A Knowledge Based Approach to Shape From Shading

Nick Barnes and Zhi-Qiang Liu
Computer Vision and Machine Intelligence Lab, Department of Computer Science,
The University of Melbourne, Parkville, Victoria, 3052, AUSTRALIA

Abstract — **In this paper, we study the problem of recovering the approximate three-dimensional shape of an object when knowledge about the object is available. The application of knowledge-based methods to image processing tasks will help overcome problems which arise from processing images using a pixel-based approach. We show that by applying domain specific knowledge, we can fully automate the derivation of the approximate shape of an object. Further, this approach can yield specific advantages over existing approaches, both in terms of computation and processing results. This is a powerful paradigm that will have applications in object recognition, robotic navigation, and domain specific scene understanding.**

I Introduction

The classical computer vision approach [1] attempts to build a model of a three-dimensional scene from a single image, without any prior knowledge of what appears in the scene. However, the importance of applying domain knowledge to artificial intelligence (AI) programs has been a fundamental observation arising from research in AI [2]. This observation has not yet been extensively applied in image processing. Most methods act only on local regions of pixels; however, even the entire image does not carry sufficient contextual information. Image processing methods still regularly fail to extract features which are obvious to the human eye. This paper presents a new and powerful paradigm for image processing based tasks: a knowledge-based approach. We demonstrate this method for the classic shape-from-shading problem, showing that it can give advantages both in terms of what problems can be solved, and of computational time.

A Shape from Shading

It is well known that the shape-from-shading problem is under-determined and difficult to solve. All available methods impose some constraints. Variational methods, require *boundary conditions*, which are points where the surface normal can be determined, in order to find a unique solution, if at all possible. These are typically singular points, or limb boundaries [3, 4]. Local methods make assumptions about local surface geometry, and do not generally produce exact estimates of shape [5, 6]. Linear approximations to the reflectance model allow a closed form solution. However, such approaches have problems when the light source is not oblique to the viewer [7], or with noise [8].

Brooks and Horn [3] derived a method for finding surface shape and light source direction for a smooth, lambertian surface where boundary conditions are set. They use a global scheme based on the image irradiance equation with a regularisation component. The surface is found iteratively by minimising the error in the surface fit.

$$I(x,y) = \int\int_{\Omega}((E - \mathbf{n}\cdot\mathbf{s})^2 + \lambda(\| \mathbf{n}_x \|^2 + \| \mathbf{n}_y \|^2) + \mu(x,y)(\| \mathbf{n} \|^2 - 1))dxdy, \quad (1)$$

where E is the image irradiance, $\mu(x,y)$ is a lagrange multiplier function to impose the constraint that $\mathbf{n}$(x,y) is a unit vector, $\mathbf{s}$ points to the light source, $\mathbf{n}_x$ and $\mathbf{n}_y$ are the partial derivatives of n, and λ rates the relative importance of the regularisation term. Minimising I is solved as a problem of variational calculus.

Malik and Maydan [4] extend this technique, combining shape-from-shading with three-dimensional line drawing interpretation. They find the shape from a single image with piecewise smooth surfaces by propagating shape across discontinuities. They assume that the image is pre-segmented, and that a sparse labelling is available.

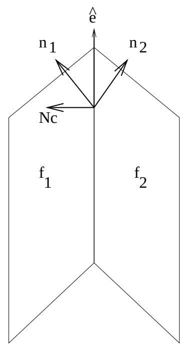


Fig. 1: Tangent vector and surface normals at a non-occluding edge.

Let $\hat{\mathbf{e}}$ be the unit tangent vector to the edge at a point on the roof, and let $\mathbf{n}_1$ and $\mathbf{n}_2$ be the unit surface normals to the tangent planes to the two faces f_1 and f_2 at that point, respectively. Let $\hat{\mathbf{e}}$ be oriented such that when one walks on the edge in the direction of $\hat{\mathbf{e}}$, the face f_1 is to the left, see Figure 1. Now $\hat{\mathbf{e}}$ is perpendicular to both $\mathbf{n}_1$ and $\mathbf{n}_2$, and thus parallel to $\mathbf{n}_1 \times \mathbf{n}_2$, provided that the variation in depth of the edge is small compared to the distance between the surface and the camera. Also, let $\hat{\mathbf{e}}_{proj}$ be the unit vector of the orthographic projection of $\hat{\mathbf{e}}$ onto the image plane. If we let $\mathbf{N}_c$ be a unit vector parallel to the image plane and perpendicular to $\hat{\mathbf{e}}_{proj}$, then:

