In *Proceedings of the IEEE Workshop on Qualitative Vision*, June 14, 1993, New York

Verifying The 'Consistency' Of Shading Patterns And 3-D Structures

Pawan Sinha[1] & Edward Adelson[2]
[1]NE43-809, Artificial Intelligence Laboratory
[1,2]Vision & Modeling Group, Media Laboratory
Massachusetts Institute of Technology
Cambridge, MA 02139

The problem of interpreting images in terms of their shading and reflectance components has traditionally been addressed as an early vision task in a simple 2D Mondrian domain. Recently it has been appreciated that in a 3D world, such conventional approaches are inadequate; more sophisticated strategies are required. One such strategy has been proposed by Sinha [22, 25], who has addressed the problem as a mid-level vision task rather than as a purely low-level one. Sinha suggested that a key computation that needs to be performed for interpreting images acquired in a 3D domain is the verification of the consistency of image shading patterns and the likely 3D structure of the scene. This is the problem we have addressed in the present paper. Considerations of robustness and generality have prompted us to discard available quantitative techniques in favor of a qualitative one. The two prime attributes of our technique are its use of qualitative comparisons of gray-levels instead of their precise absolute measurements and also its doing away with the need of an exact pre-specification of the surface reflectance function. We show that this idea lends itself naturally to a linear-programming solution technique and that results obtained with some sample images are in conformity with human perception.

1. INTRODUCTION:

Why is it that some images (such as figure 1(a)) appear to depict 'properly' shaded 3-D objects while others (such as figure 1(b)) give the impression of flat painted structures? It has recently been suggested that the percept obtained depends on the 'consistency', or lack thereof, of the gray-level patterns in the image and the likely 3-D structure of the underlying scene [22,25]. In other words, a pattern of gray-levels is interpreted as 'shading' if it can be produced by illuminating a uniform albedo 3-D structure with a single distant light-source (see [21] for a justification of the single source assumption). The considered 3-D structure must, of course, be projectionally consistent with the 2-D geometric configuration of the gray-level pattern. Of the infinitely many 3-D structures consistent with the pattern's geometric structure, we restrict our attention to those that are perceptually likely (characteristics of perceptually likely 3-D structures are discussing in [23,24]).

This research was supported in part by contracts to the MIT Media Laboratory form Goldstar Co. and the TVOT Program

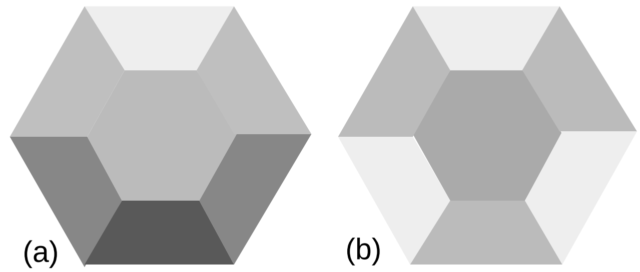

Figure 1. Two figures that have identical geometric structures but very different interpretations. Figure (a) is perceived to be a shaded 3-D truncated hexagonal pyramid while figure (b) appears to be a flat pattern of paint.

If the gray-level pattern cannot be produced simply by illuminating a 'perceptually likely' uniform albedo 3-D structure, then the hypothesis of spatial variation of reflectance needs to be invoked. In such a situation, we suggest, the image will appear to depict a pattern of paint.

From the discussion above, it is apparent that a method for interpreting gray-level patterns needs to perform two conceptually different tasks:

1. it needs to recover a set (possibly singleton) of perceptually likely 3D structures consistent with the geometric structure of the input pattern, and

2.it needs to verify whether (any of) the recovered 3D structures can be illuminated with a single distant light source so as to produce a gray-level pattern equivalent to the input pattern. This, in essence, is an attempt to determine whether the variations in image gray-levels are due to shading or changes in intrinsic surface reflectance.

Figure 2 shows this two part strategy. Schemes for accomplishing task 1 have been described in [8, 22, 24]. In section 3, for the sake of completeness, we provide a brief account of our solution to this problem. The main focus of this paper is, however, task 2. Beginning with section 4, we present a discussion of the inadequacies of existing quantitative approaches and propose a qualitative solution technique for this task.

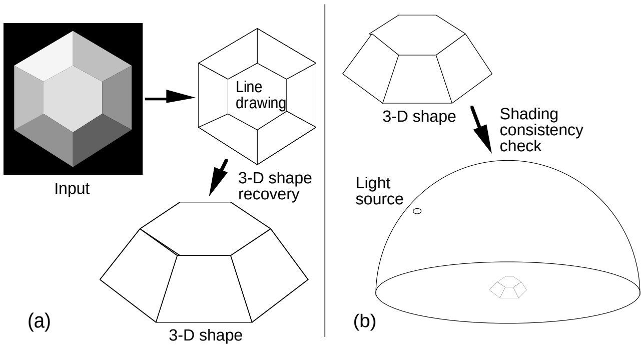

Figure 2. (a) Phase I of the proposed strategy involves deriving the likely 3-D structure(s) corresponding to the geometric configuration of the image. (b) Phase II of the strategy verifies consistency of the shading pattern in the image and the recovered 3-D shape. This involves determining whether the 3-D shape can be illuminated by a single distant light source to yield the shading pattern observed in the original image.

2. THE WORLD MODEL:

The domain we shall be concerned with in this paper comprises of painted polyhedral/origami objects such as those shown in figure 3. We assume an absence of cast shadows. The surfaces shall be assumed to be qualitatively matte without necessarily being precisely Lambertian. The scene shall be assumed to be illuminated with a diffuse ambient and a single distant light source.

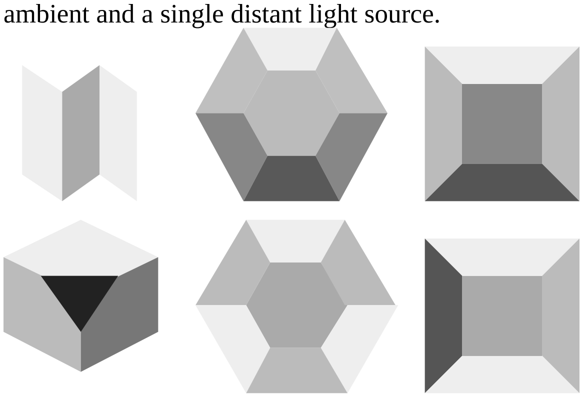

Figure 3. Sample objects from our domain of interest.

