Recovering LSHGCs and SHGCs from Stereo [1]

Ronald Chung

Department of Mechanical and Automation Engineering
The Chinese University of Hong Kong
Shatin, Hong Kong
E-mail: rchung@mae.cuhk.hk
TEL: (852)2609 8339, FAX: (852)2603 6002

Ramakant Nevatia

Institute for Robotics and Intelligent Systems
University of Southern California
Los Angeles, California, USA
E-mail: nevatia@iris.usc.edu
TEL: (213)740 6428, FAX: (213)740 7877

September 21, 1995

[1] This research was performed at Institute for Robotics and Intelligent Systems, University of Southern California. It was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under Contract No. F49620-90-C-0078. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

Recovering LSHGCs and SHGCs from Stereo

September 21, 1995

Abstract

We examine the problem of computing shape descriptions from stereo, where by shape descriptions we mean 3-D volumetric descriptions of objects rather than a $2\frac{1}{2}$-D depth map of the scene. We argue that intermediate $2\frac{1}{2}$-D depth measurements may not be always directly available from stereo, especially when there are curved surfaces in the scene, and that 3-D volumetric descriptions of objects may have to be derived directly from stereo correspondences. We then present methods to recover volumetric shape from stereo using LSHGCs and SHGCs as the shape models. Our methods are based on some invariant properties of LSHGCs and SHGCs in their monocular and stereo projections. Experimental results on both synthetic and real images of objects with curved surfaces are given. Our technique allows dense surface descriptions to be recovered even for objects without much texture, and it is not restricted to narrow stereo angles or low resolution images. Our technique can also handle objects in close range where perspective distortion in the images can be significant.

1 Introduction

A basic goal of vision is the capability of extracting shape description of objects in a given scene for recognition and manipulation purposes. Monocular methods have to make a number of assumptions owing to the inherent ambiguity in the problem. Stereo vision, in comparison, has more input information and therefore less ambiguity; it has the potential of making fewer assumptions and resulting more robust performance. This paper addresses how shape description of possibly curved objects can be extracted from stereo.

Traditionally, stereo has been thought of as a process that *precedes* high level structural descriptions, and the only goal of stereo is to come up with a depth map like the range data measured from direct range sensors. If necessary, shape information can be derived from such range data. The shape derivation process is hence independent of how the range data is acquired. A good survey for recent literatures on stereo vision can be found in [10].

Here we take a different point of view. We believe that while for some scenes such as random dot stereograms the stereo correspondence process must come before the description process, in general scenes they should work together cooperatively in a way that structural descriptions in each view help reduce ambiguity in establishing correspondences and that stereo correspondences help confirm the structural descriptions. Recovering surface boundaries by observing local depth differences [23] alone will be a problem if the scene is not densely textured; structural descriptions matched across the stereo views also relax that problem.

More importantly, there is an inherent problem in the stereo cue which is not present in direct range sensors. Depth information regarding curved surfaces slowly turning away from the viewer is not directly available from stereo. The reason is, as shown in Figure 1, at different angles of view the apparent surface edges are projected by different contours on the surface. Such boundaries in 3-D are called *limb boundaries* or *contour generators*, and the projected edges are called *limb edges*. It is incorrect to match the limb edges in the two views directly; the error increases with increasing image resolution and stereo angle which are what are desired to have more accurate measurements. On the qualitative side, the description that the edge is a limb edge but not a crease edge can also be important in many applications. For curved surfaces, 3-D information is simply not available unless global shape description is taken into consideration.

Despite these, there have been only a few attempts on recovering shape description directly from stereo. Rao and Nevatia [20, 19] described a technique of deriving volumetric descriptions of conic objects from stereo. The proposed technique, however, makes the assumption that $2\frac{1}{2}$-D depth measurements along the contour generators are available, which is not the case along the limbs of curved surfaces as discussed above.

Lim and Binford [14] were the first to explicitly address the problem caused by limb edges in the process of reconstructing curved surfaces from stereo. Their work is probably the closest to ours and we present it in more details here. They did not indicate how limb edges can be identified, but they proposed a curved surface reconstruction method using stereo. The object in the scene is assumed to be composed of a number of parallel cross-sections, each of them can be described by a conic function which involves five parameters. The curved object is first cut into a number of slices such that each slice is in an epipolar plane. Within each epipolar plane, the

four lines of sight can be recovered from the stereo images, and the cross-section is a conic which is tangential to the four lines of sight. This leaves only one free parameter for the conic in each epipolar plane. Two constraints, the extremum constraint and the terminator constraint, were proposed to determine the free parameter. The extremum constraint chooses the most compact shape among all possible conics in each epipolar plane. The terminator constraint chooses the conic which has the same eccentricity as that of the boundary of the terminator surface.

However, there are still problems with the method:

1. Epipolar planes are parallel to each other only when the projection geometry is orthographic. However, the limb problem is significant only when the curved surface is at close range where orthographic approximation is not a good one. This means the epipolar slices, not being parallel to one another, generally do not possess similar characteristics and neither terminator constraint nor extremum constraint can be applied to them as a whole.

2. Epipolar planes are in general not parallel to the terminator surface, as shown in Figure 2. The difference in the orientations strictly depends upon the orientation of the object with respect to the cameras. This implies the epipolar slices have no direct relationship with the terminator surface and thus terminator constraint is generally not applicable. On the other hand, applying extremum constraint will give a stack of conics being most compact only in the orientations of the epipolar slices. The shapes recovered therefore will be different with different angles of view of the object.