$$(\mathbf{n}_1 \times \mathbf{n}_2)_{proj} \cdot \mathbf{N}_c = 0 \quad (2)$$

Malik and Maydan form a composite functional which allows shape from shading to propagate across orientation discontinuities. Here the equation is formed for a general reflectance map $\mathbf{R}$ and is taken over the surface s:

$$I(x,y) = \int_s [\lambda_1[(E_1 - R(\mathbf{n_1}))^2 + ((E_2 - R(\mathbf{n_2}))^2] + \lambda_2[(\mathbf{n}_1 \times \mathbf{n}_2) \cdot N_c]^2 + \lambda_3(\| \mathbf{n}_1' \|^2 + \| \mathbf{n}_2' \|^2) + \mu_1(\| \mathbf{n}_1 \|^2 - 1) + \mu_2(\| \mathbf{n}_2 \|^2 - 1)] ds \quad (3)$$

By minimising I as a problem of variational calculus, they derive a scheme to propagate shape across non-occluding boundaries.

In this paper, we apply shape-from-shading to object recognition, particularly for robot visual guidance around man-made objects. Existing variational methods are inadequate as man-made objects may consist entirely of planar faces, and thus images may have no occluding limb boundaries, or singular points. The linear surface reflectance assumption is problematic for robot vision because the light-source may be mounted on the robot, close to the camera, and noise is often a problem with real cameras. We cannot tell, *apriori*, what the local surface shape will be, so local methods which require assumptions about local surface geometry are not appropriate. Finally, traditional three-dimensional reconstruction methods generally implicitly rely on user intervention.

B Knowledge Based Representation of Objects

Generally, knowledge-based approaches emphasise symbolic inferencing for reasoning and decision making [9]. However, much of common-sense knowledge about problems is in the form of procedures or processes for achieving particular goals [10]. Knowledge-based systems that reason about processes may succeed or fail. Success cannot be predicted for systems which interact with an environment, it depends on the environment. Negation-by-failure is often not adequate as failure of a process may be due to lack of evidence. Also, our shape-from-shading approach consists of a number of independent numerical processes, and so requires a somewhat different approach from the use of a simple inference engine [11].

To tackle the shape-from-shading problem, we need to represent knowledge about the visual appearance of the object as well as knowledge about its properties. Also, we need correspondences between the geometric structure of the object and perceived structures in the image. We know the light source direction and some knowledge derived from the image. This must be combined with rules which govern processing, and the numerical processes which collectively calculate shape. In specific domains we may also have knowledge about types of objects being examined.

We employ a knowledge-based architecture based on frame constructs [12]. Note, however, that this is not a traditional application of frames, shape-from-shading operates on numeric rather than symbolic data. Our knowledge here is a mixture of symbolic and numeric data. Frames allow the ability to group facts, rules and processes into associative clusters [13]. This feature is useful for problems in which much of the knowledge is in the form of procedural know-how or numerical processes, and where the facts, rules and processes can be partitioned in useful way.

We wish to use shape-from-shading in conjunction with edge-based matching in a model-based object recognition system. In deriving the surface shape of an object, the system may have access to some prior knowledge from the model. We apply shape from shading after edge matching, so the object is pre-segmented into surfaces along discontinuities and there are no discontinuities within the surfaces. This would often be the case if we use object recognition for robot navigation such as [14].

We assume that our model provides the angles between surfaces along discontinuities, and the reflectance function. We describe a general formulation, but particularly examine the lambertian case, with constant albedo.

II Using Object Model Knowledge for Shape From Shading

We assume objects have piecewise smooth faces, as is the case with many man-made objects. We consider a situation where image edges have been matched to a model. The object model knowledge we apply consists of:

- A wire frame model of the object;
- The direction of faces of the object (what is inside and outside the object);
- The angle between adjoining object faces;
- A hypothesised labelling of faces in the image; and
- The correspondence between the edges in the model and those in the actual image.

