When is the Shape of a Scene Unique Given its Light-Field: A Fundamental Theorem of 3D Vision?

Simon Baker[†], **Terence Sim**[†‡], **and Takeo Kanade** [†]

† The Robotics Institute
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15213

‡ Department of Computer Science
National University of Singapore
3 Science Drive 2, Singapore 117543
Republic of Singapore

Abstract

The complete set of measurements that could ever be used by a passive 3D vision algorithm is the plenoptic function or light-field. We give a concise characterization of when the light-field of a Lambertian scene uniquely determines its shape and, conversely, when the shape is inherently ambiguous. In particular, we show that stereo computed from the light-field is ambiguous if and only if the scene is radiating light of a constant intensity (and color, etc) over an extended region.

Keywords: 3D shape reconstruction, stereo, shape-from-silhouette, the plenoptic function, light-fields, uniqueness.

1 Introduction

Computer vision algorithms operate on measurements of light. The complete set of all such measurements is known as the *plenoptic function* [Adelson and Bergen, 1991] or *light-field* [Levoy and Hanrahan, 1996]; i.e. the radiance of light in free space given as a function of 3D position, 2D direction, wavelength, polarization, and time. Any passive 3D vision algorithm, whether it be a *stereo* algorithm, a *depth-from-defocus* algorithm, or a *shape-from-shading* algorithm, uses (a subset of) the information contained in the light-field. Perhaps the most fundamental question in 3D visual reconstruction, then, is when does the light-field uniquely determine the scene shape?

A closely related question, recently considered by [Kutulakos and Seitz, 2000], is when does the n-camera stereo problem have a unique solution and when is it inherently ambiguous? Kutulakos and Seitz showed that n-camera stereo is not always unique and proposed the concept of the *photo-hull* to quantify any ambiguity. They provided very little insight, however, into when n-camera stereo is actually ambiguous and when the solution is (theoretically, at least) unique.

We analyze stereo computed from the entire light-field. In particular, we present a characterization of when the light-field of a Lambertian scene uniquely determines its shape, and conversely when stereo is inherently ambiguous. Intuitively there are several phenomena that might cause stereo to be ambiguous. One obvious example is the presence of constant intensity regions where any pixel in the constant intensity region “matches” any other. A second example might be the presence of repeated structures in the scene like a picket fence. Stereo might then be ambiguous because a window in one image matches multiple windows in the other image(s).

Our characterization of when stereo is unique is particularly concise and intuitively very natural; stereo is unique (given the light-field) *if and only if* there is no extended region in the scene that is radiating a constant intensity, color, and polarization. Our analysis therefore shows that constant intensity regions are the only inherent ambiguities in stereo; i.e. the only ambiguities that cannot be resolved by using more visual measurements. It also shows that constant intensity regions are *always* ambiguous for stereo; i.e. for any number and arrangement of cameras.

We also show that constant intensity regions are not always ambiguous for *shape-from-*

silhouette. (Part of our analysis consists of formalizing the difference between stereo and shape-from-silhouette.) Our analysis therefore does not completely answer the most general formulation of when the light-field uniquely determines the scene shape. Answering that question is left open for future research. Our analysis also does not not prove anything about the uniqueness of n-camera stereo. The most we can say is that, given *enough* cameras, and in the absence of both constant intensity regions and non-Lambertian reflectance, stereo is *unlikely* to have inherent ambiguities. Complete absence of ambiguity cannot be guaranteed though.

Proving our claims involves addressing a number of technical details. To avoid burdening the reader with all of the details at once, we organize this paper as follows. In Section 2 we introduce the question of whether the shape of the scene is unique given the light-field. We proceed informally to outline the major arguments in the proof of our claims. In Section 3 we formally state our claims as a theorem and discuss its implications. We prove the theorem in Appendix A.

2 Stereo Uniqueness and Ambiguities

2.1 The Plenoptic Function or Light-Field

The *plenoptic function* [Adelson and Bergen, 1991] or *light-field* [Levoy and Hanrahan, 1996] is a function which specifies the radiance of light in free space. It is typically assumed to be a 5D function of position (3D) and orientation (2D). It is also sometimes modeled as a function of wavelength, polarization, and time, depending on the application. We ignore these last 3 effects because: (1) our results can easily be generalized to the case that light can be distinguished based on its wavelength or polarization, and (2) there is an implicit assumption in stereo that the images are captured at the same time, or equivalently that the scene and illumination do not change.

Assuming that there is no absorption, scattering, or emission of light through the air [Nayar and Narasimhan, 1999], the light-field is only a 4D function, a function of direction (2D) defined on a (2D) surface [Gortler *et al.*, 1996, Levoy and Hanrahan, 1996]. (Similarly, the light-field of a 2D scene is 2D rather than 3D, as illustrated in Figure 1.) We make the "no absorp-

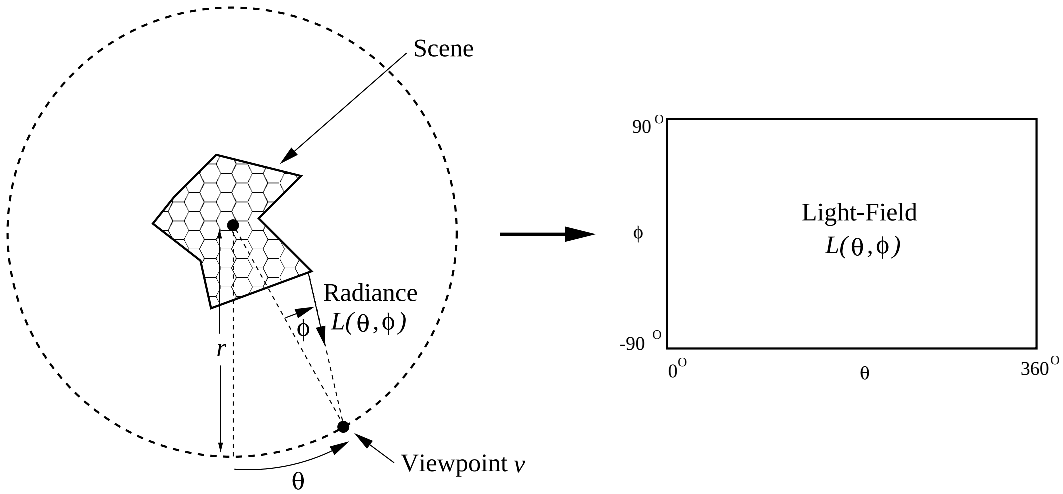


Figure 1: An illustration of the 2D light-field of a 2D scene. The scene is conceptually placed within a circle, radius r. The angle to the viewpoint v around the circle is measured by the angle θ, and the direction that the viewing ray makes with the radius of the circle by ϕ. For each pair of angles θ and ϕ, the radiance of light reaching the viewpoint is denoted $L(\theta, \phi)$, the *light-field.* Although the light-field of a 3D scene is actually 4D, we will continue to use the 2D notation of this figure for ease of explanation. Everything derived here also holds for the 4D light-fields of 3D scenes. The circle can also be replaced with an arbitrary piecewise smooth curve (surface in 3D) in more complex scenes (and need not be connected.)

tion/scattering/emission" assumption. For ease of explanation, we also assume that the surface that the light-field is defined on is a sphere (circle in 2D.) At no point is this spherical surface property required in our analysis. All of our results generalize to the case that the light-field is defined on an arbitrary piecewise smooth surface (consisting of a finite number of connected components.)