3. DERIVING 3-D SHAPES FROM 2-D LINE-DRAWINGS:

As stated earlier, our aim here is to interpret 2-D line-drawings extracted from the input gray-level patterns in terms of their perceptually/physically likely 3-D structures. The difficulty of this task arises from its highly underconstrained nature; any planar line-drawing is geometrically consistent with infinitely many 3-D structures, as shown in figure 4. In light of this observation, two questions that need to be addressed are: 1.what distinguishes the 'correct' 3-D structure from the rest?, and 2.how might we search for the 'correct' structure in the infinite space of all possible 3-D structures consistent with the given line-drawing?

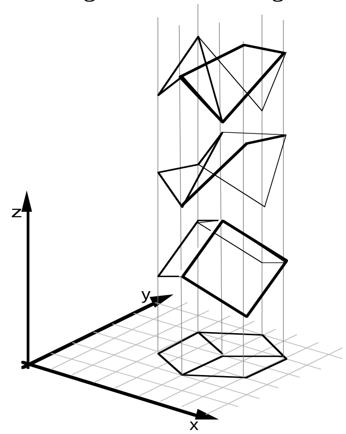

Figure 4. Any planar line-drawing is geometrically consistent with infinitely many 3-D structures.

It has long been suggested that the distinguishing characteristic of a perceptually favored 3-D interpretation is its low 'complexity'. The variance of the included angles has been proposed as a measure of complexity [2, 19]; minimizing this metric leads to perceptually correct interpretations for many line drawings. However, we find that using this metric alone results in unexpected and bizarre interpretations for certain figures (see figure 5). We propose that to properly characterise the perceptually 'correct' interpretations, three types of measures are required: angle variance, planarity of faces and overall compactness; we wish to obtain that 3-D configuration which, while having planar faces is maximally regular and compact so as not to require excessive foreshortening of any line segment to relate it to the input 2-D drawing. A similar suggestion was made in [8].

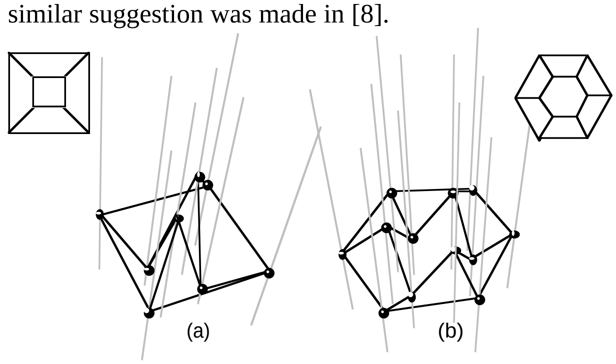

Figure 5. Perceptually incorrect shapes recovered from the input line-drawings by Marill's algorithm. The shapes are shown as states of a 'beads-on-wires' model. The 'wires' are aligned to the line-of-sight and the positions of the beads represent the depth values associated with each vertex.

The question of how to search for the desired configuration in the infinite search-space of all possible configurations is a tricky one. Traditional approaches usually involve formulating and then optimizing a composite cost function (such as a weighted sum of the relevant metrics [8]). This approach suffers not only from the need to make *ad hoc* choices for the relative weights but also has the same drawbacks that regularizing techniques have, viz., the constructed cost-function might not represent the original problem. Also, the parameter values that might be appropriate for one problem instance might be inappropriate for another.

Our search strategy belongs to the class of 'greedy' optimization algorithms [5]. Here we give only a conceptual description and refer the reader to [24] for a formal treatment of the same. Imagine that one is given a 2-D line drawing that one wishes to derive the maximally regular planar 3-D shape of. What kinds of intermediate stages should one expect to pass through on way to the final configuration? One natural way of getting to the desired shape is to incrementally modify the originally planar configuration so that at every intermediate step the most regular planar faceted configuration is obtained. This can be thought of as doing gradient descent in regularity space where the points considered in the space correspond to the different planar-faceted 3-D configurations. The first local minima reached in this fashion is reported as the recovered 3-D shape. This strategy does not require the construction of one composite cost function from different metrics. Besides obviating the need for *ad hoc* choices of parameters, this also has the desirable result of having the same algorithm work unchanged on all problem instances. Figure 6 shows two sample results.

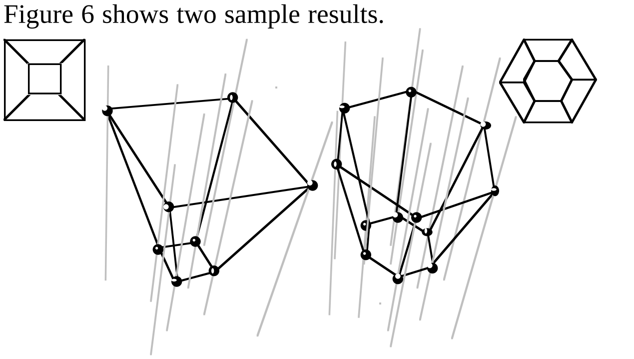

Figure 6. Two examples of 3-D shape recovery using constraints of regularity, planarity and compactness.

4. CHECKING SHADING CONSISTENCY:

The problem of checking for the consistency of shading may formally be stated as follows: Given a 3-D shape and a 2-D gray-level image (which is already known to be *geometrically* consistent with the 3-D object), we need to determine whether there exist any light source directions that would completely account for all gray-level variations in the image without having to invoke the hypothesis of surface reflectance changes. If not, then we would like to know which edges cannot be accounted for simply by illumination variations.

Given a 3-D structure and the gray-levels associated with each of its faces, the problem of determining the source direction under the assumption of a precisely specified reflectance function does not seem too difficult. Indeed, closed-form solutions for this task have already been described [12, 21, 26]. These methods, however, have some fundamental limitations. We highlight these limitations below with the aid of an example.