III Shape from Shading

Our mathematical approach is based on that of Brooks and Horn [3], and extends the idea of using a segmented image [4]. The method proposed in this paper is able to find a unique solution whenever part of an orientation discontinuity between two adjoining surfaces, and the surfaces near the edge, are visible. This is achieved by deriving additional constraints based on knowledge. Further, this method is able to find a unique solution for the limb and singular point cases without user intervention.

If the angle between two surfaces along an edge is ϕ, then at a pair of pixels along an edge common to the two surfaces, the surface normals will be at an angle of $\theta = \phi - \pi$. We denote the surface normal on the first surface, as $\mathbf{n}_1$, and the surface normal on the second surface, as $\mathbf{n}_2$. Now from vector geometry [15] we have:

$$\mathbf{n}_1 \cdot \mathbf{n}_2 = \| \mathbf{n_1} \| \cdot \| \mathbf{n_2} \| \cdot \cos\theta. \quad (4)$$

For unit vectors:

$$\mathbf{n}_1 \cdot \mathbf{n}_2 = \cos\theta. \quad (5)$$

This provides an additional constraint on the orientation of the surface normals for the pixels along an edge.

We combine this equation with the image irradiance equations for each pixel; and the triple product of the normals and the direction of the edge [4]. We have four equations and four unknowns, and so can find a solution.

We are able to prune most ambiguous solutions because we know if each edge is concave or convex, and that surface normals for all visible points must be in the camera direction. Combining pruning with knowledge-based initialisation generally leads to correct solutions.

The image irradiance equations for the two surfaces are:

$$E_1 - R_1(\mathbf{n}_1) = 0, \tag{6}$$
$$E_2 - R_2(\mathbf{n}_2) = 0, \tag{7}$$

where E_1, E_2 are pixel intensities, and R_1, R_2 are the reflectance maps, of the first and second surfaces respectively.

We assume the edges are smooth, as any discontinuities should be modelled. As smoothing is not required for a unique solution, we smooth the resulting surface normals rather than adding a smoothing component to the variation. We also constrain surface normals to be unit normals.

$$I(x,y) = \int\int_\Omega [(E_1 - R_1(\mathbf{n}_1))^2 + (E_2 - R_2(\mathbf{n}_2))^2 + \lambda_1(\mathbf{n}_1 \cdot \mathbf{n}_2 - \cos\theta)^2 + \lambda_2(\mathbf{n}_1\mathbf{n}_2\mathbf{Nc})^2 + \mu_1(x,y)(\| \mathbf{n}_1 \|) + \mu_2(x,y)(\| \mathbf{n}_2 \|)]dxdy. \tag{8}$$

To solve Equation (8) we form two components, and update $\mathbf{n}_1$ and $\mathbf{n}_1$ in turn. Using a method similar to that in [3], we take the Euler equations for two components, and derive an iterative scheme, see [16] for a full derivation. Assuming a Lambertian reflectance model gives:

$$\begin{aligned}\mathbf{m}_{1_{i,j}}^{(k+1)} = \quad & (E_{1_{i,j}} - \mathbf{n}_{1_{i,j}}^{(k)} \cdot \mathbf{s})\mathbf{s} + (E_{2_{i,j}} - \mathbf{n}_{2_{i,j}}^{(k)} \cdot \mathbf{s})\mathbf{s} \\ & + \lambda_1 \mathbf{n}_{2_{i,j}}^{(k)} (\mathbf{n}_{1_{i,j}}^{(k)} \cdot \mathbf{n}_{2_{i,j}}^{(k)} - \cos(\theta)) + \\ & \lambda_2 (\mathbf{n}_{1_{i,j}}^{(k)} \mathbf{n}_{2_{i,j}}^{(k)} \mathbf{Nc})(\mathbf{n}_{2_{i,j}}^{(k)} \times \mathbf{Nc}). \end{aligned} \tag{9}$$

$$\mathbf{n}_{1_{i,j}}^{(k+1)} = \frac{\mathbf{m}_{1_{i,j}}^{(k+1)}}{\|\mathbf{m}_{1_{i,j}}^{(k+1)}\|}. \tag{10}$$

The equation for $\mathbf{n}_{2_{i,j}}^{(k+1)}$ is similar.