In an earlier work [6, 8], we have proposed a stereo system which uses high level structural descriptions for stereo correspondence. Hierarchical descriptions up to the surface level are computed from each image using a perceptual grouping technique based on global properties like co-curvilinearity and symmetry [17, 16], and such descriptions in the two images are used for stereo correspondence. The output of such a system is therefore not merely depth estimates along edges in the scene, but also segmented surfaces. The system is also able to distinguish limb edges from crease edges by observing the behavior of the junctions at the ends of the edges. However, the shape description derived is only *qualitative*, and the focus of this paper is to describe how *quantitative* depth information about possibly curved surfaces can be estimated from stereo.

Since dense $2\frac{1}{2}$-D depth measurements are not always directly available from stereo correspondences, solution to the shape reconstruction problem requires bypassing intermediate depth measurements if necessary and reconstructing shapes directly from stereo correspondences. In addition, as there are generally an infinite number of shapes that can display the same stereo image data, assumptions about the shapes are necessary. We take the approach of looking for regularities or symmetries in the stereo image data which are unlikely to happen by accident and are properties of some primitives of shape, and inferring such shapes from the correspondences directly. Strictly speaking, such an approach is model-based. Yet the models being used are not shapes of specific objects, but shape primitives that are common and that exhibit properties unlikely to happen by accident. We propose to use generalized cylinders (GCs), introduced by Binford [3], as the shape primitives. The motivation in using GCs is three-fold. First, GC is a volumetric description which is rich and global. Second, they are important classes of shape that

Figure 2: Lateral view of the stereo camera geometry. Epipolar planes in general are neither parallel to one another nor parallel to the terminator surface of an object.

can represent many objects [2]. Third, many important properties of GCs in their projected images have been discovered recently that can help in the reconstruction process.

A generalized cylinder (GC) is defined as a volume consisting of an arbitrary planar shape called the *cross-section*, swept along an arbitrary curve in 3-D called the *axis*. In general, the size and even the shape of the cross-section may change along the axis; the rule describing the change along the axis is called the *sweep rule*. It has been argued that this representation may be too general, and specific types of generalized cylinders have been proposed to describe 3-D objects. Examples are the Linear Straight Homogeneous generalized cylinder (LSHGC) in which the shapes of the cross-sections are homogeneous with a straight axis and a linear sweep rule, the Straight Homogeneous generalized cylinder (SHGC) in which the sweep rule may not be linear, and the Constant generalized cylinder (CGC) in which the cross-section is constant along the axis. The generalized cylinder description is rich, and is naturally amenable to a hierarchical representation so that a complex object can be represented as an assembly of simpler objects.

This paper concentrates on how LSHGCs and SHGCs can be reconstructed from stereo images. They are both common classes of shape and large variety of objects can be described by composites of them, as advocated in [4]. LSHGCs, in particular, are fundamental and important as any GC can be viewed as composed of a number of LSHGC sections, with some of the sections perhaps infinitesimally thin. We present an invariant property of LSHGCs in their stereo projections which help both in hypothesizing their presence and reconstructing them from their image contours. Viewing an SHGC as a number of LSHGC sections, we extend the LSHGC recovery method to reconstruct an SHGC from stereo images.

On the other hand, in parallel with the stereo work, there are work that recovers shapes from a single image. Questions are, what are the limitations of such monocular methods, and precisely how much more can our stereo method offer?

Early monocular work was focussed on the analysis of line drawings of polyhedra [12, 9, 15, 13]. There have also been some primitive attempts to handle curved surfaces such as [1, 22, 27, 11]. Ponce *et al.* [18] examined invariant properties of 3-D shapes in their 2-D projections, and showed that the shape reconstruction problem can be simplified in the case where the objects viewed are GCs. Recently, Ulupinar and Nevatia [24, 25] proposed a general technique to recover shapes of Zero-Gaussian-Curvature surfaces (ZGC), SHGC, and CGC, based on the analysis of symmetries in their images. They are among the first to be able to actually recover the $2\frac{1}{2}$-D dense shape description of a curved object from a single line drawing. Xu *et al.* [26] have also addressed the shape recovery problem for right rotational-symmetric SHGCs.

However, these monocular methods have made the following assumptions which limit their applications:

1. Orthographic projection: so that parallel symmetries in 3-D are retained in their 2-D projections, and thus images of the cross-sections of the object can be extrapolated from the image of the cut,

2. Rotational symmetry or minimum eccentricity of the cross-sections: so that orientations of the cross-sections in 3-D can be recovered, and further,

3. Orthogonality between cross-sections and the axis of the object: so that the orientation of the axis can be determined from the orientations of the cross-sections.

Object shapes derived from one view alone can also at best be localized in 3-D to an arbitrary scaling factor.

Our work shows that with an additional view, the recovery problem is much simplified. More importantly, our stereo method is not restricted to the above assumptions.

In the following we describe separately how an LSHGC and an SHGC can be recovered from stereo. An earlier shorter version of the paper has been presented in [7].

2 Recovering LSHGCs

To reconstruct LSHGCs from stereo, we first need to look at a number of properties of LSHGCs in the images. We present the properties and propose a method for the reconstruction. Experimental results then follow.

2.1 Invariant Projective Properties

An LSHGC is a volume defined by sweeping a given cross-section function along a straight line called axis, such that the cross-section is scaled linearly along the axis (see Figure 3). Linking the points on the surface which correspond to the same unscaled arc length s along boundaries of the cross-sections, we have *meridians*. Because of the linearity in the sweeping function, all meridians are straight and intersect at a point which we call *apex* on the axis, at which the scaling is zero.

