

Boundary Conditions in Multiple Intrinsic Images

Bernhard H. Stuth, Dana H. Ballard, and Christopher M. Brown

Computer Science Department, University of Rochester
Rochester, NY 14627 U.S.A.

Abstract

Intrinsic images are viewer-centered descriptions of physical properties such as color, disparity, optic flow, and surface orientation. Algorithms for computing intrinsic images depend crucially on boundary conditions, but it is not always possible to specify boundary conditions if intrinsic images are computed independently. We explore two observations: (a) if intrinsic images are computed simultaneously, the boundary condition problem is improved; and (b) boundary conditions can be computed from global parameters.

I. Introduction

One important purpose of vision is to build useful descriptions of the environment and intrinsic images are a first step in the description-building process. Intrinsic images are so called because they represent physical (intrinsic) properties of surfaces in a viewer-centered image format.

Over the past few years, enormous progress has been made in specifying the computation of these images by parallel-iterative methods. Such methods have been applied to disparity [Marr and Poggio, 1976; Baker and Blicher, 1982; Grimson, 1981], optic flow [Horn and Schunck, 1981; Ullman, 1979; Barnard and Thompson, 1979], and shape from shading [Ikeuchi and Horn, 1981; Brown et al., 1982]. While such methods constitute an important advance, they only work under most or all of the following conditions: (1) the variation in the image intensity can be entirely explained by the intrinsic parameter being computed; (2) any large discrete variations in the intrinsic parameter are pre-specified; (3) certain free parameters (such as light source direction and reflectance function, or equivalently the reflectance map, in shape-from-shading calculations) must also be known a priori; and (4) the image resolution must be chosen appropriately.

Barrow and Tenenbaum [1978] pointed out that intrinsic images are logically computed simultaneously. They also realized the importance of boundary conditions in defining the intrinsic image computations. However, their edge catalog (page 12 in above reference) is restricted to variations in the illumination of the scene. Subsequent work in shape from shading [Bruss, 1981] and optic flow [Horn and Schunck, 1981] has shown that the estimation of boundaries *within* a given intrinsic image is

difficult. Variations in different intrinsic properties (e.g., color, disparity, optic flow, and surface orientation) interact, and computing several intrinsic images simultaneously is easier than computing them separately. We argue that in most cases, interpolation based on multiple intrinsic images (e.g., color and disparity) provides mechanisms for explicitly acknowledging property discontinuities. This allows for piecewise continuous solutions that are a closer approximation to the physical reality. In many cases the boundary conditions for a given intrinsic image are easily specified from another intrinsic image. For example, the disparity image can specify flow boundaries. We show cases where a *global property* can define boundaries. We conclude with a discussion of mechanisms whereby a multiple intrinsic image computer can detect invalid solutions.

II. When is a Property Intrinsic?

An "intrinsic image" is often defined as a spatially-distributed surface property, *depicted from the perspective of the viewer*. Thus there is some confusion between a truly intrinsic property and a property that depends on view parameters. The albedo of a matte surface is an intrinsic property, but in general the reflectance of a surface is not. The three-dimensional velocity of a surface point is an intrinsic property, but an optic flow velocity is not, since it depends on the view parameters. And of course the disparity map is not, since it depends critically on the viewer.

To pursue the consequences of this distinction further, consider the constraint derived for optic flow by Horn and Schunck [1981]. Given a time-varying image $f(x,y,t)$, one can expand the intensity function $f(x,y,t)$ in a Taylor series, i.e.,

$$f(x+dx, y+dy, t+dt) = f(x,y,t) + f_x \Delta x + f_y \Delta y + f_t \Delta t + \text{higher-order terms} \quad (1.1)$$

where f_x , f_y , and f_t are the appropriate partial derivatives and higher-order terms are ignored. (For an analysis of the validity of the linearity assumption, see [Snyder, 1981].) If indeed the image at some time $t+\Delta t$ is the result of the original image at time t being moved translationally by Δx and Δy , then

$$f(x+\Delta x, y+\Delta y, t+\Delta t) = f(x,y,t) \quad (1.2)$$

Consequently, from (1.1) and (1.2),

$$f_x(\Delta x/\Delta t) + f_y(\Delta y/\Delta t) + f_t = 0 \quad (1.3)$$

The approximation (1.2) does not hold at boundaries between different objects (or at boundaries arising from a self-occluding object), but may be expected to hold for large patches of the image if the variations Δx , Δy , and Δt are appropriately small. Now the partial derivatives are all measurable quantities, and $\Delta x/\Delta t$ and $\Delta y/\Delta t$ are estimates of the optic flow (u, v). Thus

$$\nabla f \cdot \mathbf{u} + f_t = 0 \quad (1.4)$$

where ∇f is the spatial gradient of the image and $\mathbf{u} = (u, v)$ the optic flow.