Thus, we have boundary conditions, and so can find a unique solution in the following cases:

1. a discontinuity between two surfaces where the angle between the surfaces is known;
2. an occluding limb; and
3. a singular point.

For occluding limbs, we may set the surface normal by calculating the normal to the tangent of the extracted edge. A singular point has only one possible solution to the irradiance equation. Once we have determined boundary conditions we can propagate the shape across surfaces, using the variational technique of [3].

IV Knowledge-Based Implementation

Our system has three main types of knowledge: facts, rules, and processes. Facts consist of correspondences between the object's geometric structure and perceived structures in the image, properties of the object itself, and facts about the environment. Rules include knowledge of how to apply the facts and processes, rules about proceeding in cases of failure, and standard deductive rules to reach the system goals. Processes are the image processing tasks required to perform shape from shading, and the know-how of how these are performed, e.g. find boundary conditions before solving surfaces.

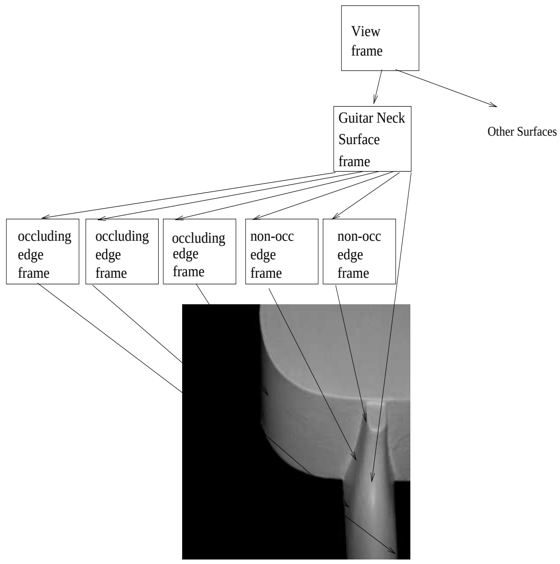


Fig. 2: Instantiation of frames for the guitar used in our final experiment. Five surfaces are visible in this view. The guitar neck has three occluding limb boundaries and two non-occluding edge boundaries visible (other edges are modelled for this surface, but have no shape calculations). The edge frames instantiate the edge pixel locations, and calculate shape at these locations.

A Knowledge / Frame Topology

Object knowledge is structured as perceived, because shape derivation begins from a projected image. Our representation method is called *Canonical views*, see [14]. Briefly, canonical-views model a series of disjoint views of the object; the views are generic in the sense that a single view represents a finite proportion of the object view space. Within a view we represent information about surfaces and edges. By modelling each canonical-view with a frame, the transition of the viewer from one viewpoint to another can be modelled as the transformation between a frame representing each viewpoint.

Our frame system consists of Object, Surface, View, and Edge frames to represent objects. An object frame represents a single object, and has terminals, or subframes for each of the represented views of the object. Each view frame has terminals, or subframes for surfaces and edges. Overlapping view frames may have surfaces in common, and adjoining surfaces have edges in common.

B Fact Knowledge

To recognise an object, it is matched to a view frame. Matching instantiates image features to edges and surfaces in the view frame. If a frame fails to match sufficient slots, it is exchanged for a spatially neighbouring view frame, and matching begins again. View frames also store type descriptions for each edge. Surface frames contain a list of attached edges and their types for the current view, which is passed to them by the view frame. They have information about the reflectance model, which may be the same for all surfaces. They may also hold a description of expected surface shape if the system is employed for model matching. Finally, they contain a normal map for the surface, which is a product of the process. Our processing treats limb edges and orientation discontinuities differently, but a single edge subframe supports both types, as an edge may be of either type in separate views. The frames contain the edge pixel locations, which also delineate the surfaces. For limb edges, the tangent will be calculated at every point along the edge. For orientation discontinuities, the edge frame has a set of pixel locations for each surface, these are typically a few pixels away from the edge to avoid narrow bands of high curvature. Frames for orientation discontinuity also store the angle between the surface normals of the edges.