Lemma 1 *The image contour of an LSHGC is the projection of one of its meridians under orthographic or perspective projection.*

Proof Shafer and Kanade [21] have shown algebraically that the contour generator of an LSHGC is a straight line under orthographic projection. Here we give a more intuitive proof that the contour generator is in fact one of the meridians *regardless* of the projection geometry. Because of the linearity in the sweeping function, the surface normals to the LSHGC surface at points along a meridian are all parallel and perpendicular to the meridian. Say a line of sight from the optical center C of a camera touches the LSHGC surface at point P as shown in Figure 4. Let m_P be the meridian passing through P. Then both CP and m_P are perpendicular to the surface normal at point P and they define the tangent plane π to the surface at point P. Since all surface normals along a meridian are parallel, the tangent plane π is orthogonal to all the surface normals along the meridian m_P, and so is any line on the plane π. As a result any point Q on the meridian is a point on the contour generator since the line of projection CQ lies on the plane π. □

Theorem 1 *Given four image contours of an LSHGC in a stereo pair of images, the points of intersection among their extensions in the two images are projections of the apex of the LSHGC and they fall on corresponding epipolar lines under orthographic or perspective projection (An example illustrating the theorem is shown in Figure 5).*

Figure 4: Contour generator of an LSHGC is a meridian.

Proof Since all meridians intersect at the apex in 3-D, the projection of meridians to any image will also intersect at the projection of the apex to that image. By lemma 1 the image contours in the stereo images are projections of meridians; their extensions therefore intersect at the images of the same point in 3-D, the apex. As a result the points of intersection fall on corresponding epipolar lines. □

2.2 Recovery

Theorem 1 helps both in hypothesizing the presence of an LSHGC and in reconstructing it. We recover the shape of an object as if it is an LSHGC if:

1. both images separately satisfy the monocular property described in lemma 1, i.e., two of the image contours of the surface are linear, and

2. the images together satisfy the stereo property described in theorem 1.

We then first recover the apex of the LSHGC by matching the image apices at the intersection of the image contours, according to theorem 1. Notice that a contour in 3-D on the surface of the LSHGC can be recovered by matching the terminator contours in stereo. The apex and the 3-D contour therefore uniquely define an LSHGC which projects to the four image contours, regardless of how the cylinder is cut at the two ends. If the cuts are important, we can first recover their partial descriptions in 3-D by matching their images, and the cuts are where the partial descriptions intersect with the cylinder in 3-D.

Notice that as the viewpoint changes, the contour generator may change from a limb boundary to a crease boundary if the cross-section function is not smooth everywhere. However, this does not affect our analysis; both edges are projections of meridians and their extensions still go to the image apex.

Since we have not specified any particular cut on the LSHGC, the method works even for cuts being non-planar or non-orthogonal to the axis.

2.3 Experimental Results

Results on the stereo image pair of a cone, as shown in Figure 6, are presented in Figure 7. We extract a hierarchy of structural descriptions from each image using a perceptual grouping technique and match those descriptions in stereo using the stereo system described in [8]. *Edges* are detected from each image using Canny's edge-detector [5], and are linked into edge-contours using eight-neighbor connectivity. Edge-contours are segmented into *curves* at curvature extrema so that every curve is smooth in itself, and curves are grouped into *contours* based on co-curvilinearity. *Symmetries* are then detected from each pair of approximately symmetrical contours, and they form *ribbons* if they have proper end closures at both ends of the symmetries. Closure at the end of a symmetry can be composed of a contour, a set of multiple contours, or the ends of other symmetries. A number of conflicting ribbons are computed and selection of ribbons is done to resolve conflicts.

The output of the hierarchical stereo system is segmented and matched surfaces in stereo. Junctions have also been labelled as either junctions along limb boundaries or junctions along

Figure 5: Stereo correspondence of LSHGC contours.

creases, from which limb edges have also been identified. We then extend the limb edges to see if their points of intersection in the stereo views fall on corresponding epipolar lines. If they do, we hypothesize the involved surface to be the surface of an LSHGC and the points of intersection in the stereo images are the projections of the apex. Using the apex and a cut of the object in 3-D, recovered by matching the image apices and the image terminators respectively, we can recover the volumetric description of the cone. Notice that although surface markings have not been used in the process, their presence can be exploited by checking if the recovered volumetric description is consistent with depth measurements along those surface markings, and if not, the volumetric description can be deformed to fit the depth data.

In Figure 8 we overlay the recovered descriptions on the left image to illustrate the performance of our method. Note also that the perspective distortion in the images are significant; the eccentricities of the projected ellipses change gradually along the axis from one end to the other.

3 Recovering SHGCs

Viewing an SHGC as a number of LSHGC sections, we can also reconstruct it from stereo images. We present orderly how we hypothesize an object visible in a stereo image pair to be an SHGC, how we establish point correspondences across the stereo views which are projected by the same cross-section, and how we reconstruct the SHGC from such correspondences of different cross-sections.

