view (c) The light source direction is initially assumed to be head-on, as diagrammed here by position on the surface of a sphere. (d) Since the object is initially taken to be flat, there is no variation in illumination across the surface, and so the shading image is constant. (e) The reflectance image is left with the task of explaining all of the luminance variation. (f) After the algorithm has arrived at a minimal cost configuration, the object is represented as a set of square panels which join each other in a zig-zag configuration. (g) The light source moves up into a position such that shading can account for as much luminance variation as possible. (h) The shading image accounts for the horizontally oriented luminance edges. (i) The reflectance image accounts for the remaining luminance variation. Only two colors are required. (j) An alternate final state, with the same minimum cost, occurs when the object is reversed in depth. (k) The light source moves down, rather than up, in this case. (l) The shading image and (m) the reflectance image are exactly the same in the depth-reversed case as they are in the first case.

specialist finds that by placing the light at the position shown in Figure 11.7(g), a maximum number of luminance edges can be explained in terms of shading rather than reflectance. This leads to the shading image of Figure 7(h). Finally, the reflectance specialist takes care of the remaining luminance variation by assigning the reflectances shown in Figure 7(i). The final interpretation is similar to that reported by human observers.

Another interesting aspect of the human percept is bistability: if one looks at the figure for a while it will spontaneously reverse in depth. Although our algorithm does not undergo spontaneous reversals, it does assign the same cost to each of the reversed states, and considers each to be an equally good global minimum. It will randomly settle into one or the other interpretation depending on the starting conditions. The reversed-depth percept is shown in Figure 7(j). When the shape specialist chooses this interpretation, the lighting specialist automatically moves the light source to the correspondingly reversed angle, thereby maintaining a consistent interpretation, as shown in Figure 7(k). This again is consistent with the perceptual reports of human observers. Note that when the 3-D shape and the light source direction both reverse, the shading and reflectance images, Figure 7 (l) and (m), settle into the same states as before, as they should.

The search for a minimum cost in the above example is particularly simple because the cost function for this scene has no local minima, and has only two global minima, corresponding to the two depth-reversed solutions. For this reason we were able to solve the example without confronting the complexities that will necessarily emerge with more convoluted cost functions. In order to deal with more complex scenes, such as those involving curved surfaces or occlusions, as well as complicated forms of lighting, one will need more sophisticated specialists and more sophisticated control structures. Nonetheless we are encouraged by the capabilities of the simple algorithm described here.

11.5 Conclusions

Most models of human lightness perception have been designed to deal with images consisting of simple arrays of grey patches on a flat field. These stimuli are devoid of cues about three dimensionality, and the corresponding models are unable to deal with the percepts of reflectance edges and illuminance edges as they are seen in ordinary three dimensional scenes.

At the same time, models for shape-from-shading typically assume that the world is all of a constant reflectance, and that three-dimensional shape is responsible for all the luminance variation observed. Such models cannot deal with patches of varying reflectance.

To model the perception of lightness and shading in three dimensional scenes, we must turn to systems that have richer vocabularies. This means that there are many ways to explain the luminance variation in a given image; the representation language is highly overcomplete. We introduce the "workshop" metaphor as a way of exploring the possibilities and problems of such systems. We imagine a set of specialists who have particular expertise about paint, or lighting, or 3-D shape, and each of which can explain the observed images without the help of the others. The problem then becomes selecting a description that makes proper use of these various sources of expertise. We have also developed an algorithm based on these ideas to illustrate one concrete instantiation. The output consists of a three-dimensional model, a reflectance image, a shading image, and a light source direction. For some simple scenes, the algorithm can produce interpretations similar to those reported by human observers; it is also consistent with the perception of reversible figures, and when depth reversal occurs it infers an appropriate change in lighting conditions.

References

Arnheim, R., (1954). *Art and Visual Perception*. Berkeley: University of California Press.

Attneave, F. (1959). *Applications of Information Theory to Psychology*. New York: Rinehart and Winston.

Barrow, H. G., and Tenenbaum, J. M. (1978). Recovering Intrinsic Scene Characteristics from Images. In *Computer Vision Systems*, ed. Hanson, A. R., and Riseman, E. M., 99. 3-26. Academic Press.

Barrow, H. B., and Tenenbaum, J. M. (1981). Interpreting line drawings as three-dimensional surfaces. *Artificial Intelligence,* **17**, 75-116.

Dickenson, S., Pentland, A., and Rosenfeld, A. (1992). 3-D shape recovery using distributed aspect matching, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **14**(2):174-198.

Fischler, M. A., Leclerc, Y. G. (1992). Recovering 3-D wire frames from line drawings. *Proceedings of the Image Understanding Workshop.*

Gilchrist, A., Delman, S., and Jacobsen, A. (1983) The classification and integration of edges as critical to the perception of reflectance and illumination. *Perception &Psychophysics*, **33**, 425-436.

Helmholtz, H. von (1962). *Treatise on Physiological Optics*, Vol. III (trans. from the 3rd German ed., J. P. C. Southall), New York: Dover.

Hochberg, J., and McAlister, E. (1953). A quantitative approach to figural goodness, *J. Exp. Psych*., **46**, 361-364.

Horn, B. K. P. (1974). Determining lightness from an image. *CGIP*, **3**, 277-299.

Land, E. H., and McCann, J. J .(1971). Lightness and retinex theory. *J. Opt. Soc. Am*., **61**, 1-11.

Leclerc, Y., (1989) Constructing simple stable descriptions for image partitioning. *Intl. J. Comp. Vision*, **1**, 73-102.

Leeuwenberg, E. (1969). Quantitative specification of information in sequential patterns. *Psych. Rev*., **76**, 216-220.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Marill, T. (1991). Emulating the human interpretation of line-drawings as three-dimensional objects. *Intl. J. of Comp. Vis*., **6**, 147-161.

Pentland, A. (1989). Part segmentation for object recognition. *Neural Computa-*

tion, **1**, 82-91.

Restle, F. (1982). Coding theory as an integration of Gestalt psychology and information processing theory. In *Organization and representation in perception*, ed. J. Beck, pp. 31-56. Hillsdale NJ: Erlbaum.

Sinha, P., and Adelson, E. H. (1992). Recovery of 3-D shape from 2-D wireframe drawings. *Assoc. for Res. in Vision and Opth.*, May 1992, *Inv. Opth. and Vis. Sci.* (supp.) **33**, p. 825.

Sinha, P., and Adelson, E. H., (1993) Recovering reflectance in a world of painted polyhedra. *Proc. Fourth Intl. Conf. Comp. Vision*, Berlin, May 1993.

Ullman, S. (1984). Maximizing rigidity: the incremental recovery of 3-D structure from rigid and rubbery motion. *Perception*, **13**, 255-274.