There are several ways (1.2) could be satisfied. For example, (1.2) holds if $f(x, t)$ is proportional to the matte reflectance of the surface: $f(x, t) \simeq r(x, t)$. However, if the surface had a spatially varying reflectance function, then $f(x, t) \neq r(x, t)$ and this constraint would be invalid. Another way to satisfy (1.2) would be to have a bumpy, shiny surface (e.g., orange skins). In this case the specular reflections would move with surface motions. These observations point out that although one can keep the notion of intrinsic image, the following further distinctions are helpful:

- a *surface property* (e.g., curvature, reflectance) is truly intrinsic. Such a property may or may not satisfy (1.2).
- a *viewer-centered property* (e.g., relative depth, p-q space surface orientation, optic flow).
- a *correspondence property* is one used for matching. By definition, it is any property satisfying (1.2). Viewer-centered properties are usually not correspondence properties.

One would like surface properties and viewer-centered properties to have the correspondence property but this is not guaranteed, and each case must be tested individually.

Sometimes an image intensity function that does not satisfy the correspondence property can be fixed. For example, is there any way to satisfy (1.4) given a surface of varying albedo? One sample way would be to filter out slow variations in the intensity to get back to lightness [Horn, 1975], but this requires knowing the spatial resolution of the lightness variation compared to spatial variations in surface reflectance and orientation. Less delicate methods involve multiple intrinsic images. For example, Stuth [1982] uses edge density for highly-textured surfaces, and Wohn et al. [1982] have applied the $df/dt=0$ using parameters other than intensity. Another proposal is to use the color image $f(x, t) = (r(x, t), g(x, t), b(x, t))$. Transforming this image into an intensity-saturation-hue basis tends to isolate the non-intrinsic achromatic component from the more intrinsic chromatic components. To see that this would work, consider the definition of intensity, saturation, and hue by [Tenenbaum and Weyl, 1975]. Where r , g , and b are primary color measurements,

$$\text{intensity} := r + g + b \quad (1.5)$$

$$\text{saturation} := \frac{1 - 3\min(r, g, b)}{\text{intensity}}$$

$$\text{hue} := \begin{cases} \cos^{-1} \left\{ \frac{\frac{1}{2}((r-g) + (r-b))}{((r-g)^2 + (r-b)(g-b))^{\frac{1}{2}}} \right\} & \text{if } g \geq b \\ 2\pi - \text{hue} & \text{if } g < b \end{cases}$$

If the intensities are subject to variation in surface albedo, then the imaging process will measure $(r/x, g/x, b/x)$ for some x that affects only the intensity measurement. Specular variations are (approximately) manifested as $r+x, g+x, b+x$ for some x and this affects intensity and saturation. Thus hue is a true intrinsic image. Of course, Equation (1.4) must also be satisfied nontrivially for this to be helpful.

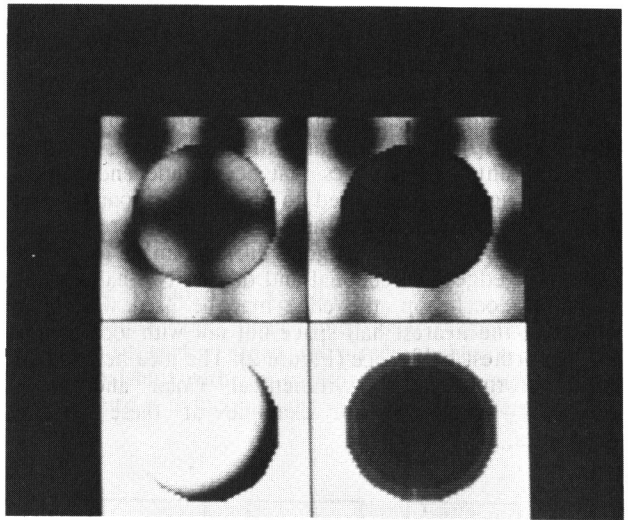


Figure 1. Examples of different properties of a sphere seen against a flat background: (a) mottled surface reflectance which is both an intrinsic property and satisfies the correspondence property; (b) color, which can be used as a correspondence property under certain transformations; (c) surface shading of a matte surface, which is neither a surface property nor a correspondence property; and (d) depth (darker = nearer) which is also neither a surface property nor a correspondence property.