4.1 Quantitative techniques for illuminant direction computation - an example:

Let us assume that the surfaces of the object considered satisfy the Lambertian reflectance model, i.e. the brightness E of a surface whose normal is inclined ß degrees with respect to the incident light direction is µ cos(ß).

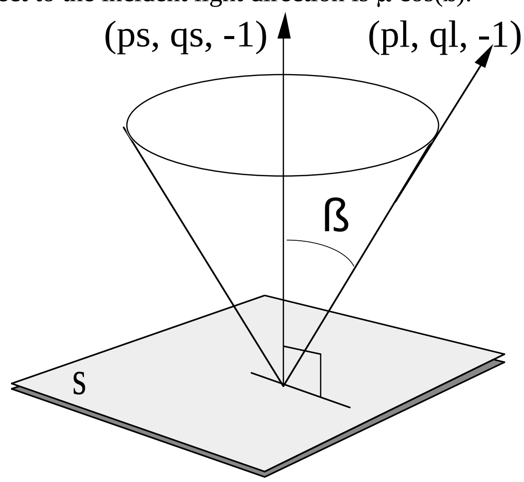

Figure 7. Under the Lambertian reflectance model, knowledge of the surface normal and the brightness E of the surface constrains the set of valid light directions to lie on a cone with a half angle of arc cos (E).

µ is the surface albedo and for simplicity is assumed to be unity. It is clear then, that knowing the orientation of the surface normal and the precise brightness of the surface, the set of valid light directions define a cone with a half angle of ß (where ß = arc cos (E)) as shown in figure 7. Now consider a multifaceted polyhedral object. Assume that the normal vector and measured brightness of surface *i* are n_i and E_i respectively. Each facet then defines a cone of valid light directions. To determine whether there exists a light source direction that can simultaneously account for the measured brightnesses of *all* the surfaces, we need to verify whether the cones associated with each surface have any common directions. The idea is easier to understand in gradient space where a cone of directions maps to a 2-D curve. Thus, for the case of a 'sliced-cube' shown in figure 8, we have four curves (see figure 10) and depending on whether or not they have a common point of intersection, the shading pattern observed in the 2-D image should be reported as consistent or inconsistent with the 3-D shape. Despite the reasonableness of the ideas so far, we can begin to see some chinks in the armor of such an approach.

Figure 8. A 'sliced-cube'.

The major limitation of the quantitative approach is its critical dependence on precise measurements of image brightnesses. To understand why this is a limitation, consider an object with four faces (such as the 'sliced-cube' of figure 5) with surface normals n_1, n_2, n_3, and n_4 and let the light direction be n_l. Let the image-irradiance corresponding to face i be e_i and let the intensity of the light source be I. Then, assuming Lambertian surfaces, $e_i = I \times Cos\theta_i$ where $Cos\theta_i = \overline{n_i} \cdot \overline{n_l}$

Now imagine altering the gray-level of one of the four faces slightly leaving the others and their relative brightness ordering unchanged. The question we are interested in is whether in this altered situation, we can find a light direction $\overline{n'_l}$ and illumination intensity I' that will completely account for the observed brightnesses of the faces. For convenience and without loss of generality, we may assume I to be unity. Then,

$$\overline{n_1} \cdot \overline{n_l} = I' \, \overline{n_1} \cdot \overline{n'_l} \equiv e_1$$
$$\overline{n_2} \cdot \overline{n_l} = I' \, \overline{n_2} \cdot \overline{n'_l} \equiv e_2$$
$$\overline{n_3} \cdot \overline{n_l} = I' \, \overline{n_3} \cdot \overline{n'_l} \equiv e_3$$
$$\alpha\overline{n_4} \cdot \overline{n_l} = I' \, \overline{n_4} \cdot \overline{n'_l} \equiv e'_4$$

where α is the factor by which the brightness of face 4 has been altered. We clearly have four equations in three unknowns (I' and the two components of the unit vector $\overline{n'_l}$). This, therefore, constitutes an overdetermined system of equations which may not admit any solution. Thus, a slight alteration of the brightness of a face may make the quantitative approach conclude that the observed shading pattern in the image is inconsistent and that the given object cannot be illuminated in the manner shown. The hypothesis of reflectance variations would be invoked to account for the image brightness patterns.

For a more intuitive understanding of this discussion, let us study its geometric correlates. Consider what happens when we slightly alter the gray-level of one of the faces of the sliced-cube in figure 8. The before and after versions of the sliced-cube are shown in figure 9. This alteration of the gray level alters the half angle of the associated cone of valid light directions and consequently, the 2-D curve corresponding to the cone in gradient space (see figure 10). The new curve no longer intersects the other three curves at the same position; there is now no common point of intersection of the four curves and the quantitative approach is forced to conclude that the observed shading pattern in the image is inconsistent. That is certainly not the way it looks like to a human subject when he/she examines the two images in figure 9. Both the objects look consistently shaded. Minor perturbations in the precise quantitative values of the surface gray-levels do not seem to be of much consequence perceptually. In fact, the gray-levels of the surfaces can be altered significantly without producing the percept of inconsistency, so long as their relative ordinal relationships are left intact. The ordinal relationships seem to be far more important than the precise gray-levels of the individual surfaces. Dependence on precise gray-level values is one major limitation of quantitative approaches.

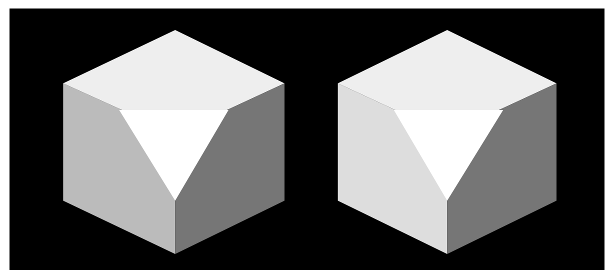

Figure 9. The image on the right is derived from the one on the left by a slight alteration of the gray-level of one of the faces.