Aside: It might be asked at this point why we work with the 4D light-field rather than the 5D plenoptic function. There are 2 answers to this question: (1) The domain of the 5D plenoptic function is free space. The shape of the scene is therefore trivially unique given the 5D plenoptic function. (2) Assuming no absorption or scattering of light, etc, knowing the 4D light-field and the shape of the scene is equivalent to knowing the 5D plenoptic function. In this paper, we assume that we know the 4D light-field and ask whether we can uniquely compute the shape of the scene.

2.2 Assumptions Made About the Scene

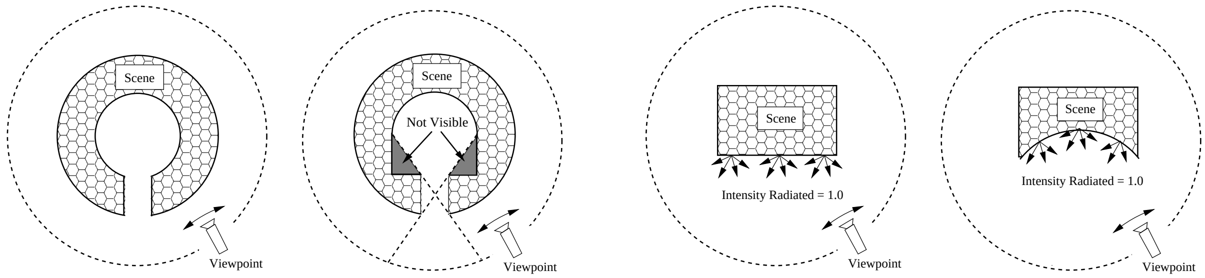
Without making any assumptions about the scene it is impossible to say almost anything. There are three components in the formation of a light-field: (1) the shape of the scene, (2) its reflectance properties, and (3) the illumination conditions. We make piecewise smoothness assumptions about the shape of the scene, piecewise smoothness assumptions about the illumination, and the Lambertian assumption (made by all brightness constancy algorithms) on the reflectance properties. See the formulation of the theorem in Section 3 for more details of these assumptions.

2.3 Uniqueness of Stereo from Lambertian Light-Fields?

Light-fields of Lambertian scenes contain shape information in that any two rays which image the same scene point must necessarily have the same intensity. Is this information enough to constrain the scene shape uniquely? Conversely, of course, just because two points in the light-field have the same intensity does not mean that they necessarily correspond to the same scene point; there may be multiple points radiating the same intensity. This is just the well-known *correspondence problem*, and we are simply asking whether the correspondence problem has a unique solution.

Kutulakos and Seitz studied this question (for n-camera stereo) and showed that *if there is any ambiguity,* there is one special solution (the *photo hull*) which is the union of all of the solutions and therefore contains them all [Kutulakos and Seitz, 2000]. (They also proposed an algorithm to estimate the photo-hull.) But, is there ever any ambiguity? Kutulakos and Seitz did give one example where there is ambiguity, albeit ignoring inter-reflections. (See Figure 3 in [Kutulakos and Seitz, 2000].) Is this generally the case, or is the solution unique most of the time? We answer this question by characterizing when the light-field of a Lambertian scene uniquely determines its shape, and conversely when there are multiple scenes that share the same light-field.

We formulate our characterization in terms of the light radiating outwards from the surface of the scene (i.e. the light-field) rather than in terms of the scene shape, reflectance, and illumination conditions. This leads to a much more concise characterization than would otherwise be



(a) Non-Visible Regions Can Be Ambiguous   (b) Constant Intensity Regions Can Be Ambiguous

Figure 2: Examples of situations in which the shape of the scene is ambiguous even given the light-field: (a) if there are parts of the scene that are not visible in any part of the light-field, the shape of those regions cannot be uniquely determined, and (b) if the light radiated from the scene is constant over an extended region, in this case the entire bottom face of the rectangle on the left and the entire curved bottom of the "carved-out" rectangle on the right, the shape of the two scenes can sometimes be in-distinguishable; i.e. their light-fields can sometimes be set up to be identical (with an appropriate choice for the albedo variation.)

possible. The light-field is also the information that a stereo algorithm has to work with. Our characterization will therefore be more useful to a stereo algorithm that is trying to determine whether there is a unique solution or not, rather than a characterization in terms of the quantities that are unknown to the algorithm (i.e. shape, reflectance, and illumination.)

2.4 The Ambiguous Cases

There are two simple scenarios in which the shape of the scene cannot be uniquely determined from the light-field. The first such case, illustrated in Figure 2(a), occurs when there are points in the scene that are not visible in the light-field. (The shape of the scene is, of course, completely unconstrained anywhere it is not visible.) This trivial case is really just an artifact of our parameterization of the light-field by points on a sphere (or by a circle in 2D). Since the light-field can always be defined on a piecewise smooth surface, possibly disconnected, so that every point in the scene is visible somewhere, this scenario can be ignored.

The second ambiguous case is more significant. It occurs when the intensity of light radiated from the scene is constant over an extended area. See Figure 2(b) for an example of such a scenario. In this case, the rectangle and the modified "carved-out" rectangle with the concavity both have the same *visual hull* [Laurentini, 1994]. If the albedo variation in the concavity is set up in a way that the intensity of light radiated outwards is constant across the concavity (and is the same as that for the rectangle), the two scenes have *exactly the same light-field.* The shape of the scene is therefore ambiguous, even given the entire light-field. (The ambiguous example presented in [Kutulakos and Seitz, 2000] is similar and also has extended constant intensity regions.)