III. Interpolation

In any particular intrinsic image, the points that contain measures of the physical property can be sparse. One such example is the disparity image; the system must interpolate disparity measures between the data points. Such interpolation does not add information to the data, but an explicit representation of the surface is helpful in such tasks as tactile manipulation.

Much work has been done on defining good interpolation methods [Grimson, 1981; Terzopoulos, 1983], but as Baker and Blicher [1982] pointed out, these

methods usually treat all points uniformly. They do not automatically acknowledge the natural, sharp discontinuities that arise in the scene (for example, from occlusion). Thus, while interpolation is desirable, the key problem is to limit the process to an appropriate set of points. Figure 2 shows this problem.

Another problem raised by interpolation across boundaries is that it raises erroneous expectations as to how the scene would look if viewed from a different angle. If the interpolation could be piecewise through appropriate points, then those data would still be valid from different view angles, once transformed (Figure 2). Recent work by Terzopoulos [1983] takes breakpoints into account, and even discovers them automatically within a single image. Multiple intrinsic images may be more helpful in the interpolation process as they can serve to isolate breakpoints in the interpolation grid more simply, and the breakpoints can be based on physical rather than mathematical features. As an example, consider the use of color image together with the depth image.

We assume that a depth edge (a sharp change in the disparity map) in conjunction with a color edge indicates a breakpoint. Depth edges that occur independently of color changes are not used to signify breakpoints. The idea is that discontinuities may be expressed as *conjunctions* of spatially-aligned intrinsic image gradients. The color edge is directional and defines two half spaces. Owing to occlusion, the edge may be used with local points in the nearest half space but not with local points in the furthest half space (Figure 2). The idea here is that edge information is asymmetrical. Color and depth gradients are only one example of these general principles.

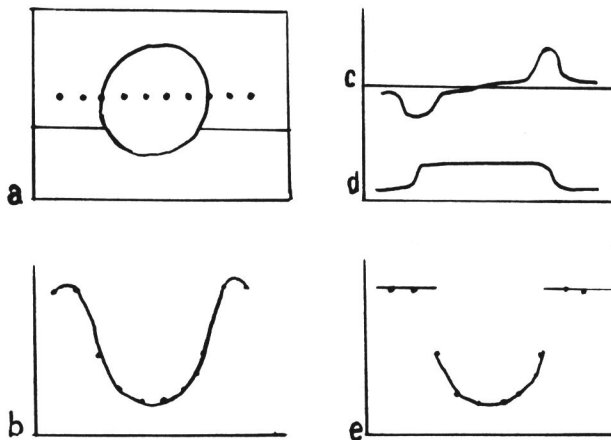


Figure 2. Problems with interpolation: (a) a depth image of the sphere such as Fig. 1d; (b) smooth interpolation in depth through the line of points shown in (a) showing the problems in not recognizing discontinuities; (c) gradients in the depth image; (d) schematic gradients in a registered (not shown) color image; and (e) depth interpolation corrected by using conjunctions of edges in (c) and (d).

IV. Dynamic Boundaries

Consider the optic flow example mentioned earlier. In the case where the sphere does not translate the flow is stationary (does not change from frame to frame), even though the intensities change. This allows the flow to be set to zero in places where $f_t = 0$ and not propagated into regions of zero flow. However, when the sphere translates across the field of view, the image intensities at the boundaries are time-varying, ($f_t \neq 0$) but at the same time the intrinsic property is violated ($df/dt \neq 0$). The constraint equations do not hold at the boundary. This can prove disastrous when the erroneous estimates propagate to neighboring points via relaxation methods.

One solution to this problem uses the time-varying disparity image. Large differences in temporal disparity change are likely to be occlusion boundaries. These large differences can be used to delimit areas in the flow image where the continuity constraint $df/dt = 0$ is unlikely to hold.

Experiments in computing flow [Stuth, 1982] have shown that if the surface reflectance is varying (e.g., the surface is textured), then estimates for flow can be obtained from very local measurements. Consequently, the segmentation based on depth discontinuities can be very conservative (there can be more regions than necessary) without disturbing the flow computations.