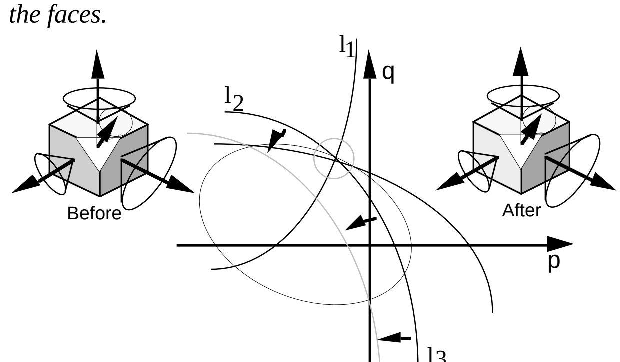

Figure 10. The various faces of a multi-faceted polyhedral object define curves of valid light directions in gradient space [17]. The common point of intersection of all these curves corresponds to the light direction that would account for the brightnesses of all faces simultaneously. A slight alteration in the gray-level of one of the faces shifts the corresponding locus of valid light directions in gradient space and the four curves (corresponding to the valid source directions for the four visible faces in this example) no longer have a common point of intersection suggesting that the sliced-cube on the right is no longer consistently shaded. Perceptually, however, the minor gray-level alteration is inconsequential.

Quantitative approaches suffer from yet another limitation, which is related to their critical reliance on a precisely pre-specified surface reflectance function. Minor alterations in this function profoundly influence the computed solution. This is a serious drawback considering that in most situations, the choice of the reflectance function is at best an educated guess. It seems unreasonable to demand that the detailed mathematical characteristics of surfaces' reflectance functions be known before their images can be interpreted.

What we seek to have is a method that is more generally applicable (by not assuming a precise reflectance function) and more robust (by not being thrown off by small perturbations of the gray-levels in an image). To do so, it is not sufficient to simply patch-up the existing quantitative approaches with, say, an idea like using least-squares error minimization; doing so is like trying to push the basic causes of the drawbacks under the rug and then pounding on it to make the bump disappear. We need a qualitatively different paradigm.

4.2 Determining light directions from shaded images of polyhedral objects - a qualitative approach:

One of the key motivating observations behind our approach is that our perceptual apparatus is far more sensitive to detecting relations like 'brighter than'/'darker than' between pairs of adjacent surfaces than to estimating their absolute brightnesses. Perceptual interpretations of images are quite stable over alterations in image gray-levels that leave the binary relations between adjacent pairs of surfaces unaltered (in a sense, these relations define perceptual equivalence classes for images). In our approach, we use only such binary relations extracted from the underlying images.

The other key idea is to use these relations to constrain the source direction in a manner that least commits us to a particular reflectance function. Consider figure 11. If *S1* and *S2* are two surfaces with normals *n1* and *n2* (remember that the 3D shape of the object has already been recovered in phase-1 of the strategy) and *S1* appears darker than *S2* in the image, then the valid light directions can be represented on a Gaussian sphere by one of the two hemispheres formed by a plane passing through the origin that perpendicularly bisects the vector joining the tips of the normals *n1* and *n2* on the sphere surface. This set of light directions is valid for any reflectance function that results in a monotonically decreasing relationship between image irradiance and angle of incidence. We may further constrain the light directions to lie above the horizontal plane. A light direction chosen from this set will maintain the ordinal relationship between the brightnesses of surfaces *S1* and *S2*. Other pairs of surfaces will similarly define hemispheres of valid light directions. The complete shading pattern can be considered consistent if the intersection of all the hemispheres corresponding to different adjacent surface pairs yields a finite set. The problem of checking for the consistency of the observed shading pattern is thus rendered equivalent to determining whether a set of hemispheres have a non-null intersection. The non-null intersection set, if obtained, represents the valid set of light directions. Since (as shown below) each hemisphere is equivalent to a linear constraint on the possible positions of the source vector, this approach lends itself naturally to a linear programming solution method such as the Fourier-Motzkin elimination technique [9, 6, 14]. Interestingly, the Perceptron Learning algorithm [7, 20] is also perfectly suited to solving this problem. This approach also has the desired properties of not being critically dependent on precise measurements of the absolute surface brightness values and not having to assume a precisely specified formal reflectance model.

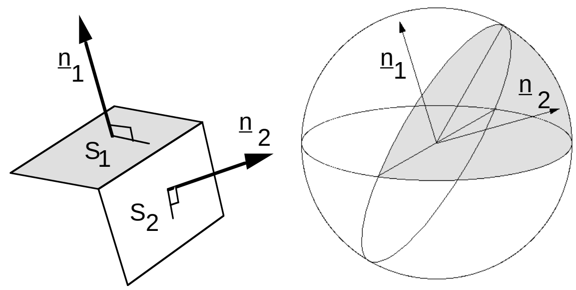

Figure 11. Two surfaces of known orientation and their relative brightness values constrain the light source to lie in a particular sector of the Gaussian sphere. See text for details.

<u>*4.2.1. A linear programming approach to determining light directions from shaded images of polyhedral objects:*</u>

Consider the pair of surfaces S1 and S2 with normals $\bar{n}_1$ and $\bar{n}_2$ respectively. Compute a vector $\bar{s}$ such that

$$\bar{s} \;.\; ((\bar{n}_1 + \bar{n}_2) / 2) = 0;$$
$$\bar{s} \;\mathrm{x}\; (\bar{n}_1 \;\mathrm{x}\; \bar{n}_2) = 0;$$
$$\bar{s} \;.\; \bar{n}_1 > 0 \text{ and } \bar{s} \;.\; \bar{n}_2 < 0$$
if S1 is brighter than S2 in the image
$$\bar{s} \;.\; \bar{n}_1 < 0 \text{ and } \bar{s} \;.\; \bar{n}_2 > 0$$
if S2 is brighter than S1 in the image

The hemisphere of valid directions defined by the surface pair S1 and S2 then is precisely the set of vectors t satisfying the inequality s . t > 0. To constrain the valid directions to lie above the horizontal plane, we may wish to enforce the additional inequality z . t ≥ 0 (assuming without loss of generality that the ground plane is the X-Y plane). For each adjacent pair of surfaces Si and Sj, we get one such linear inequality , viz. sij . t > 0. We wish to find a vector t (if it exists) that satisfies *all* these inequalities. This is a simple linear programming problem. There are 'e' linear inequalities for a polyhedral object with 'e' internal edges. Since we are interested only in the direction of t, there are only two degrees of freedom to be solved for. As no objective function is being extremized, there will exist an infinite number of solutions if there are any solutions at all. All of these solutions will lie in a convex polygon on the unit sphere. The sides of this polygon are portions of great circles corresponding to constraints imposed by some surface pairs (see figure 12). If a unique solution is desired, the center of the solution polygon may be chosen to be the one.