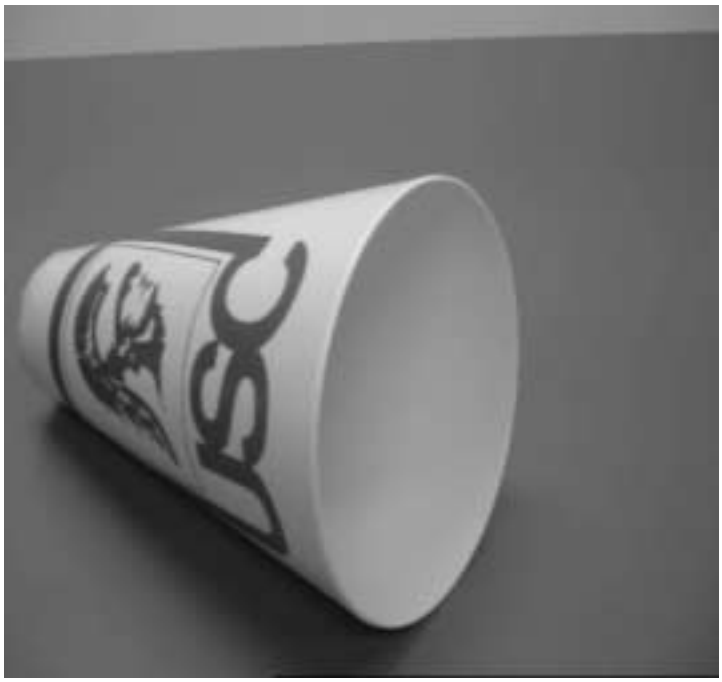
3.1 Hypothesizing Presence of SHGCs

Ponce *et al.* [18] have derived an important theorem regarding the tangents to the image contours of an SHGC under orthographic projection, which was later extended to perspective projection by Ulupinar and Nevatia [25]. The theorem can be stated as this: *If two image contour points of an SHGC are from the same cross-section, the tangents to the contours at these points when extended intersect on the projection of the axis.* Using this property of SHGCs, we can hypothesize the existence of an SHGC by establishing pairwise correspondences between points on the image contours in each image such that their tangents intersect on the same straight line, with the straight line so derived being the projection of its axis.

The pairwise correspondences and the projected axis in each image can be estimated using Hough Transform as in [18], and confirmed by checking whether all corresponding pairs of points follow the same order and are continuous along the contours. A cheaper method is to first hypothesize the axis from two initial known correspondence pairs if they are available, and infer the rest of the correspondences from the hypothesized axis. The initial correspondence pairs can be from junctions at the ends of the image contours, as they are from the same cross-sections which are the terminator surfaces. Zero-curvature points on the image contours are another possibility as shown by Ponce *et al.* [18].

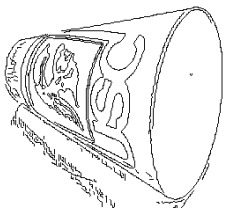
Another evidence is that if the cuts of an SHGC are both along the cross-sections, then their boundaries exhibit *parallel symmetry* in 3-D as defined by Ulupinar and Nevatia [25]: there exists a linear corresponding function between points on the boundaries of the cuts such that

left image

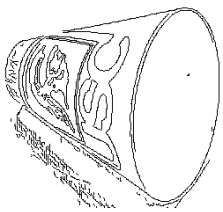


right image

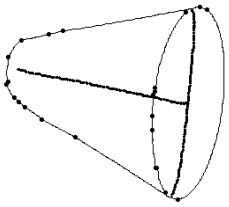
Figure 6: A stereo image pair of a cone.



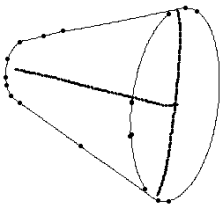
left edges



right edges



matched ribbons (left)



matched ribbons (right)

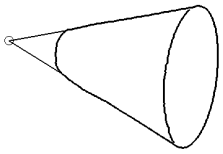


image apex (left)

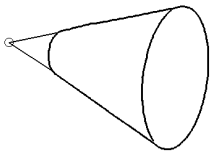
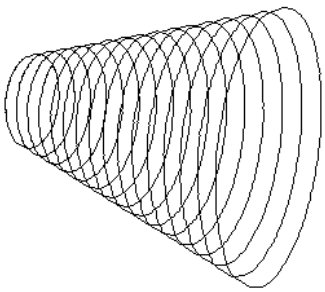


image apex (right)



volumetric description overlaid on the left image

Figure 7: Results for the scene of a cone.