V. Boundary Conditions Based on Global Criteria

A well known method for delimiting prominent color image boundaries is to use a global color space. For a survey of this method, see [Shafer and Kanade, 1982]. The basic idea is to construct a histogram in a color space $h(r, g, b)$ where $h(r_o, g_o, b_o)$ is the number of pixels in the image with color (r_o, g_o, b_o) . The histogram may be processed into a set of local maxima $h_1, \dots, h_n$. Then each pixel in the image is assigned a label k where the color denoted by h_k is the maximum nearest the pixel's color. With regard to boundary conditions, there are two key points. First, a global criterion--nearness to prominent color features in color space--is used to define boundaries. Second, the process of delimiting the boundaries may be seen as a two-way transformation: the transformation to color space defines color features and the transformation back to the color image defines boundaries. This process is familiar from histogram-based segmentation schemes (e.g., [Shafer, 1980]).

The central idea of finding boundary conditions via global transformations that is captured by the color example has been extended to several other intrinsic images [Ballard, 1983]. These applications involve several technical refinements of the color technique. The two most important are: (1) instead of a single global feature space there may be several that must be computed sequentially; and (2) the boundary must be computed sufficiently fast to help the intrinsic image relaxation computations. Our intent here is to suppress these details but at the same time describe the basic interactions that allow boundary conditions to be computed.

A. Example 1: Shape from Shading

Shape from shading has been studied extensively by Horn [1975] and a relaxation algorithm for surface orientation has been described by Ikeuchi and Horn [1981]. In this algorithm, boundary conditions and reflectance map were specified. Both of these conditions can be relaxed if additional information is specified (Fig. 3). First, oriented intensity discontinuities (edges) constrain normals at boundaries of corresponding surfaces. However, there may be many edges that are not on the boundary of surfaces. A second constraint uses a global feature space of possible light source orientations to constrain smooth surface boundaries further. Boundaries that are inconsistent with a global light source direction are disallowed. This algorithm has been extensively tested [Brown et al., 1982]. The essential steps are as follows:

1. Edges in the body give possible space curves in three dimensions. (In the experiments by Brown et al. [1982], the three-space curve was assumed known, but the surface normal direction in the plane orthogonal to the curve was not. Methods of estimating surface tilt from boundary have been developed (e.g., [Brady and Yuille, 1983].
2. An initial estimate for the illuminant, together with the constraint that normals at the boundary must be perpendicular to the space curve, determines normal estimates.
3. These estimates refine the estimate for the illuminant, and smoothness and irradiance constraints allow more normals to be determined and refined.

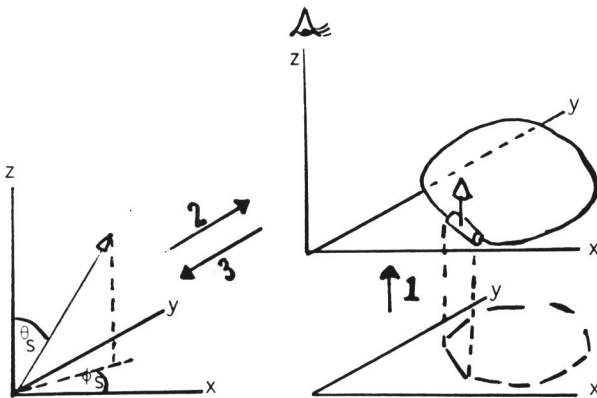


Figure 3. Use of local edge boundary and global illumination constraints in determining surface normals (see text).

B. Example 2: Rigid Body Motion

Consider the problem raised by a disk sliding over a planar surface. Here there are only small depth changes. One approach to the 2-D problem is to invoke a rigid body constraint and determine global motion parameters by a transformation technique [Ballard and Kimball, 1983]. Briefly, the method works as follows in 3-D (Figure 4):

1. The origin of the body, x_b , is assumed to be initially positioned at the viewer.
2. Estimates for flow and depth are used to compute estimates for three-dimensional velocities $v(x,t)$.
3. Three successive frames of three-dimensional velocities can determine the direction of rotation ω .
4. This direction, together with the 3-d velocities, can determine the magnitude of rotation.
5. Knowing $|\omega|$, ω , and v_T allows $x_b(x,t)$ to be updated.
6. The parameters x_b , $|\omega|$, ω , and v_T , can segment the three-dimensional velocities (and therefore flow) into a region where u is consistent with the parameters and a region where u is inconsistent with the parameters.