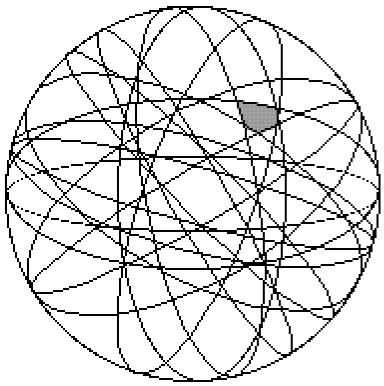

Figure 12. The solutions to the system of constraints set up by the various surface pairs lie on a convex polygon on the unit sphere. The sides of this polygon are portions of great circles corresponding to constraints imposed by some surface pairs.

4.2.2 Determining the illuminant direction - some examples:

In this section, we present some examples illustrating the use of the aforementioned ideas for checking the consistency of the observed shading pattern in the image and for recovering the illuminant direction.

The first two examples comprise of a cube illuminated from two different directions. The graphical solutions (figures 13 and 14) show that the recovered sets of valid light directions are consistent with human perception.

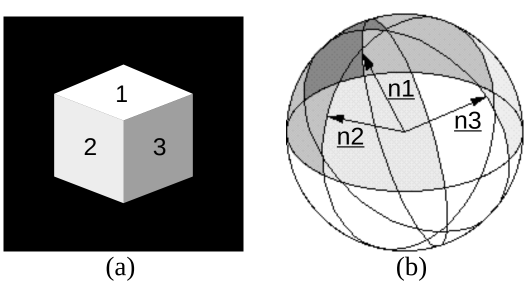

Figure 13. Computing the set of valid light directions for a cube illuminated in the manner shown in figure (a). The darkest sector of the Gaussian sphere in (b) represents the solution set. The ni's are the surface normals corresponding to the three visible faces of the cube.

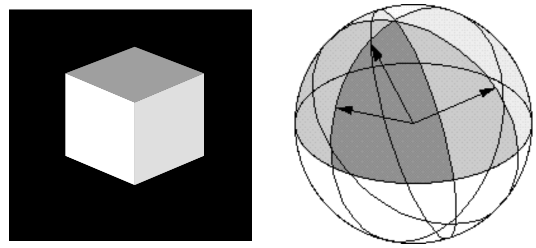

Figure 14. Varying the ordinal relationships between the gray-levels of the faces of the cube changes the computed solution set for illuminant directions in a manner consistent with human perception.

The next two examples are more interesting in that they highlight the difference between a consistently and an inconsistently shaded image. The reader may recall having seen these images in the introductory section as figures 1(a) and 1(b).

The 3D structures constructed by the shape recovery module described in [24] from the geometric configurations of the two images are truncated hexagonal pyramids. The task at hand is to verify the consistency of the image gray-level patterns with respect to these 3D structures. The graphical solutions, shown in figures 15 and 16, suggest that while the shading pattern in figure 1(a) is consistent with the shape recovered by the module responsible for 3-D shape recovery from 2D line-drawings, that of figure 1(b) is not.

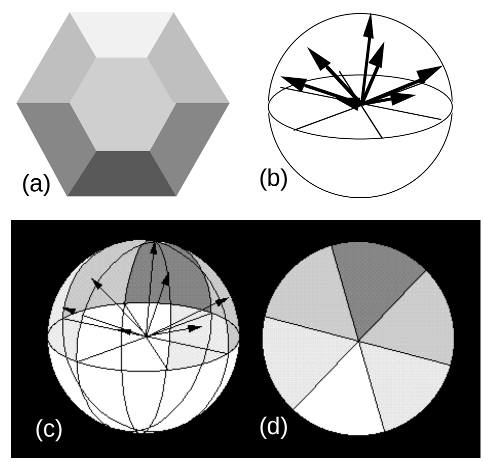

Figure 15. Determining the valid light directions for figure (a) with associated 3-D structure shown in figure (b) using the Gaussian sphere/surface normal representation. The 3D structure assumed (and the one that is recovered by the 3D shape recovery module described in [Sinha & Adelson, 1993a]) is that of a truncated hexagonal pyramid. The darkest sectors in (c) and (d) represent the computed solution set. Figure (d) shows the Gaussian sphere in (c) viewed with the line of sight aligned to the polar axis.

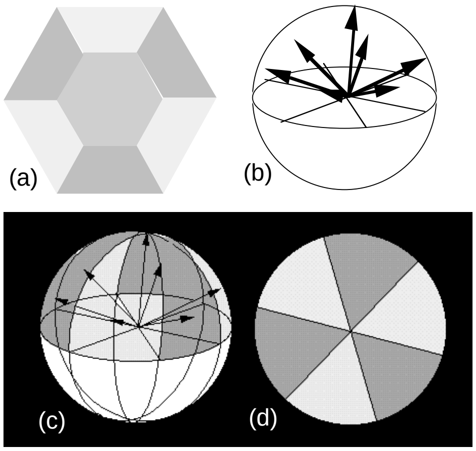

Figure 16. The system of constraints set up by the surface pairs of figure (a) do not admit a solution. No region of the Gaussian sphere satisfies all the constraints. Each of the three dark sectors of the Gaussian sphere satisfies only four of the six constraints simultaneously.