To complete the proof that this is an ambiguous case we must show that it is actually possible to configure the albedo variation to achieve a constant radiance over an extended region. Suppose x is a point in the region that we wish to make radiate a constant intensity. (See Figure 3 for an illustration.) Suppose x has albedo $alb(x)$. The light falling on x can be divided into two components, that inter-reflected from the constant intensity region, and that coming from the rest of the scene (which may be either direct illumination or inter-reflected from some other part of the scene.) Suppose that the radiance of the constant intensity region is $inter$. Denote the foreshortened solid angle (see [Horn, 1986]) subtended by the non-constant "rest" of the scene by $fssa(x)$. The foreshortened solid angle subtended by the constant intensity region is therefore $\pi - fssa(x)$. Finally, denote the foreshorten-weighted average incoming illumination radiance from the rest of the scene $illum(x)$; i.e. the total incoming irradiance from the rest of the scene divided by $fssa(x)$. The total incoming irradiance is therefore:

$$fssa(x) \times illum(x) + [\pi - fssa(x)] \times inter. \tag{1}$$

Multiplying this expression by the albedo gives the amount of light radiated by point x. The constraint to be solved to obtain a constant intensity region is therefore:

$$inter \; = \; alb(x) \left(fssa(x) \times illum(x) + [\pi - fssa(x)] \times inter \right). \tag{2}$$

This equation can always be solved for $alb(x)$. We just have to make sure that the resulting albedo is valid (the surface does not radiate more light than it received); i.e. we just need to check that $alb(x) \leq 1/\pi$ [Horn, 1986]. Substituting this inequality into Equation (2) and simplifying gives:

$$inter \; \leq \; illum(x). \tag{3}$$

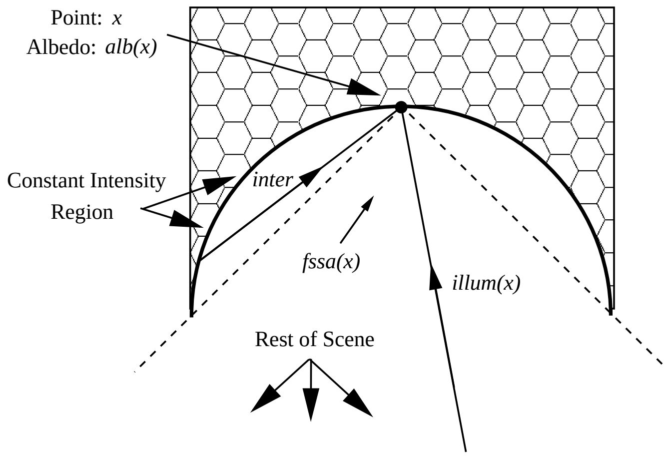


Figure 3: A derivation of the albedo variation required to create a constant radiance region. Consider a point x with albedo $alb(x)$. Suppose that the required constant radiance is $inter$, the foreshortened solid angle of the scene that does not have the constant radiance is $fssa(x)$, and that the foreshorten-weighted average incoming radiance from that part (i.e. the rest) of the scene is $illum(x)$.

Equation (3) means that we can always find an albedo variation to create a constant intensity region so long as the constant radiance desired is less than or equal to the minimum (computed over the constant radiance region) foreshorten-weighted average incoming radiance $illum(x)$ (the total incoming irradiance divided by the foreshortened solid angle subtended by the rest of the scene.) The scenario in Figure 2(b) is therefore realizable in practice. It is a valid ambiguous case where more than one differently shaped scenes share the same light-field.

2.5 Uniqueness of Stereo from Light-Fields

We have just shown that there are cases where two differently shaped scenes share the same light-field. In the concrete example that we exhibited there are extended regions that are radiating a constant intensity. We now argue that if there are no such regions the light-field uniquely determines the shape of the scene; i.e. all ambiguous cases include a constant intensity region.

The main step in the (2D) proof is illustrated in Figure 4. Suppose that θ and ϕ are a pair of angles defining a point in the light-field $L(\theta, \phi)$. If θ changes to $\theta + \mathrm{d}\theta$ and ϕ changes to $\phi + \mathrm{d}\phi$ in a way that the same point in the scene is imaged, the geometry is as sketched in the middle of

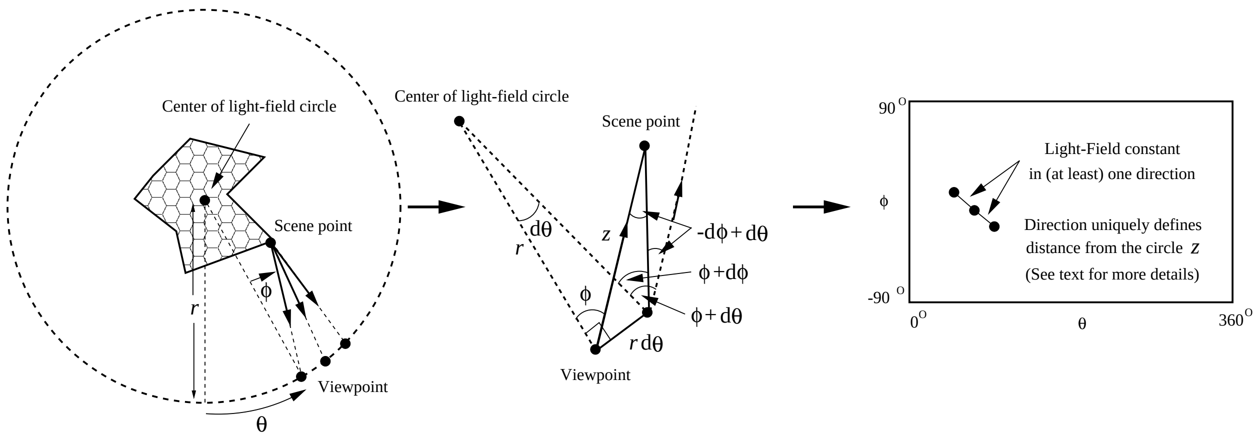


Figure 4: If the 2D light-field is differentiable at (θ, ϕ), it is locally constant in at least one direction (the direction orthogonal to the gradient.) If the gradient is non-zero, the light-field is locally constant in only one direction and that direction *uniquely* defines the distance to the point in the scene (i.e. the depth z.) The depth is therefore uniquely defined wherever the gradient exists and is non-zero. The depth (scene shape) is also globally unique if the light-field is not constant in an extended region (i.e. one containing an open subset) since under this assumption (and appropriate smoothness assumptions about the scene shape), there can only be "isolated points and curves" where $L(\theta, \phi)$ is either not differentiable or has zero gradient. The depth at these points is then uniquely determined because of the assumed piecewise continuity of the scene. This argument can be extended to the 4D light-fields of 3D scenes where there is a 3D hyper-plane in which the light-field is locally constant and which uniquely defines the scene depth. (See Appendix A.1.)