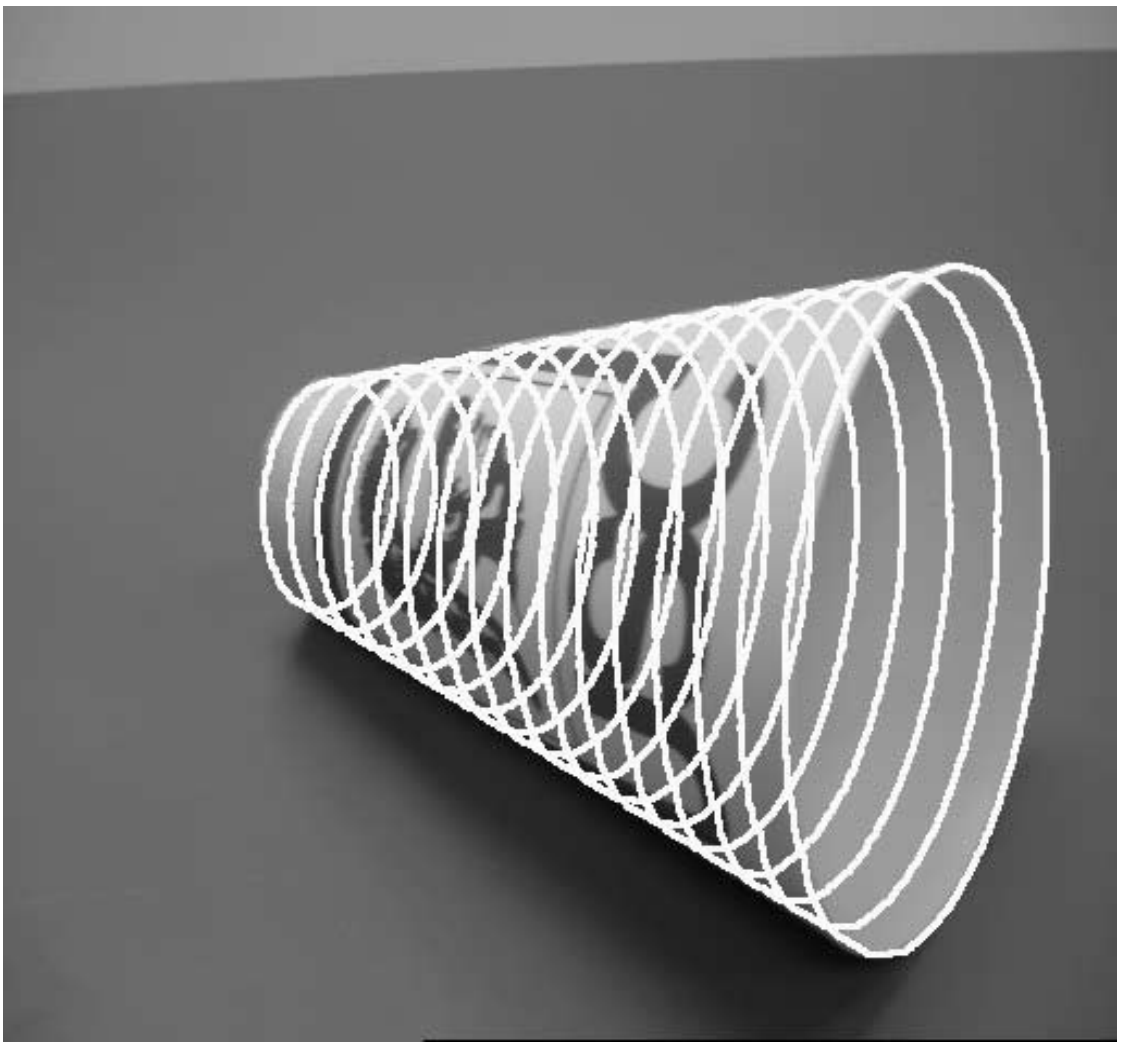


Figure 8: The left image of the cone overlaid by the projection of the recovered volumetric description.

the tangents to the boundaries at the corresponding points are parallel. Since parallel symmetry is preserved under orthographic projection, Ulupinar and Nevatia have used this property to confirm an SHGC in an orthographically projected image. As here we have stereo images of the cuts, we can check the parallel symmetry in 3-D to confirm the existence of an SHGC as well as whether the cuts are along the cross-sections.

3.2 Establishing Correspondences across Stereo Images

The next step is to see if we can construct an SHGC with stereo projections consistent with the scene data. The idea is simple. Matching the projections of the axis in the stereo images will automatically recover the 3-D position of the axis. Similarly, matching the projections of the terminator boundaries will recover the 3-D shape of the cross-section function. What is left is to scale the cross-section function in 3-D so that it touches the lines of projections from the same cross-section of the object.

The problem is that, we have to first set up correspondences across the stereo images so that we know which pair of points in the left image and which pair of points in the right image are projected from the same cross-section. This is not trivial as in case if the contour generators are limbs, the corresponding points in the images are indeed projections from four different points on the surface of the SHGC. Here let us call the corresponding points in the stereo images projected from any given cross-section p_{l1}, p_{l2}, and p_{r1}, p_{r2} respectively.

The correspondence problem is simpler when the contour generators are creases instead of limbs, in which case p_{l1} and p_{r1} are projections of the same point in 3-D and will fall on corresponding epipolar lines, and so will p_{l2} and p_{r2} (see Figure 9). This shows the importance of identifying the nature of the visible surface boundaries as being creases or limbs.

To solve the correspondence problem for the limb edges is more involved. We will first state a few previously proved properties of SHGCs, from which we will derive a new theorem useful for establishing the correspondences across stereo views. Here let us first define the tangent to a surface at a point P *in the direction of* a line L in 3-D to be the tangent which lies on the plane containing the point P and the line L.

The first property we would like to bring up here is:

Lemma 2 (Shafer and Kanade [21]) *Given points on the surface of an SHGC that belong to the same cross-section, the tangents to the surface at these points in the direction of the axis when extended intersect at a common point on the axis (see Figure 10).*

Following Shafer and Kanade, we call the common point of intersection on the axis the *apex*, and the tangents in the direction of the axis the *apex tangents* of the given cross-section. We also call the 2-D projections of the apex and the apex tangents on any image plane the *image apex* and *image apex tangents*. Notice that different cross-sections of an SHGC generally have different sets of apex tangents and different apices on the axis.

Another property is:

Lemma 3 (Ulupinar and Nevatia [25]) *All the tangent lines to a surface at a point, say P, which is on a limb edge of the surface under any given projection geometry, project as the same line on the image plane.*

Figure 10: The apex of a given cross-section of an SHGC.

This property, in combination with lemma 2, imply that tangents to the limb edges at points which belong to the same cross-section, say p_{l1} and p_{l2}, are in fact equivalent to the 2-D projections of the apex tangents at those points. As a result their point of intersection in the image will also be the 2-D projection of the point of intersection of the apex tangents. This has been included in the work of [18] and [25] when they prove that tangents to the limb edges intersect on the projection of the axis. We rephrase it as below:

Lemma 4 (Ponce *et al* [18], Ulupinar and Nevatia [25]) *Given two points on the limb edges of an SHGC that belong to the same cross-section, the point of intersection between the tangents at those points is the image apex of that particular cross-section.*

Combining lemmae 2 and 4, we get the following theorem:

Theorem 2 *If four points on the limb edges of an SHGC in a stereo pair of images belong to the same cross-section, the points of intersections among the tangents at those points in the two images fall on corresponding epipolar lines (An example illustrating the theorem is given in Figure 11).*

Proof By lemma 4 tangents at the image points intersect at the image apex of that cross-section in each image. Since the four points are from the same cross-section, and by lemma 2 apex is uniquely defined for each cross-section, the two image apices in the stereo images are in fact projections of the same point. As a result the two image apices fall on corresponding epipolar lines. □

Theorem 2 allows establishing correspondences among points on the limb edges which belong to the same cross-section. What is remaining is to scale the cross-section function in 3-D to project to the four corresponding points. This is still nontrivial as the four points correspond to four lines of projection which in general are not coplanar. The next section describes a non-iterative method of how we can compute the cross-section in 3-D for each set of correspondences. Remember that the cross-section function in 3-D can be computed by matching the terminator contours in stereo. Similarly, the axis of the SHGC in 3-D can be recovered by matching the image axes.