In other words, while a distant light source can be positioned so as to illuminate the truncated hexagonal pyramid shape to make it look similar to figure 1(a), there is no way that it may be made to look like figure 1(b) by solely manipulating the light direction. This consistency or inconsistency of the 2-D shading pattern in the image, we suggest, determines whether the percept obtained is one of a solid 3-D shape illuminated in a particular manner or simply a 2-D pattern of paint.

4.2.3. 'Largely-solvable' systems of linear inequalities:

According to the ideas developed so far, a system of linear inequalities (that represent constraints on light source directions) is either 'consistent' or 'inconsistent' and the shading pattern in the image is interpreted *completely* in terms of illumination or reflectance variations respectively. There is no middle ground. Intuitively, this binary decision strategy seems overly harsh. In this section, we consider 'largely solvable' sets of linear inequalities. As the name suggests, these are systems where a majority, but not all, of the inequalities are simultaneously solvable. When mapped onto a Gaussian sphere, such a system leads to a situation akin to that depicted schematically in figure 17. If we choose a polygon that satisfies the maximum number of inequalities as the solution set, a natural question to ask is what the inequalities that are not satisfied in the chosen polygon represent perceptually/physically. The answer turns out to be quite interesting - these inequalities represent 'compound edges' - image brightness transitions that are perceived as being caused due simultaneously to changes in illumination and surface reflectance.

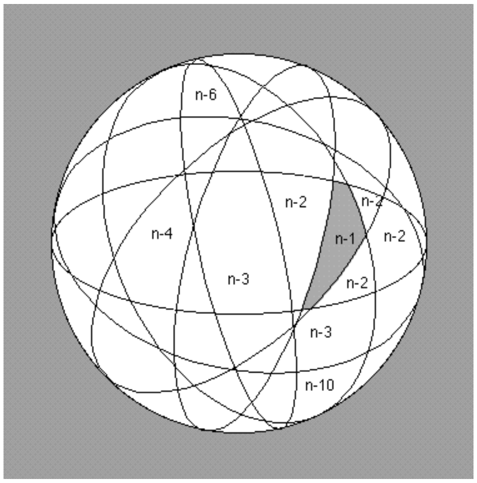

Figure 17. For a 'largely solvable' system of linear homogeneous inequalities, while no finite region simultaneously satisfies all the inequalities, some regions satisfy most. The numbers within the polygons in the figure above represent the number of inequalities satisfied within that region. The shaded region represents the polygon that satisfies the maximum number of inequalities.

Figures 18 and 19 show objects that exhibit compound edges. By setting up the systems of constraints for such objects and identifying which inequalities are not satisfied in the polygon of maximal overlap in the Gaussian sphere, we can pin-point compound edges, as shown in figure 19.

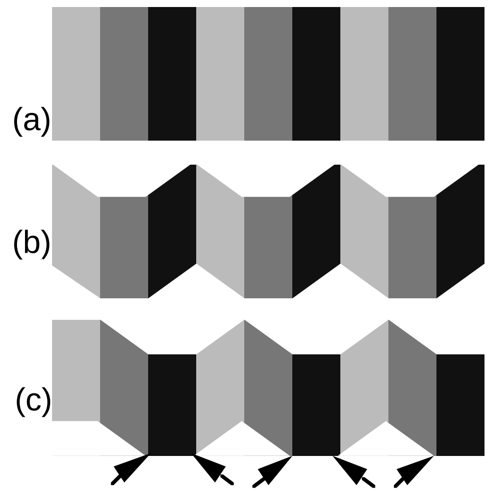

Figure 18. (a) Reflectance edges, (b) Illumination edges, and (c) Illumination and compound edges (indicated by arrows).

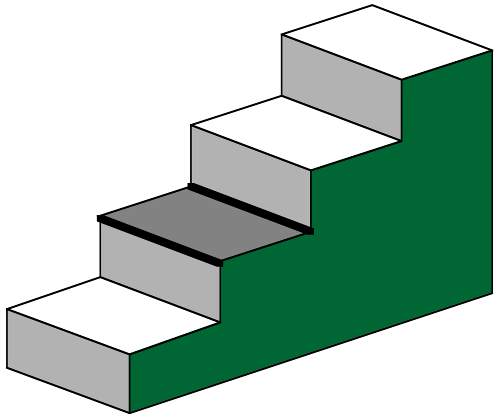

Figure 19. The inequalities not satisfied within the region of maximal overlap on the Gaussian sphere correspond to (and may be used to detect) compound edges in an image.

Why do the unsatisfied inequalities perceptually correspond to compound edges? The answer is quite straightforward. The 3-D shape recovery module suggests that surface orientation changes across all observed edges in the given image (edges that do not represent such orientation transitions are immediately labelled reflectance edges and not considered any further in the shading consistency checking process (see [22] for details)). Moreover, the 'largely solvable' system of linear inequalities suggests that the observed pattern of shading in the image is quite consistent with the 3-D shape. Data that does not fit well into this largely consistent illumination based interpretation (such data is represented by the unsatisfied inequalities which correspond, say, to edge set {e}) is then interpreted in terms of reflectance variations. But, since all the edges already represent orientation changes (due to their geometric structure), the reflectance and illumination interpretations are superimposed for the edges belonging to set {e} and they are perceived as being due to a simultaneous change in surface orientation and reflectance.

4.3 Justifying the shading consistency checking approach:

The strategy for global processing that we have employed above comprises of two parts: (1) deriving the likely 3-D structures using the geometric cues in the 2-D image, and (2) checking whether the 2-D pattern of shading in the original image is consistent with the derived 3-D structure(s). The shape derivation process, in other words, does not make use of the gray level information in the image. At first thought, this may seem to be an odd way to proceed. Would not the shape recovery process be facilitated if we were to provide it with gray-level information ? Not necessarily. Since the image gray-level patterns can be arbitrary (could be due to reflectance *or* illumination variations), they often may not correspond to any 3-D shape. A shape recovery process that relied critically on shading cues would be misled by such patterns. To avoid this possibility, it makes sense to not confound geometric and shading cues in the shape recovery process. The geometric cues, being perceptually more powerful than the shading information for signaling structure [27] may be used in isolation and the consistency of the two sources of information may be verified subsequently.