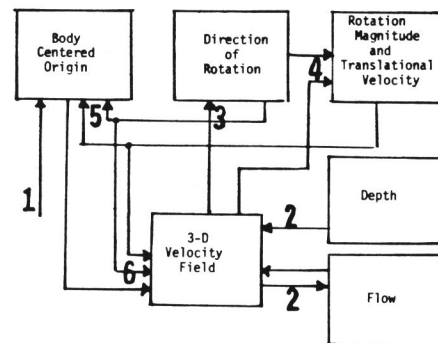


Figure 4. Use of global rigid body motion to determine flow boundaries (see text).

VI. Conclusions

The original concept of a boundary in the computation of multiple intrinsic images was a luminance boundary [Barrow and Tenenbaum, 1978]. Since then in the process of computing intrinsic images, the general problem of having appropriate boundary conditions has become more apparent. However, intervening work has concentrated on boundary conditions in single intrinsic images. Our work shows that the collaborative computation of several intrinsic images allows the possibility of: (1) boundary conditions being defined as conjunctions of multiple intrinsic properties; (2) a well-defined boundary in one image helping the computation of another; and (3) boundaries being determined by global features.

References

- Baker, H. and P. Blicher, Review of: *From Images to Surfaces: A Computational Study of the Human Early Visual System* by W.E.L. Grimson, *SIGART* 82, October 1982.
- Ballard, D.H., "Parameter networks: Towards a theory of low-level vision," *Proc.*, 7th IJCAI, Vancouver, B.C., August 1981b; to appear, *Artificial Intelligence*, 1983.
- Ballard, D.H. and O.A. Kimball, "Rigid body motion from depth and optical flow," TR 70, Computer Science Dept, U. Rochester, 1981; to appear, *CGIP*, Special Issue on Computer Vision, 1983.
- Barnard, S.T. and W.B. Thompson, "Disparity analysis of images," TR 79-1, Computer Science Dept, U. Minnesota, January 1979.
- Barrow, H.G. and J.M. Tenenbaum, "Recovering intrinsic scene characteristics from images," in A.R. Hanson and E.M. Riseman (Eds). *Computer Vision Systems*. New York: Academic Press, 1978.
- Brady, M. and A. Yuille, "An extremum principle for shape from contour," submitted to IJCAI-83, 1983.
- Brown, C.M., D.H. Ballard, and O.A. Kimball, "Constraint interaction in shape-from-shading algorithms," *Proc.*, DARPA Image Understanding Workshop, Palo Alto, CA, September 1982; 1982-83 *Research Review*, Computer Science Dept., U. Rochester, 1982.
- Bruss, A.R., "The image irradiance equation: its solution and application," Ph.D. thesis and MIT AI Memo 623, June 1981.
- Grimson, W.E.L., "A computational theory of visual surface interpolation," MIT AI Memo 613, June 1981.
- Horn, B.K.P., "Obtaining shape from shading information," in P.H. Winston (Ed). *Psychology of Computer Vision*. New York: McGraw-Hill, 1975.
- Horn, B.K.P. and B.G. Schunck, "Determining optical flow," *Artificial Intelligence* 17, 185-203, 1981.
- Ikeuchi, K. and B.K.P. Horn, "Numerical shape from shading and occluding boundaries," *Artificial Intelligence* 17, 141-184, 1981.
- Marr, D. and T. Poggio, "Cooperative computation of stereo disparity," *Science* 194, 283-287, 1976.
- Shafer, S.A. *MOOSE: User's Manual, Implementation Guide, and Evaluation*. IIF-HH B-70/80, Fachbereich Informatik, U. Hamburg, West Germany, April 1980.
- Shafer, S.A. and T. Kanade, "Using shadows in finding surface orientations," *Proc.*, DARPA Image Understanding Workshop, 90-102, Palo Alto, CA, September 1982.
- Snyder, W.E., "Computer analysis of time varying images," *IEEE Computer* 14, 8, August 1981.
- Stuth, B.H., "Local computation of motion," TR 116, Computer Science Dept., U. Rochester, July 1982.
- Tenenbaum J.M. and S. Weyl, "A region-analysis subsystem for interactive scene analysis," *Proc.*, 4th IJCAI, 682-687, September 1975.
- Terzopoulos, D., "The role of constraints and discontinuities in visible-surface reconstruction," MIT AI Lab, January 1983; submitted to IJCAI, 1983.
- Ullman, S., "Relaxation and constrained optimization by local processes," *CGIP* 10, 115-125, 1979.
- Wohn, K., L.S. Davis, and P. Thrift, "Motion estimation based on multiple local constraints and nonlinear smoothing," TR-1188, Computer Vision Laboratory, U Maryland, August 1982.