Figure 4. Applying the sine rule to the solid triangle gives:

$$\frac{\sin\left[180^\circ - (-\mathrm{d}\phi + \mathrm{d}\theta) - (90^\circ - \phi)\right]}{z} = \frac{\sin\left[-\mathrm{d}\phi + \mathrm{d}\theta\right]}{r \cdot \mathrm{d}\theta} \tag{4}$$

where z is the distance to the scene point (as measured from the viewpoint circle.) Rearranging this expression gives:

$$\frac{\mathrm{d}\phi}{\mathrm{d}\theta} = 1 - \frac{r}{z} \cdot \cos\phi. \tag{5}$$

This equation means that there is a *one-to-one* relationship between the direction $\frac{\mathrm{d}\phi}{\mathrm{d}\theta}$ in the light-field and the distance z to the point in the scene. The 2D light-field of a Lambertian scene will therefore be locally constant in the direction defined by Equation (5). If the light-field is differentiable and has non-zero gradient, this direction will be the direction orthogonal to the gradient. The gradient of $L(\theta, \phi)$, if it exists and is non-zero, therefore uniquely determines the distance to the scene. The direction of the gradient of the light-field therefore could (theoretically at least) be used as a cue for shape recovery. Note that this cue and the analysis in this section are very closely related to the extraction of depth from *epipolar-plane images* (EPIs) [Bolles *et al.*, 1987].

In general there will be many points at which the light-field $L(\theta, \phi)$ is either not differentiable or has zero gradient. The argument above does not apply at these points. Assuming that the light-field is not constant in an extended region (i.e. one containing an open subset or equivalently a small disk with non-zero radius), the gradient can be zero (or undefined) only on a set of "isolated points and curves." (See Appendix A.1 for more details.) The depth will then also be uniquely defined at these isolated points and curves using the assumed piecewise continuity of the scene. So, assuming that every point in the scene is visible somewhere in the light-field, the light-field uniquely defines the scene shape *so long as there are no extended constant intensity regions.*

2.6 Stereo Versus Shape-From-Silhouette

In Section 2.4 we showed that there are ambiguous cases where the light-field does not uniquely define the shape of the scene. In Section 2.5 we argued that all ambiguous cases contain an extended region that is radiating a constant intensity. Are all cases that contain such a region ambiguous? Perhaps surprisingly, the answer is no. Figure 5 illustrates a scene which contains an extended constant intensity region (the black disk, or sphere in 3D) and which has a unique light-field.

The fact that the light-field of this scene cannot be generated by any other scene can be deduced using a similar argument to that used in *shape-from-silhouette* algorithms [Giblin and Weiss, 1987, Martin and Aggarwal, 1983]. Since the walls are textured (with no black and no constant intensity regions), the argument of the previous section implies that the volume outside the black disk (the region to the right of the dashed ray in the figure) is empty space. We can then deduce that there is a black object in the middle of the room, the *visual hull* [Laurentini, 1994] (defined by an infinite number of cameras) of which is a disk (a sphere in 3D). Since the only shape which has a disk as its visual hull is the disk itself, the object in the middle of the room must be a disk. The shape of the scene is therefore unique. (This argument extends to spheres in 3D. Also note that if the disk is replaced with a rectangle, the situation is exactly the same as in

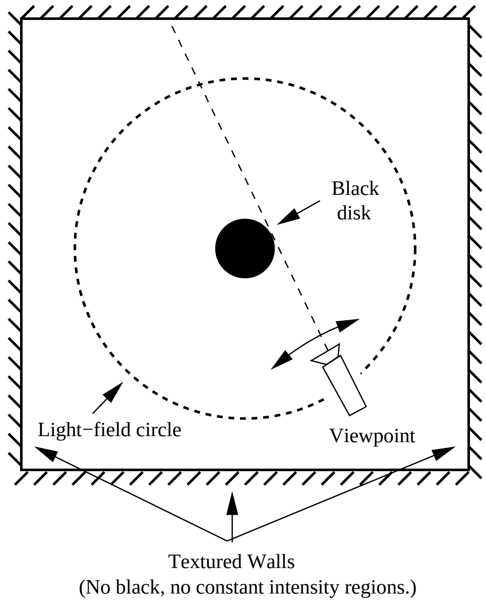


Figure 5: An example of a scene which contains an extended constant intensity region (the black disk in the center of the scene) and yet which has a unique light-field. The uniqueness of the light-field can be deduced using a *shape-from-silhouette* like argument [Martin and Aggarwal, 1983, Giblin and Weiss, 1987]. (See the body of the text for more details.) To distinguish stereo from shape-from silhouette, we exclude the use of any tangent rays (or rays that are infinitesimally close to tangent rays) from stereo. Once such rays are excluded, the shape of the scene is then *always ambiguous* if there is an extended constant intensity region.

Figure 2(b) and the shape becomes ambiguous again.)

The uniqueness of the scene shape in this example is deduced using a shape-from-silhouette like argument. We would like to distinguish stereo algorithms and shape-from-silhouette algorithms. The key distinguishing aspect of shape-from-silhouette is that it uses rays that are tangent to the surfaces of the objects in the scene. Stereo, on the other hand, primarily only "matches" rays that are radiated from the "bodies" of the objects in the scene. Tangent rays are not used.

It turns out that if we exclude tangent rays from the light-field, and thereby remove the source of information used by shape-from-silhouette algorithms, the situation is changed completely. The shape of the scene is then *always ambiguous* if there is an extended constant intensity region. The proof of this fact is constructive and operates by "carving" out a concave hole in the constant intensity region *without changing the visual hull.* The albedo is then modified in the carved region to restore the constant intensity region. The fact that this can always be done was shown in Section 2.4. The details of the proof are included in Appendix A.2.

3 Formalization as a Theorem

So far we have presented our arguments informally to provide the reader with an overview. We now formally state our claims as a theorem. First we need a definition:

Definition: A *Lambertian Scene* is a 3D sub-manifold of $[0, 1]^3$ with boundary. (See [Bröcker and Jänich, 1982] for a definition of these terms.) The boundary consists of a finite collection of continuously differentiable surface patches, each a connected 2D manifold with (possibly overlapping) 1D boundaries. Each surface patch is a perfect Lambertian reflector and has continuously differentiable albedo variation. Each surface patch is also an area light-source and radiates (finite valued, possibly zero) light in an isotropic manner, continuously differentiably across the patch.