Besides the need to avoid the possibility of being misled by arbitrary gray-level patterns, a more compelling argument for not relying overmuch on the shading cues is provided by the results of a new psychophysical experiment we designed expressly to determine the relative significance of geometric and shading cues for the task of 3-D shape recovery. The stimuli we used in the experiments were different states of a structure we call the 'Random Height Tesselation'.

4.3.1 Random Height Tesselations:

To be able to determine the relative significance of gray-level and gemetric cues for the task of 3-D shape recovery, we need a stimulus that would allow us to vary one of the cues while keeping the other constant. Furthermore, we would not want our results to be 'contaminated' by higher-level cognitive processes i.e., the stimuli should be such as to avoid the possibility of the use of high-level object specific knowledge in their interpretation. Both these requirements are met elegantly by the use of random height tesselations.

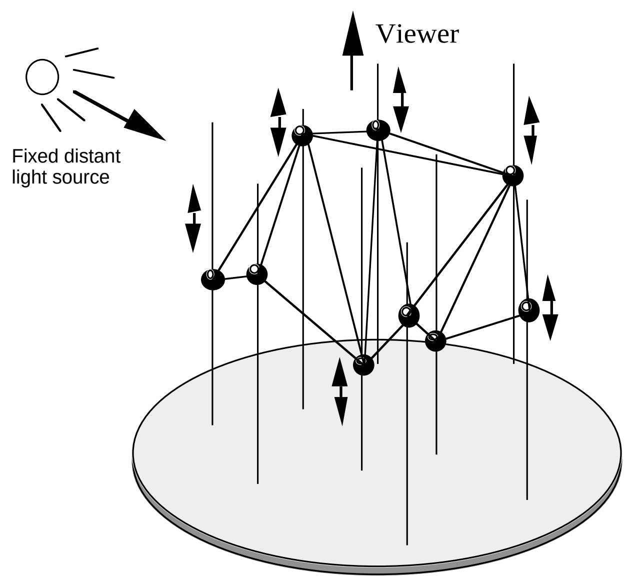

Figure 20. The physical setup underlying a 'random height tesselation' display. The beads are free to move along the wires that are aligned with the line of sight. The changing surface orientations of the triangular facets leads to changes in image irradiance while leaving the geometric structure of the display unaffected.

The physical setup of a random height tesselation is shown in figure 20. It is inspired by the beads on wires idea described previously, only now the x-y positions of the wires are chosen randomly. The bead connectivities are such that sets of three neighboring beads are grouped together to define planes. The z-values of the beads are arbitrary and can be varied at will without changing the geometric structure of the image as observed by an observer positioned vertically on top of the whole setup (the 'wires' are aligned with the line of sight). As the beads move (independently and in an arbitrary fashion) along the wires, the orientations of the triangular planar facets defined by groups of three beads change. Assuming a fixed light direction and a particular reflectance function for the planar facets (say, Lambertian, as used in our experiments), the changing surface orientation leads to a change in the observed image irradiance in the corresponding image patch. Over the whole image, the triangular facets change brightnesses as the beads move along the wires; the geometric structure of the figure, however, remains completely unaffected. Thus, here we have a technique that allows us to vary the shading information in an image while maintaining the geometric cues constant. The interesting question to ask, obviously, is whether an observer can recover the 3-D configuration of the beads at different times by the varying shading information in the image.

Figure 21 shows two typical stimuli generated using random height tesselations. Observe that the geometric structure of both the stimuli is identical; the gray-levels associated with the different faces are, however, very different due to the different surface orientations. The results of the psychophysical experiment were quite interesting. Recovery of 3-D structure from such static stimuli and even their continuous animations was found to be an extremely difficult perceptual task by all the subjects. The animations were perceived as showing a flat pattern with the surface grays of different regions changing over time; the changes in shades were *not* correlated with changes in 3-D structure of the whole setup.

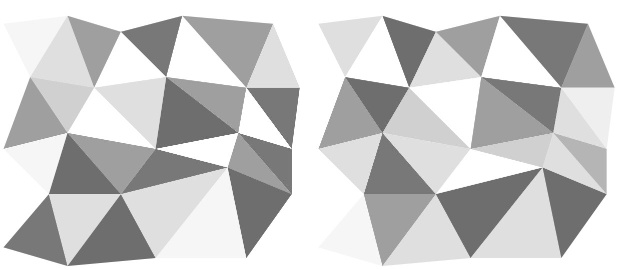

Figure 21. Two typical random height tesselation displays. Notice that the geometric cues in both the displays are identical (the figures reproduced here are not actual experimental stimuli. They are intended simply to provide a qualitative 'feel' for the appearance of the actual patterns).

We are now left with the task of exploring the situation where geometric cues are varied while keeping the shading cues constant. This, as the reader will surely realize, is identical to the conventional 'structure-from-motion' experiments. It has been repeatedly established that 3-D structure can be easily and reliably extracted from the information provided by the transformations in the geometric structure of the image over time.

The above results suggest then that the human visual processes involved in 3-D shape recovery are more critically dependent on the geometric structure of images than on their shading cues. To be fair, we can say this with confidence only for polyhedral objects. For smoothly curved objects, where geometric cues like lines and edges are not too readily available (especially inside the object boundaries), shading information might turn out to be important for structure recovery. While we do not wish to suggest that shading information is completely irrelevant for shape recovery of polyhedral objects, we would like to propose that it is not made use of in a very big way by the human visual system; more importance is attached to the geometric cues.

The results described above provide some justification for our approach of first deriving 3-D shape using geometric cues and then verifying the consistency of the gray-level pattern in the image. From another point of view too, this approach seems valid: in deriving the 3-D structures from only the line drawing, the set of structures S we come up will include all interpretations perceptually consistent with the line drawing. If we had used shading information during shape recovery along with line drawings, we might have come up with a structure set S'. Since S was derived under less restrictive conditions than S', the latter necessarily has to be a subset of the former. Thus, we shall not miss the true solution by not making use of the shading information though we might have to do some extra work. If the cardinalities of S and S' are not too different (as is usually the case), the problem of extra-work is really a non-issue.