C Procedural Knowledge

Procedural knowledge consists of: numeric processes that find shape, the order and way in which they are performed, and the process of initialisation. The numerical processes are Brooks and Horn's method for propagating shape across surfaces, our method of calculating the surface normal along orientation discontinuities, and the method of setting the surface normal along limb boundaries to the tangent to the edge. The required order is calculate normals for solvable edges, then solve for surfaces, and finally combine these for the view. Default values for surface terminals can generally be set using surface direction based on estimates of pose extracted during model matching, and the assumption of a flat surface. However, better initialisation can be achieved in specific domains.

V Synthetic Image Experiments

The dodecahedron (Figure (4)) has no curved occluding boundaries or singular points. With a noiseless image the error was 2.9 %. With guassian noise added such that the magnitude of the noise is equal to the magnitude of the signal, the error was 15.9%. The surfaces are still visibly planar (Figure 4 (d)), despite the image being quite distorted (Figure 4 (b)). The surfaces were initialised individually, consistent with the knowledge-based approach. The leftmost surface was initialised to point to the left-hand side of the image, similarly, the rightmost surface was pointing to the right-hand side. The surface facing the viewer was pointing directly toward the viewer, and the upper surface was pointing straight up. In a typical case for a recognition system, these surfaces would be initialised based on an estimate of object orientation.

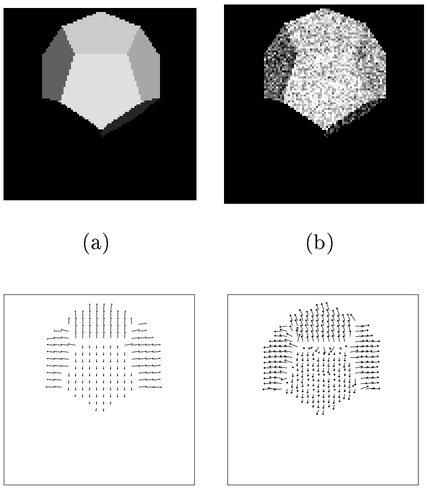
VI Camera Image Experiment

Figure 5(b) shows the surface normals extracted from the guitar image, Figure 5(a). This shows our method can de-

```
Frame Object {
        integer                 NumViews
        Frame View              *Views
}

Frame View {
        integer                 NumSurfaces
        integer                 NumEdges
        Shape_description       *Surface_shape_list
        Frame Surface           *Surfaces

        struct {
               Frame    Edge edge
               integer  Edge_type
        }                       *Edges
}

Frame Surface {
        Surface_reflect_model  R
        /* Shape is for model-based match only */
        Surface_shape_desc      Shape
        Vector                  *Derived_Surf_Norms

        struct {
                Frame   Edge edge
                integer Edge_type
        }                       *Edges
}

Frame Edge {
        integer                 NumPixels
        struct {
                /* Calculated at run time */
                Vector  Unit_tangent
                integer x1
                integer y1
                /* Only used for non-occluding */
                integer x2
                integer y2
        }                       *Pixel_info

        /* non-occluding only */
        Angle_map               Angle_bw_surfaces
        /* occluding only */
        Boolean                 Edge_orientation
}
```

Fig. 3: There are six frame types in our system. Object frames store all the information we require about objects, which consist of most our other frame types: views; surfaces; and edges. Within the view frame all surface and edge frames are instantiated to sets of pixel locations in the image. There is also an environment frame, and an image frame to mediate access to the intensity image. The rulebase assumes as a starting point that an edge matching process has instantiated a view frame to an image.

rive the shape of complex objects, with a combination of planar surfaces, occluding limb boundaries, and orientation discontinuities. However, as the full boundary information for the large surface on the body of the guitar cannot be uniquely inferred, it has not fully converged to the correct solution. With planar surfaces, initialisation based on known edges ensures this is not a problem.