We assume that the scene is contained in $[0, 1]^3$ to avoid certain technical problems with rays of light that go off into infinity and do not intersect the scene. Each boundary surface patch is assumed to have continuously differentiable albedo and radiated light. Nothing is gained by generalizing this assumption to *piecewise* continuously differentiable albedo and radiated light because the surface patches can always be sub-divided giving a larger number of smaller ones where the albedo and radiated light are continuously differentiable. Ideal "point" light-sources are also excluded. There are, of course, no such things in reality. Moreover, it is always possible to approximate a point light-source arbitrarily well with a very small source. We then have:

Theorem: *Suppose that S_1 is a Lambertian Scene and L_1 a light-field of that scene with the following properties: (1) L_1 is defined on a finite collection of continuously differentiable surface patches in $[0, 1]^3$, (2) L_1 is open or equivalently every ray in L_1 is contained in a 4D open subset of rays in L_1 (i.e. across the 2D surface on which L_1 is defined and the 2D space of directions), (3) every ray in L_1 intersects S_1 somewhere, (4) every point in free space and on the surface of S_1 is visible in L_1 somewhere, and (5) no ray in L_1 is tangent to S_1 or to the surface patch on which the light-field is defined. Then, there is another Lambertian Scene S_2 (i.e. that has a different occupied volume) that also has light-field L_1 if and only if L_1 is constant in a 4D open subset.*

Of the five assumptions, the first states that the light-field is defined on a 2D surface rather than everywhere in 3D space; i.e. we are assuming that there is no absorption of light in free space.

If the light-field were defined everywhere in free-space the problem would become trivial. The domain of the light-field would immediately determine which parts of the scene are free-space and which are occupied. See the aside at the end of Section 2.1 for more discussion of this point. The second assumption states that there is a 2D open set of directions for which the light-field is defined around each ray making the entire light-field 4D. The third assumption outlaws rays that go off to infinity without intersecting the scene. There is no control over such rays and so they must be eliminated from consideration. The fourth assumption is the visibility assumption. Every point on the scene boundary and in free-space must be visible in the light-field. Every point in free space must be visible in the light-field, because if not, is it possible to add new occupied regions to the scene without changing the light-field. The fifth assumption states that no "shape-from-silhouette" information can be used.

3.1 Interpretation of the Theorem

Figure 6 contains a schematic diagram illustrating the results proved in this paper. The *only if* half of the theorem shows that the shape of the scene is unique, so long as the light-field is not constant in an extended region. This result holds whether the shape-from-silhouette tangent ray information is used or not. It does require that the Lambertian assumption holds and that the entire light-field be given. This result does not apply to the n-camera stereo problem. It does, however, indicate that there are unlikely to be ambiguities if every point in the scene is imaged by enough cameras.

The *if* half of the theorem shows that the shape of the scene is *always ambiguous* whenever there is an extended region in the scene that is radiating a constant intensity, *even if the entire light-field (without tangent rays) is given.* This result also holds in the n-camera case, and under weaker assumptions, such as in the absence of the Lambertian assumption. Constant intensity regions are therefore *inherent ambiguities* in the sense that they cannot be resolved by adding more visual measurements. So, if we encounter a constant intensity region, we can immediately deduce that we will need to use *a priori* assumptions (or silhouette information) to resolve the inherent ambiguity.

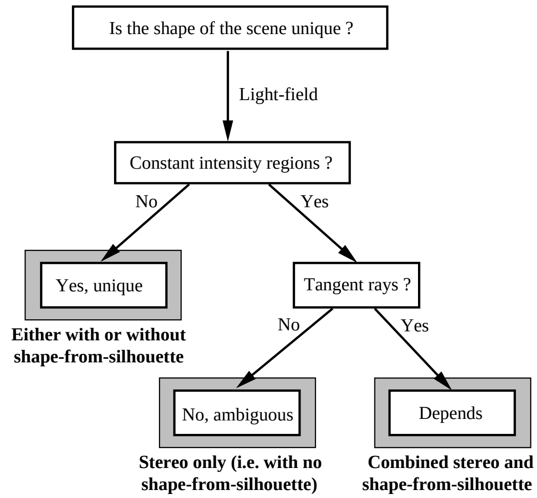


Figure 6: Our characterization of when the light-field of a Lambertian scene uniquely determines its shape. If the light-field contains no extended constant intensity regions, the shape of the scene is *always unique.* If the light-field does contain any such regions, but does not contain any tangent rays, the scene shape is *always ambiguous.* If the light-field also contains tangent rays, the shape may or may not be unique.

3.2 Open Questions

This last result only applies to stereo algorithms that do not take advantage of the shape-from-silhouette information provided by tangent rays. When this information is added, the shape of the scene may or may not be unique. Providing a concise characterization of this case is probably difficult to do (in terms of the light-field at least) and is left as an open question. (This question is much harder than characterizing the information in silhouettes [Martin and Aggarwal, 1983, Giblin and Weiss, 1987, Laurentini, 1994] because it is the light-field that is given, not the silhouettes.)

As an example of why this question is probably quite hard consider the scene in Figure 5 again (the light-field of which is unique.) This scene can be turned into an ambiguous case by simply adding a pair of black hemispheres to the scene and placing them against the walls opposite each other. See Figure 7. Although the shape of the two hemispheres that were added can be uniquely determined, the shape of the disc in the center of the scene is now ambiguous. Parts of it can now be cut away, as shown in Figure 7 (right), without the light-field changing. This example shows that whether the scene shape is unique or not depends on global properties of the light-field. Global properties are generally far harder to analyze than the local ones used in this paper.