5. *CONCLUSION:*

The problem of interpreting images in terms of their shading and reflectance components has traditionally been addressed as a low-level vision task in a highly simplified 2D Mondrian domain [15, 16, 11, 18, 3, 4].Only very recently has it been appreciated that in a 3D world, such conventional approaches are of very limited use; new, more sophisticated solution strategies are required [10, 1, 13]. One such strategy has been proposed by Sinha [22, 25], who has addressed the problem as a mid-level vision task rather than as a purely low-level one. Sinha suggested that a key computation that needs to be performed for interpreting images acquired in a 3D domain is the verification of the consistency of image shading patterns and the likely 3D structure of the scene. This is the problem we addressed in the present paper.

Considerations of robustness, generality and the characteristics of the human perceptual system prompted us to discard available quantitative solution techniques in favor of a qualitative one. The two most important attributes of our technique are its use of qualitative comparisons of gray-levels instead of their precise absolute measurements and also its doing away with the need of an exact pre-specification of the surface reflectance function. We showed that this idea lends itself naturally to a linear-programming solution technique. Results obtained with some sample images were seen to be in conformity with human perception.

The idea of verifying the consistency of shading is ideally suited for use in situations wherein the 3D structure of the scene can be well estimated either by employing heuristics (as was done in [22, 24]) or by the use of other information such as stereo. The notion of verifying the consistency of shading patterns instead of using them to recover 3D structure is rather unconventional. It is, however, a reasonable approach to take when the scene is likely to have a spatially non-uniform reflectance distribution. Some psychophysical results too support this approach.

A limitation of the current implementation of this approach is that it is designed to work solely with images of polyhedral objects. There is, however, no fundamental reason why it cannot be extended to handle smoothly curved objects which can be thought of as finely tesselated polyhedra. In fact, the 'needle-diagram' representation of surface orientation for curved objects uses just such an intuition.

References:

[1] Adelson, E. H. 1990. Perceptual organization and the judgement of brightness. *Invest. Ophthalmol. Vision Sci. Suppl.* 31:265.

[2] Barrow, H. G., and Tenenbaum, J. M., 1981. Interpreting line-drawings as three-dimensional surfaces. *Artificial Intelligence*, Vol. 17, Nos. 1-3, pp. 75-116.

[3] Blake, A. 1985a. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics and Image Processing*, 32: 314-327.

[4] Blake, A. 1985b. On lightness computation in a Mondrian world. *Central and Peripheral Mechanisms of Color Vision*, T. Ottoson and S. Zeki, eds., Macmillan, New York, 45-59.

[5] Cormen, T., Lieserson, C., Rivest, R., 1991. *Fundamentals of Algorithms*, McGraw Hill: New York.

[6] Dines, L. L. 1918-19. Systems of linear inequalities. *Annals of Mathematics*, 2nd series, vol. 20, pp. 191-199.

[7] Duda, R. O. & Hart, P. E. 1973. *Pattern Classification and Scene Analysis.* Wiley: New York.

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

[9] Fourier, J. J.-B. Oeuvres II, Paris, 1890.

[10] Gilchrist, A., Delman, S., and Jacobsen, A. 1983. The Classification and Integration of Edges as Critical to the Perception of Reflectance and Illumination. *Perception and Psychophysics*, 33, 425-436.

[11] Horn, B. K. P., 1974. Determining Lightness from an Image. *CGIP*, 3, 277-299.

[12] Horn, B. K. P. 1977. Understanding image intensities. *Artificial Intelligence*, 8(2) 201-231.

[13] Knill, D. C., & Kersten, D. 1991. Apparent surface curvature affects lightness perception, *Nature*, vol, 351, pp. 228-230.

[14] Kuhn, H. W. 1956. Solvability and consistency for linear equations and inequalities. The American Mathematical Monthly, 63: 217-232.

[15] Land, E. H. , and McCann, J. J., (1971). Lightness and Retinex Theory. *J. Opt. Soc. America*, 61, 1-11.

[16] Land, E. H. 1985. Recent advances in Retinex theory. *Central and Peripheral Mechanisms of Color Vision*, T. Ottoson and S. Zeki, eds., Macmillan, New York, 5-17.

[17] Mackworth, A. K. 1973. Interpreting pictures of polyhedral scenes. *Artificial Intelligence,* Vol. 4, No. 2, pp. 121-137.

[18] Marr, D. 1974. The computation of lightness by the primate retina, *Vision Research*, 14: 1377-1388.

[19] Marill, T. 1991. Emulating the human interpretation of line-drawings as three dimensional objects. *Intl. J. of Comp. Vis.* 6:2 pp. 147-161.

[20] Minsky, M. & Papert, S. 1969. *Perceptrons: An Introduction to Computational Geometry*. MIT Press: Cambridge, MA.

[21] Pentland, A. P. 1982. Finding the illuminant direction. *Journal of the Optical society of America*. 72(4): 448-455.

[22] Sinha, P. 1992. The perception of shading and reflectance. *Masters thesis*, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA.

[23] Sinha, P. & Adelson, E. 1992. The recovery of 3D shapes from 2D line-drawings. Investigative Ophthalmology and Visual Science; ARVO annual meeting abstract issue. Volume 33/4: 666.

[24] Sinha, P. & Adelson, E. 1993a. Recovering 3D shapes from 2D line-drawings. In *Intelligent Robotics; Proceedings of the International Symposium on Intelligent Robotics*, January 7-9, 1993, Bangalore, India; Editor: M. Vidyasagar.

[25] Sinha, P. & Adelson, E. 1993b. Recovering reflectance and illumination in a world of painted polyhedra. *Proceedings of the 1993 International Conference on Computer Vision*, Berlin, Germany.

[26] Yang, Y. & Yuille, A. L. 1990. Sources from shading. *Technical Report TR-90-6*, Harvard Robotics Lab., Harvard University.

[27] Yuille, A. L. 1987. Shape from shading, occlusion and texture, Memo 885, MIT AI Laboratory.