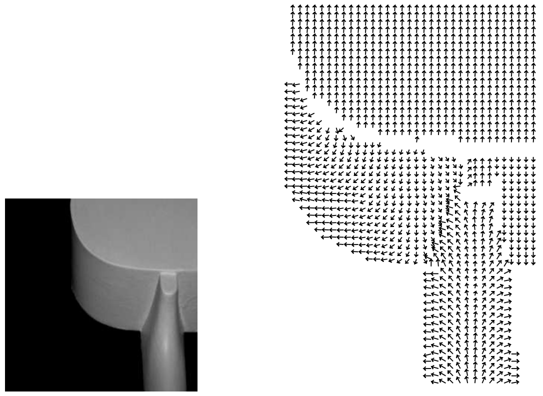
(a) (b)

(c) (d)

Fig. 4: Dodecahedron image. (a) Clean image, (b) Image with gaussian noise added. (c) Calculated surface normals for the clean image, (d) Calculated surface normals for the noisy image.

VII Conclusion

We presented a novel approach to the shape from shading problem, which applies knowledge of an object to aid in inferring shape. By using a knowledge-based approach we are able to derive the shape for complex objects which are only piecewise smooth, and may have no limb boundaries or singular points. We derived a method for uniquely finding the shape along orientation discontinuities within the image, where the angle of the adjoining surfaces is known. This method can be applied to object recognition and can be used without user intervention.



(a) (b)

Fig. 5: Back of a guitar (a) camera image (b) surface normals.

References

[1] D Marr, *Vision : a computational investigation into the human representation and processing of visual information*, W.H. Freeman, New York, 1982.

[2] R Fikes and T Kehler, "The role of frame-based representation in reasoning", *Communications of the ACM*, vol. 28, no. 9, pp. 904–920, 1985.

[3] M J Brooks and B K P Horn, "Shape and source from shading", in *Shape From Shading*, B K P Horn and M J Brooks, Eds., pp. 53–68. The MIT Press: Cambridge Massachusetts, London England, 1989.

[4] J Malik and D Maydan, "Recovering three-dimensional shape from a single image of curved objects", *IEEE Transactions on Pattern Analysis and Machine*, vol. 11, no. 6, pp. 555–566, 1989.

[5] A Pentland, "Local shading analysis", in *Shape From Shading*, B K P Horn and M J Brooks, Eds., pp. 443–487. The MIT Press: Cambridge Massachusetts, London England, 1989.

[6] C H Lee and A Rosenfeld, "Improved methods of estimating shape from shading using the light source coordinate system", in *Shape From Shading*, B K P Horn and M J Brooks, Eds., pp. 323–347. The MIT Press: Cambridge Massachusetts, London England, 1989.

[7] A P Pentland, "Linear shape from shading", *International Journal of Computer Vision*, vol. 4, no. 2, pp. 153–162, 1990.

[8] P-S Tsai and M Shah, "Shape from shading using linear approximation", *Image and Vision Computing*, vol. 12, no. 8, pp. 487–498, Oct 1994.

[9] O N Garcia and Y-T Chien, "Introduction", in *Knowledge-Based Systems: Fundamentals and Tools*, O N Garcia and Y-T Chien, Eds., pp. 1–6. IEEE Computer Society Press, Los Alamitos, California, 1992.

[10] M P Georgeff and A L Lansky, "Procedural knowledge", *Proceedings of the IEEE*, vol. 74, no. 10, pp. 1383–1398, 1986.

[11] R Davis, "Expert systems: Where are we? and where do we go from here?", in *Knowledge-Based Systems: Fundamentals and Tools*, O N Garcia and Y-T Chien, Eds., pp. 7–26. IEEE Computer Society Press, Los Alamitos, California, 1992.

[12] M Minsky, "A framework for representing knowledge", in *The psychology of computer vision*, Patrick H. Winston, Ed., pp. 211–277. McGraw-Hill, New York, 1975.

[13] A J Gonzalez and D D Dankel, *The Engineering of Knowledge-Based Systems: Theory and Practise*, Prentice Hall, 1993.

[14] N M Barnes and Z Q Liu, "Vision guided circumnavigating autonomous robots", *IEEE Trans. Robotics and Automation*, 1997, revised.

[15] D A Holton and J W Lloyd, *Algebra and Gemetry*, Charles Babbage Research Centre:Manitoba, Canada, 1978.

[16] N M Barnes and Z Q Liu, "A study of a knowledge-based approach to 3-d computer vision problems", Tech. Rep. 10, Dept. of Computer Science, Univ. of Melbourne., 1997.