3.3 Reconstruction as a stack of LSHGC sections

We have obtained for every cross-section of the SHGC two corresponding image apices and four image apex tangents in the stereo images. Here we treat the recovery problem as one to recover a virtual LSHGC whose apex is the apex of the cross-section, whose meridians are the apex tangents of that cross-section, and whose cut is the cross-section itself. Following the idea in section 2, we first derive the LSHGC and then determines the proper cut that is consistent with the scene data.

By matching the image apices we can recover the apex A of the virtual cylinder (see Figure 12). We then move the cross-section function down the axis in 3-D to some distance, say t from the apex. We call the new axis point $C(t)$. Pick one of the four image apex tangents, say the one at point p_{l1}. The image apex tangent and the optical center of the corresponding camera form a plane of projection which we call Π, to which the virtual cylinder should touch.

Figure 12: Recovering a cross-section.

For each point s along the boundary of the cross-section function, we compute two measures: the distance $r(s)$ from $C(t)$ to the boundary point s, and the distance $R(s)$ from $C(t)$ through the point s to the plane Π. The fraction $R(s)/r(s)$ is the scaling of the cross-section such that the point s touches the tangent plane Π. The proper scaling of the cross-section function to touch with the plane Π, regardless of whether the edges are limb edges or crease edges, can therefore be computed as:

$$\text{scale}(t) = \min_s R(s)/r(s)$$

The apex A, the axis, and the scale function scale(t) at distance t along the axis uniquely define a virtual LSHGC which gives rise to the image apex tangents.

The next step is to recover the proper cut of the cylinder to project to the correspondence quadruple. From the above process we can recover the point of contact $P(t)$ between the projection plane Π and the scaled cross-section of the cylinder at distance t from the apex. The line $AP(t)$ then defines the contour generator on the cylinder which projects to the given image apex tangent. Notice that the contour generator of an LSHGC has to be a straight line. Let us say that the proper cut is at a distance t_1 from the apex along the axis, and the proper scaling of the cross-section function on that plane is scale(t_1). The point $P(t_1)$ on the surface of the cylinder which projects to point p_{l1} is then given by the intersection of two lines: the line of projection through point p_{l1}, and the contour generator $AP(t)$. Finally, using property of similar triangles we have:

$$t_1 = (AP(t_1)/AP(t)).t$$

$$\text{scale}(t_1) = (AP(t_1)/AP(t)).\text{scale}(t)$$

The above process of recovering the virtual cylinder and the cut can be applied to any one of the four image apex tangents. In principle they should all return the same cross-section, unless the object is in fact not an SHGC but merely "looks" like an SHGC in monocular views, or the cut being used as the cross-section function is actually not along one of the cross-sections. This serves as an additional test from the reconstruction process to check if the object is indeed an SHGC or not. To summarize, we recover the shape of an object as if it is an SHGC only if:

1. its stereo images separately satisfy the monocular property of an SHGC described in section 3.1, and

2. correspondence quadruples can be established along the image contours that satisfy the stereo property described in theorem 2, and further,

3. the virtual cylinder recovered for each cross-section is independent of the choice of the starting image apex tangent as described above.

Regarding the shape reconstruction process, for each cross-section our system computes t_1 and scale(t_1) from each of the four image apex tangents separately, and uses their averages to recover the cross-section if the values are consistent with one another.

As in the case of LSHGCs, as the viewpoint changes, the contour generator may change from a limb boundary to a crease boundary if the cross-section function is not smooth everywhere. Again, this does not affect our analysis. We can still treat each cross-section as an infinitesimally small LSHGC and the same claim applies: tangents to the contours in the direction of the axis

all go to the apex of that particular cross-section, and tangents to the image contours are projections of such apex tangents in 3-D.

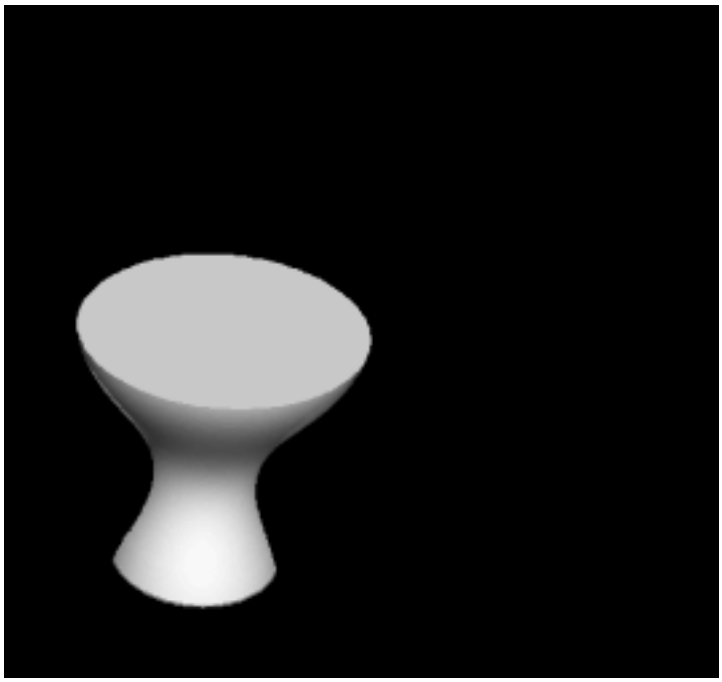
3.4 Experimental Results

Results on a synthetic stereo image pair of an SHGC, as shown in Figure 13, are presented in Figure 14. We extract the hierarchical descriptions from each image, and match those descriptions as in the scene of an LSHGC. We hypothesize the image axes of the SHGC from the tangents at the identified limb-junctions, and determine the correspondence pairs between the image contour points from the hypothesized axis. From the positions where the tangents at the corresponding image contour points intersect at the image axes, correspondences among the image contour points can be set up across the stereo images. From these correspondences we are able to recover the cross-sections of the SHGC in 3-D as shown in Figure 14.