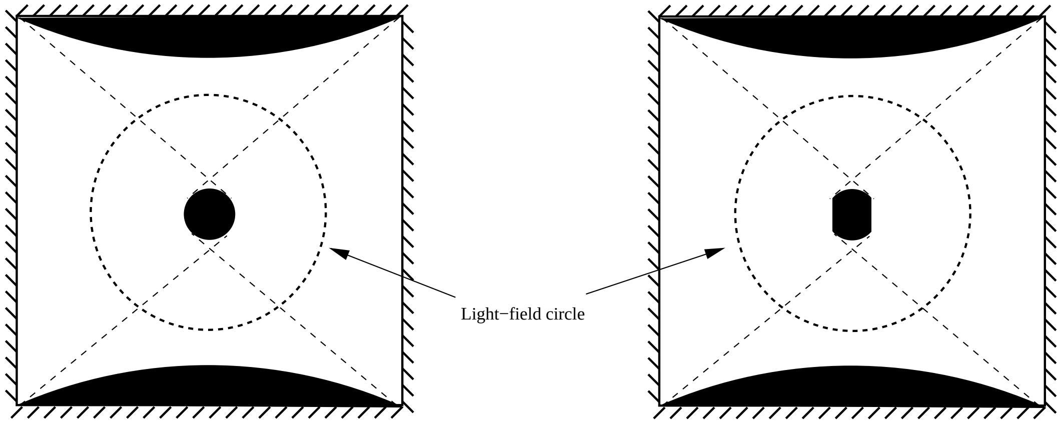


Figure 7: If the unique example presented in Figure 5 is augmented with two black hemispheres, it becomes ambiguous. Although the shape of the hemispheres can be recovered using a shape-from-silhouette argument, the shape of the disc in the center of the scene becomes ambiguous. Parts of it can be cut away, as illustrated on the right, without the light-field changing. This example shows that analyzing the ambiguity of constant intensity regions in the presence of tangent rays depends on global properties of the light-field.

3.3 Summary: A Characterization of Stereo Ambiguities

Stereo ambiguities (in Lambertian scenes) can be categorized into two types: (1) those caused by constant intensity regions and (2) those caused by the fact that only a finite number of cameras are available rather than the entire light-field. One way to interpret our analysis is that ambiguities of the first kind are *inherent*, whereas ambiguities of the second kind will only be a problem in practice if too few cameras are used, or if the Lambertian assumption is not a reasonable one.

3.4 Practical Implications

It goes without saying that this paper is a theoretical paper. This does not mean, however, that there are no practical implications. Some of the implications are as follows:

- The *only if* part of the theorem shows that if an image ever contains a constant intensity region, the shape of the scene must be ambiguous. There is nothing that can be done to resolve this ambiguity. The shape cannot be recovered using visual measurements alone. Prior information must be used instead. This point clearly motivates further study into model based stereo algorithms such as those based on piecewise planarity [Devernay and Faugeras, 1994, Belhumeur, 1996, Faugeras and Keriven, 1998, Baker *et al.*, 1998].

- The *if* part of the theorem shows that the shape of the scene is unique if the entire light-field is known and there are no constant intensity regions. Because of the idealized assumptions and the requirement that the entire light-field is known, the practical implications of this part of the theorem are more limited. This part does suggest however how to arrange a stereo-rig, providing extra insight into previous work [Okutomi and Kanade, 1993, Bhat and Nayar, 1995]. From the proof of Section 2.5 and Appendix A.1.3, it is indicated that in practice three non-colinearly arranged cameras imaging each point should be enough to avoid inherent stereo ambiguities. The reasoning is that in the limit that the locations of these cameras move to the same point, there is enough information to estimate the direction of the gradient of the light-field. Since the depth is unique given this quantity, it is reasonable to assume that if we set up the cameras so that we can approximate this quantity, the scene shape will not be ambiguous. This reasoning pays no attention to noise or to the discrete nature of real measurements, but it does provide some insight into how to set up a stereo rig.

- The pathological case in Section 2.6 provides some insight into the difference between *stereo* and *shape-from-silhouette*. These two 3D reconstruction paradigms use fundamentally different sources of information. Section 2.6 shows the power of silhouette information and Figure 6 clearly illustrates one difference between these two sources of information. This distinction motivates research that tries to combine stereo and shape-from-silhouette in the best possible way, taking maximum advantage of the positive aspects of the two types of information. We have recently proposed an algorithm to do exactly this [Cheung, 2002].

Acknowledgments

An initial version of this paper appeared in the 8th International Conference on Computer Vision [Baker *et al.*, 2001]. We thank Sundar Vedula for feedback on that paper and the anonymous ICCV and PAMI reviewers for their helpful comments and suggestions. The research described in this paper was supported by U.S. Office of Naval Research (ONR) contract N00014-00-1-0915.

References

[Adelson and Bergen, 1991] E.H. Adelson and J. Bergen. The plenoptic function and elements of early vision. In *Computational Models of Visual Processing*. MIT Press, 1991.

[Baker *et al.*, 1998] S. Baker, R. Szeliski, and P. Anandan. A layered approach to stereo reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1998.

[Baker *et al.*, 2001] S. Baker, T. Sim, and T. Kanade. A characterization of inherent stereo ambiguities. In *Proceedings of the 8th International Conference on Computer Vision*, 2001.

[Belhumeur, 1996] P. Belhumeur. A bayesian approach to binocular stereopsis. *International Journal of Computer Vision*, 19(3):237–260, 1996.

[Bhat and Nayar, 1995] D.N. Bhat and S.K. Nayar. Stereo in the presence of specular reflection. In *Proceedings of the 5th International Conference on Computer Vision*, 1995.

[Bolles *et al.*, 1987] R.C. Bolles, H.H. Baker, and D.H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–56, 1987.

[Bröcker and Jänich, 1982] T.H. Bröcker and K. Jänich. *Introduction to differential topology*. Cambridge University Press, 1982.

[Cheung, 2002] K.M. Cheung. Visual hull construction, alignment and refinement across time. Technical Report CMU-RI-TR-02-05, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, January 2002. Ph.D. Thesis Proposal.

[Devernay and Faugeras, 1994] F. Devernay and O.D. Faugeras. Computing differential properties of 3-D shapes from stereoscopic images without 3-D models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1994.

[Faugeras and Keriven, 1998] O. Faugeras and R. Keriven. Variational principles, surface evolution, PDE's, level set methods and the stereo problem. *IEEE Transactions on Image Processing*, 7(3):336–344, 1998.

[Giblin and Weiss, 1987] P. Giblin and R. Weiss. Reconstruction of surfaces from profiles. In *Proceedings of the 1st International Conference on Computer Vision*, 1987.

[Gortler *et al.*, 1996] S.J. Gortler, R. Grzeszczuk, R. Szeliski, and M.F. Cohen. The lumigraph. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH)*, 1996.

[Horn, 1986] B. Horn. *Robot Vision*. McGraw Hill, 1986.

[Kutulakos and Seitz, 2000] K. Kutulakos and S.M. Seitz. A theory of shape by space carving. *International Journal of Computer Vision*, 38(3):197–216, 2000.

[Laurentini, 1994] A. Laurentini. The visual hull concept for silhouette-based image understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(2):150–162, 1994.

[Levoy and Hanrahan, 1996] M. Levoy and M. Hanrahan. Light field rendering. In *Computer Graphics Proceedings, Annual Conference Series (SIGGRAPH)*, 1996.