In Figure 16 we show another set of results on a stereo pair of real images of a typical desk lamp (Figure 15). The cameras were configured so that the optical axes were parallel with a baseline of approximately 25cm long. The lamp was about 75cm away from the cameras. Both cameras have a spatial resolution of 512 by 480 with 8 bits of grey scale. Here we show more details about the intermediate steps used in the hierarchical stereo matching system proposed in [8]. *Edges* are detected from each image using Canny's edge-detector [5], and are linked into edge-contours based on eight-neighbor connectivity. Edge-contours are segmented into *curves* at curvature extrema so that every curve is smooth in itself, and curves are grouped into *contours* based on continuity. *Symmetries* are then detected from each pair of approximately symmetrical contours, and they form *ribbons* if they have proper closures at both ends of the symmetries. The closure at the end of a symmetry can be composed of a curve, a set of multiple curves, or the ends of other symmetries. Very small symmetries are ignored to save computation time. Still a large number of symmetries are left and they form many conflicting ribbons. The ribbons then go through a selection process based on a number of constraints among the ribbons. The selected ribbons and the hierarchies of descriptions in the two images are then used for stereo correspondence. Junctions are extracted from the matched ribbons and labeled as limb-junctions or real junctions from their behavior across the stereo images. Limb edges are also identified during this step. The details of such perceptual grouping and stereo matching processes is given in [8].

We then group neighboring ribbons which share smooth boundaries into objects. Notice that the lamp object consists of two neighboring sections of curved surfaces. Using the Hough Transform method mentioned in section 3.1, we are able to derive the SHGC axes of both curved sections and identify that they both share the same axis. We then treat the two curved sections as one single SHGC and derive the volumetric descriptions as in the previous example. In Figure 17 we overlay the recovered descriptions on the left image to illustrate the performance of our method. Notice that the perspective distortion in the images are significant; the eccentricities of the projected ellipses change gradually along the axis from one end to the other.

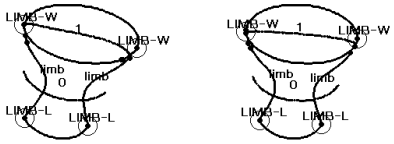
left image



right image

Figure 13: A stereo image pair of an SHGC.

left edges right edges



matched ribbons, junctions (left) matched ribbons, junctions (right)

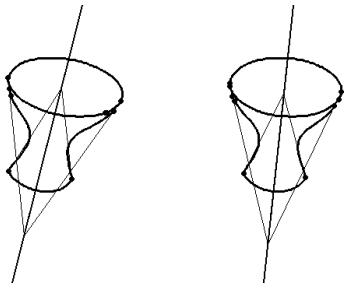
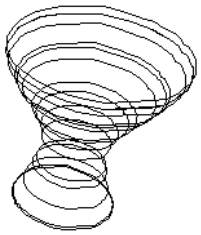


image axis (left) image axis (right)



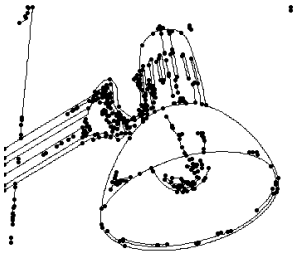
volumetric description projected to the left view

Figure 14: Results for the synthetic scene of an SHGC.

left image

right image

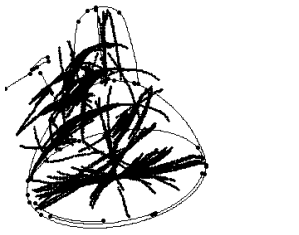
Figure 15: A stereo image pair of a lamp.



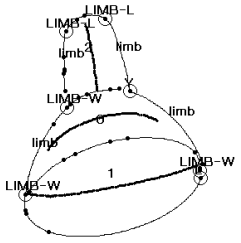
left edges

right edges

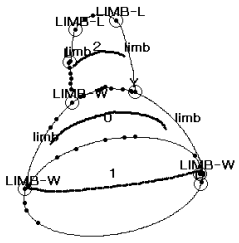
left symmetries



right symmetries



matched ribbons, junctions (left)



matched ribbons, junctions (right)

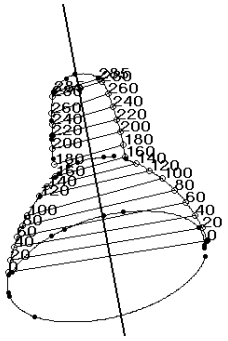


image axis and contour point correspondences (left)

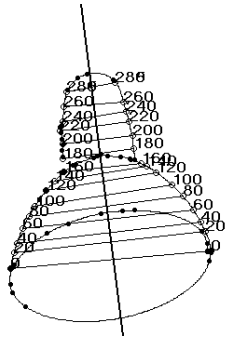


image axis and contour point correspondences (right)

Figure 16: Results of hierarchical stereo matching and volumetric shape recovery for the scene of a lamp.

Figure 17: The left image of the lamp overlaid by the projection of the recovered volumetric description.

4 Conclusion and Future Work

In this paper we have examined the problem of deriving volumetric shape descriptions from stereo images. We emphasize that intermediate $2\frac{1}{2}$-D dense depth measurements may not be always directly available from stereo, which is basically why shape from stereo cannot be treated as merely a sequence of two modules: depth from stereo, and shape from range data. As a result, the volumetric reconstruction may have to be computed directly from stereo correspondences. We have described how volumetric shape can be reconstructed from stereo using some primitives of shapes such as LSHGCs and SHGCs. The methods are based on some invariant properties of the shape models in their 2-D projections. Such properties are not all monocular; we have proposed some properties in stereo which further help confirm and reconstruct LSHGCs and SHGCs from stereo images. In particular, the reconstruction method for SHGCs is based on viewing an SHGC as a stack of LSHGC sections.

Our technique allows dense surface descriptions to be recovered even for objects without much texture, and it is not restricted to narrow stereo angles or low resolution images. Our stereo method also offers a number of advantages over the monocular methods [24, 25, 26]:

1. It allows perspective distortion in the images, i.e., the object can be in close range of the cameras, which in turn allows more accurate shape information of the object to be recovered.

2. It recovers LSHGCs with any cut, and SHGCs even with oblique cross-sections, whereas the monocular methods require planar cuts of an LSHGC and always interpret oblique SHGCs as right SHGCs.

3. It makes no assumption, like rotational symmetry or minimum eccentricity, about the shapes of the cross-sections.

We have shown results for objects with circular cross-sections, but our method is not restricted to them. In addition, we believe the properties of LSHGCs and SHGCS in their stereo images can also themselves be used in a grouping system like the one in [28].

We have not made use of the surface markings if there are any during the shape reconstruction process. In fact they can be used either to confirm the recovered volumetric descriptions, or to deform the descriptions to fit the depth measurements along the markings. LSHGCs and SHGCs are merely primitives of shape, and we are interested in how shapes of composite objects can be recovered. We would also like to explore how the technique can be extended to reconstruct other classes of generalized cylinders.

References

[1] H.G. Barrow and J.M. Tenenbaum. Interpretting line drawings as three dimensional surfaces. *Artificial Intelligence*, 17:75–116, 1981.

[2] I. Biederman. Recognition by components. *Psychological Review*, 94:115–147, 1987.

[3] T. O. Binford. Visual perception by computer. In *IEEE Conference on Systems and Controls*, Miami, Florida, December 1971.

[4] R. A. Brooks. Model-based three-dimensional interpretations of two-dimensional images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 5(2):140–150, 1983.

[5] J. F. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, November 1986.

[6] R. Chung and R. Nevatia. Use of Monocular Groupings and Occlusion Analysis in a Hierarchical Stereo System. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 50–56, Maui, Hawaii, June 1991.

[7] R. Chung and R. Nevatia. Recovering LSHGCs and SHGCs from stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 42–48, Champaign, Illinois, June 1992.

[8] R. Chung and R. Nevatia. Use of Monocular Groupings and Occlusion Analysis in a Hierarchical Stereo System. *Computer Vision, Graphics, and Image Processing: Image Understanding*, 1995. To appear.

[9] M. B. Clowes. On seeing things. *Artificial Intelligence*, 2(1):79–116, 1971.

[10] U. R. Dhond and J. K. Aggarwal. Structure from stereo–A review. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1489–1510, November/December 1989.

[11] R. Horaud and M. Brady. On the Geometric Interpretation of Image Contours. *Artificial Intelligence*, 37:333–353, 1988.

[12] D. Huffman. Impossible objects as nonsense sentences. In B. Meltzer and D. Michie, editors, *Machine Intelligence 6*, pages 295–323. Edinburgh University Press, Edinburgh, 1971.

[13] T. Kanade. Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17:409–460, 1981.

[14] H. S. Lim and T. O. Binford. Curved surface reconstruction using stereo correspondence. In *Proceedings of the DARPA Image Understanding Workshop*, pages 809–819, Cambridge, Massachusetts, April 1988.

[15] A. Mackworth. Interpreting pictures of polyhedral scenes. *Artificial Intelligence*, 4:121–137, 1973.

[16] R. Mohan and R. Nevatia. Segmentation and Description Based on Perceptual Organization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 333–341, San Diego, California, June 1989.

[17] R. Mohan and R. Nevatia. Using Perceptual Organization to Extract 3-D Structures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(11):1121–1139, November 1989.

[18] J. Ponce, D. Chelberg, and W. B. Mann. Invariant properties of Straight Homogeneous Generalized Cylinders and their contours. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(9):951–966, September 1989.

[19] K. Rao. *Shape Description from Sparse and Imperfect Data.* PhD thesis, University of Southern California, December 1988. IRIS Technical Report 250.

[20] K. Rao and R. Nevatia. Computing volume descriptions from sparse 3-D data. *International Journal of Computer Vision*, 2(1):33–50, June 1987.

[21] S. A. Shafer and T. Kanade. The theory of straight homogeneous generalized cylinders. Technical Report CMU-CS-083-105, Carnegie-Mellon University, 1983.

[22] K. A. Stevens. The visual interpretations of surface contours. *Artificial Intelligence*, 17:47–73, 1981.

[23] D. Terzopoulos. Regularization of inverse visual problems involving discontinuitites. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:413–424, 1986.

[24] F. Ulupınar and R. Nevatia. Inferring shape from contour for curved surfaces. In *Proceedings of the International Conference on Pattern Recognition*, volume 1, pages 147–154, Atlantic City, New Jersey, June 1990.

[25] F. Ulupınar and R. Nevatia. Recovering shape from contour for SHGCs and CGCs. In *Proceedings of the DARPA Image Understanding Workshop*, pages 544–556, Pittsburgh, Pennsylvania, September 1990.

[26] G. Xu, H. T. Tanaka, and S. Tsuji. Right Straight Homogeneous Generalized Cylinders with Symmetrical Cross-Sections: Recovery of Pose and Shape from Image Contours. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 692–694, Champaign, Illinois, June 1992.

[27] G. Xu and S. Tsuji. Inferring surfaces from boundaries. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 716–720, 1987. London.

[28] M. Zerroug and R. Nevatia. Scene segmentation and volumetric descriptions of SHGCs from a single intensity image. In *Proceedings of the DARPA Image Understanding Workshop*, Washington,D.C., April 1993.