[Martin and Aggarwal, 1983] W.N. Martin and J.K. Aggarwal. Volumetric description of objects from multiple views. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):150–158, 1983.

[Nayar and Narasimhan, 1999] S.K. Nayar and S.G. Narasimhan. Vision in bad weather. In *Proceedings of the 7th International Conference on Computer Vision*, 1999.

[Okutomi and Kanade, 1993] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, April 1993.

A Proof of the Theorem

We divide the proof into two halves. We first show that stereo is ambiguous *only if* there is at least one extended constant intensity region; i.e. stereo is unique if there are no such regions. Afterwards we consider the *if* half and show that stereo is ambiguous if there are any such regions.

A.1 Stereo is Unique in the Absence of Constant Intensity Regions

The proof that stereo is unique in the absence of constant intensity regions is an extension of Section 2.5 by a continuity argument. We first need one technical result. We show in Section A.1.1 that the light radiated from a Lambertian Scene S_1 is differentiable everywhere except in a 1D subset of its boundary. We then assume that there is a second Lambertian Scene S_2 that is different from S_1 and derive a contradiction, thereby proving the theorem. In Section A.1.2 we show that if there is such a scene S_2 there is a point z in one of S_1 and S_2 (and which is not in the other) for which there is a light-ray in L_1 that passes through z and intersects S_1 at a point that is radiating light: (1) differentiably and (2) that is not locally constant. This is a contradiction because the argument in Section 2.5 shows that no such point can exist; the depth along that light-ray through z is uniquely defined by L_1. To complete the proof we extend Section 2.5 to 3D scenes in Section A.1.3

A.1.1 The Light-Radiated from a Lambertian Scene is Differentiable Almost Everywhere

The argument in Section 2.5 that the distance to the surface of the scene is unique relies on the light-field being differentiable (with non-zero gradient.) We therefore need to show that the light radiated by the scene is differentiable almost everywhere across its bounding surface.

The incoming irradiance at any point on a surface in the scene is a (foreshorten-weighted) integral of the incoming scene radiance over the hemisphere of incoming directions. See Figure 8 for an illustration. Since it is an integral of a real valued function, it is continuous (except at the boundary of the surface patches where the surface normal may be discontinuous.) The light radiated from the surface patches is therefore also a continuous function, again except at the boundaries

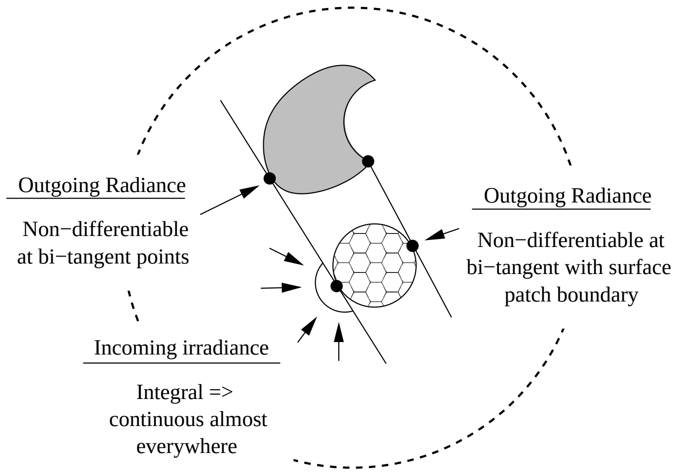
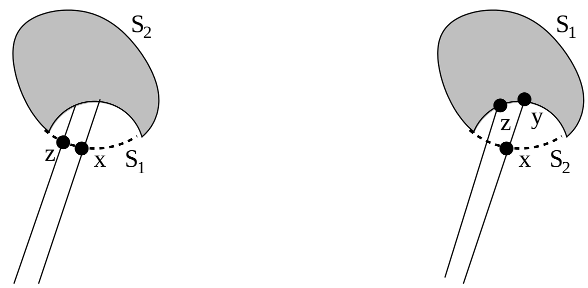


Figure 8: The incoming irradiance is continuous since it is an integral of the incoming light over the hemisphere of incoming directions. The only exception is at the boundaries of the surface patches where the surface normal is discontinuous. The radiated light is therefore also continuous except at the boundaries of the surface patches. The incoming irradiance is therefore differentiable since it is the integral of a piecewise continuous function. The only exceptions are where the visibility changes at bi-tangency points with the surface patches and their boundaries. The radiated light is therefore differentiable everywhere except at these bi-tangency points (and at the surface patch boundaries where the albedo may not even be continuous.)

of the surface patches where the albedo and source radiated light may be discontinuous.

The incoming irradiance is therefore a differentiable function (since it is the integral of a continuous function, the scene radiance) except where the “visibility” of surface patches changes. This only occurs at points of bi-tangency between surface patches, and at points of bi-tangency between surface patches and the 1D boundaries of surface patches. The radiance of light is differentiable at the same points that the incoming irradiance is differentiable. The radiated light is therefore differentiable on the surfaces of the scene, except: (1) at the surface patch boundaries, (2) at points of bi-tangency, and (3) at points of bi-tangency with the surface patch boundaries. All of these types of points are 1D manifolds at most. The first follows by definition. The set of bi-tangency points of a pair of surfaces and the set of bi-tangency points of a surface and a curve are also at most subsets of 1D curves. So, if we consider any open subset of the surface of the scene, there will be a point in that open subset where the radiated light is differentiable. In fact there will be an entire open subset around that point where the radiated light is differentiable.



Case 1 Case 2

Figure 9: Given two different Lambertian scenes S_1 and S_2 there must be a point x that is in one of them and not in the other. There are two cases. Case 1 is where x is in S_1 and not in S_2. In case 2, x is in S_2 and not in S_1. In either case it is possible under the assumptions of the theorem to find a point z that is in S_1 and not in S_2 for which the radiated light is differentiable and not locally constant.

A.1.2 If Two Scenes are Different there must be Non-Constant Differentiable Point

The argument of Section 2.5 only applies to points where the radiated light is: (1) differentiable and (2) not locally constant. We now show that if there is a second scene S_2 that is different from S_1 there is such a point on the surface of S_1 for which the radiated light is differentiable and not locally constant. The argument is essentially a simple continuity argument. The proof of the theorem is then complete except for showing that Section 2.5 can be extended to 3D.

Suppose that S_2 is a second Lambertian scene different from S_1. There is then a point x that lies in one of S_1 and S_2 but not the other. There are two cases. See Figure 9. The first case (Case 1) is when x lies in S_1 but not in S_2. By assumption (4) in the theorem there is then a ray in the light-field that first intersects x and then (possibly) S_2 some non-zero distance afterwards. By assumption (2) there is an open neighborhood of rays around this ray in the light-field that pass through an open neighborhood of x on the surface of S_1. By the argument of the previous section and the assumption that the light-field is not constant in an extended region there has to be a second point z in this open region for which the radiated light is differentiable and not locally constant. Moreover, there has to be a light ray in L_1 that first intersects S_1 at z and then (possibly) intersects S_2 later. If it does intersect S_2 it is at a non-zero distance later. See Figure 9 for an illustration.

The second case (Case 2) occurs when x lies in S_2 but not in S_1. By assumption (4) of the

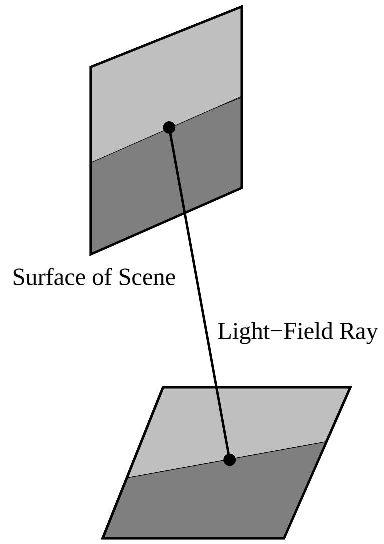


Figure 10: Since we perform a differential analysis we can treat the scene surface and the surface on which the light-field is defined as infinitesimal planar patches. The light radiated from the scene varies linearly across the surface patch. This is projected onto the surface on which the light-field is defined. There is one direction in which the intensity in the light-field will be constant. To extend the argument of Section 2.5 to 3D we just need to make sure we work in some other plane containing the light-field ray.

theorem x is visible in the light-field and by assumption (3) that ray later intersects S_1 at the point y. By a similar continuity argument to Case 1, there is then a open neighborhood of rays about the ray through x and y and an open neighborhood of S_1 which contains another point z where the radiated light is differentiable and not constant. The open neighborhoods can also always be chosen small enough to ensure that one of the rays in L_1 that intersects z first intersects S_2 a non-zero distance before it intersects S_1 at z. See Figure 9 for an illustration.

A.1.3 Extension of Section 2.5 to 3D Scenes

The extension of the argument in Section 2.5 to 3D scenes is straight-forward. Any plane in the 3D world can be used to apply Section 2.5. The only time that this is a problem is when that plane happens to line up with the direction in which the intensity radiated from the surface is constant (the direction orthogonal to the gradient of the radiated intensity on the surface in the world.) Unfortunately we do not know the surface of the scene or the distribution of light radiated from it.

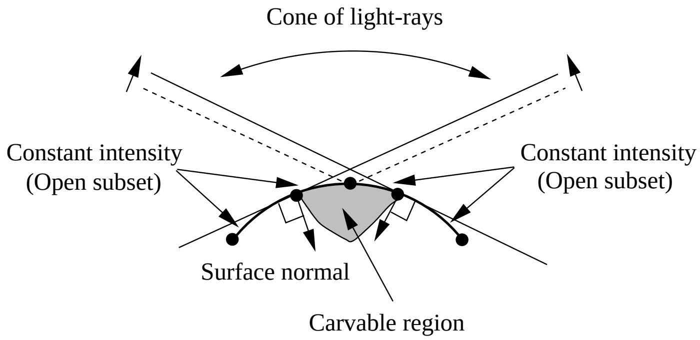


Figure 11: The proof that constant intensity regions are always ambiguous is constructive and operates by finding a point on the surface of S_1 in a constant intensity open region where the surface can be "carved" away and made more concave. To be able to do this without the light-field being changed requires the assumption that there are no tangent rays in the light-field. The incoming rays for an entire open region therefore all lie in a cone (bounded away from the tangent plane). The occupied volume of S_1 can be carved away beneath the open region without the visual hull being changed. As was shown in Section 2.4 the albedo in the carved region can always be set up to reproduce the original constant intensity.

We now show how to avoid this degenerate plane using only the light-field.

Because we are performing a differential analysis we can simply treat both the surface of the scene and the surface on which the light-fields is defined as infinitesimal planar patches. See Figure 10. The variation in radiated light can be regarded as varying linearly across the patch in one direction. There is then one direction in which the radiance variation is constant. This is projected onto the surface on which the light-field is defined. The direction in which the radiated light in the scene is constant and the light-field ray define a plane. The intersection of this plane with the surface on which the light-field is defined has constant intensity variation in the light-field. This degenerate plane is therefore also defined by the light-field ray and the direction in which the light-field is constant (along the surface on which is it defined.) The argument in Section 2.5 can be conducted in any plane that contains the light-field ray except this constant intensity plane.

A.2 Stereo is Ambiguous in the Presence of Constant Intensity Regions

The proof that constant intensity regions are always ambiguous is a combination of the arguments in Sections 2.4 and 2.6. Suppose that there is a constant open subset in the light-field. We can then find a point on the surface of S_1 and an open neighborhood of that point such that every point in the open neighborhood is the end point of a light-field ray in the constant open region of the light-field. We can then find an open subset of the surface and a cone in which all of the incoming light-rays lie, everywhere in that open region. This step requires assumption (5) of the theorem.

The open subset is then further reduced until the boundary of the cone is tangent to the surface at the boundary of the open region. (This reduction is needed to avoid the "visual hull" changing when we carve the region below.) A region is then carved out of the occupied volume of S_1 beneath this open region of the surface and the albedo changed there as in Section 2.4. (Carving means removing the open region from the surface patch and some of the scene volume beneath it and finally adding in an additional surface patch to form the new surface. Note that carving the surface in this way requires the fact that the scene is not infinitely thin. This possibility was excluded when we assumed that the scene is a 3D sub-manifold of $[0, 1]$ in Section 3.) As is illustrated in Figure 11, neither the visual hull nor the light-field is changed in this carving step; i.e. we have modified S_1 to create a different scene S_2 with the same light-field.

There is a minor technical detail to deal with here in that changing the shape of the surface changes $illum(x)$ and may make it smaller than $inter$ and violate Equation (3). This problem can be removed by noting: (1) there is a global scale ambiguity between the illumination and the albedo and so we can always find a new stereo solution with $illum(x)$ multiplied by an arbitrary constant, and (2) it is always possible to change the shape by a finite amount without $illum(x)$ dropping to zero because of the continuity of the light-field. (See Figure 8 and Section A.1 above.)