

- [12] P. Rousseeuw and A. Leroy, *Robust Regression and Outlier Detection*. New York: Wiley, 1987.
- [13] S. Barnard and W. Thompson, "Disparity analysis of images," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-2, pp. 333-340, July 1980.
- [14] W. Thompson and J. Kearney, "Inexact vision," in *Proc. Workshop Motion: Representation Analysis*, 1986, pp. 15-21.

Impossible Shaded Images

Berthold K. P. Horn, Richard S. Szeliski, and Alan L. Yuille

Abstract—In this correspondence, we show that images that could not have arisen from shading on a smooth surface with uniform reflecting properties and lighting exist. Much work has been done on recovering surface shape from images, and there has been some attention paid to the question of uniqueness. It has been shown, for example, that singular points curtail ambiguity. However, little has been said about the existence of solutions, perhaps because in practice, the given image is assumed to have arisen from a uniform, smoothly curved surface, and therefore, one knows that there must be at least one solution. What if, however, the reflecting properties of the surface vary from place to place? What if the actual surface does not reflect light the way one has assumed or that the light source is not where it was thought to be? Will the solution only be warped by these departures from the ideal model, or may there in fact be situations where there is no smooth surface that could have given rise to the given shading pattern? Can the fact that a shaded image of some surface with spatially varying surface reflectance is impossible in this sense be used to detect surface albedo variations?

Index Terms— Existence, impossible images, shape from shading, uniqueness.

I. INTRODUCTION

The problem of *shape from shading* has a history almost as long as that of computer vision itself [13]. Aside from the development of algorithms for recovering shape from shaded images, some attention has been paid to the problem of uniqueness of the solution. It has been shown that *singular points* of brightness in the image (corresponding to isolated global extrema in the reflectance map) play an important role in limiting the number of possible solution surfaces [2]-[5], [19].

Thus far, little has been said, however about the existence of solutions (but see [19]). Surfaces with continuously varying surface orientation give rise to shaded images. Are there brightness patterns that could not have arisen this way? Can such *impossible shaded images* be detected directly from their brightness patterns without explicitly solving the shape-from-shading equations? We show here that this is indeed the case.¹ In this correspondence, we will assume

Manuscript received December 30, 1989; revised November 12, 1991. Recommended for acceptance by Associate Editor J. Mundy.

B. K. P. Horn is with the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139.

R. S. Szeliski is with the Digital Equipment Corporation, Cambridge, MA 02139.

A. L. Yuille is with the Division of Applied Sciences, Harvard University, Cambridge MA 02138.

IEEE Log Number 9206560.

¹This work arose from a conjecture by one of us. Following initial confirmation of the conjecture by the co-authors and others (see Acknowledgment), we received note of work by Oliensis [17], that demonstrates that small fluctuations in intensities could indeed make most images "impossible."

that the distribution of light sources and the reflecting properties of the surface are known and that the reflecting properties of the surface are uniform. We also assume that the surfaces are *smooth*, by which we will mean that they have continuous first derivatives.

II. REFLECTANCE MAP AND IMAGE IRRADIANCE EQUATION

The brightness at a point in the image (the image irradiances) is proportional to the brightness of the corresponding point on an object (the scene radiance [11]). The latter depends on a) the reflecting properties of the surface material, b) the distribution and intensity of the light sources, and c) the surface orientation.

Surface orientation has two degrees of freedom and can be specified in several different ways. The slopes $p = (\partial z / \partial x)$ and $q = (\partial z / \partial y)$ in two orthogonal directions are convenient for this purpose, where $z(x, y)$ is the height of the surface above some reference plane perpendicular to the direction of projection of orthographic image formation.² Surface orientation may also be specified by means of the unit normal, which can be obtained from the components of the gradient p and q as follows $\hat{n} = \left(-p, -q, 1 \right)^T / \sqrt{1 + p^2 + q^2}$.

The *reflectance map* $R(p, q)$ gives scene radiance as a function of surface orientation and encodes information about both the surface reflecting properties and light source distribution [10]. It can be computed given the *bi-directional reflectance distribution function* (BRDF) [14], [11] or determined experimentally using the image of a calibration object.

The (normalized) *image irradiance equation* is $E(x, y) = R(p(x, y), q(x, y))$, where $E(x, y)$ is the image irradiance at the point (x, y) in the image, whereas $p(x, y)$ and $q(x, y)$ are the partial derivatives of $z(x, y)$ at the corresponding point on an object in the scene [10], [11].

The shape-from-shading problem is that of recovering the surface $z(x, y)$ given the image $E(x, y)$ and the reflectance map $R(p, q)$. The image irradiance equation can be viewed as a first-order nonlinear partial differential equation; therefore, it can be solved using the method of characteristic strips [7], [9], [11].

A. Phenomenological Models of Reflection

Some of the impossible images we will present depend on particular properties of a class of reflectance maps, such as rotational symmetry. At other times, it is useful to have a very specific reflectance map in mind, such as that of a Lambertian surface under point source illumination. Let us consider a simple imaging situation where we are dealing with an idealized surface material that satisfies two conditions: a) It appears equally bright from all viewing directions, and b) it reflects all incident light. Such a surface is called an (ideal) Lambertian surface, and it can be shown that when illuminated by a single light source, it satisfies Lambert's cosine law [11]. In this case, brightness depends on the cosine of the incident angle, the angle between the incident rays, and the surface normal and is independent of the direction towards the viewer. If there is a single light source in the direction given by the unit vector $\hat{s} = \left(-p_s, -q_s, 1 \right)^T / \sqrt{1 + p_s^2 + q_s^2}$, then we use the fact that the cosine of the incident angle is equal to $(\hat{n} \cdot \hat{s})$ and therefore obtain

²The shape-from-shading problem can be formulated in the case where the imaging system performs a perspective projection [18], [7] and when the light sources are near the objects being viewed, but this makes the analysis harder since scene radiance then depends on position as well as surface orientation.

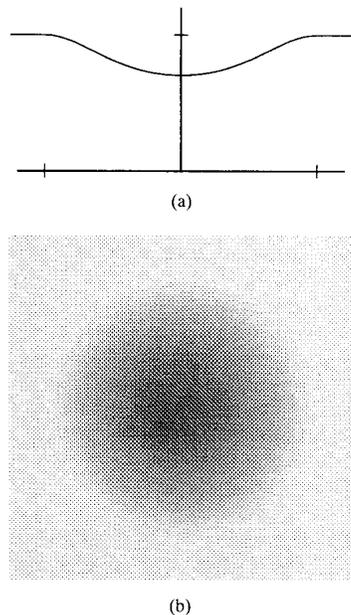


Fig. 1. (a) Cross-section of rotationally symmetric shaded image, (circular) shown in (b), which could not have arisen from a (smooth) surface with continuous first derivatives if we assume that the reflectance map is $R(p, q) = 1/\sqrt{1+p^2+q^2}$.

the (normalized) reflectance map

$$R(p, q) = \max \left[0, \frac{1 + p_s p + q_s q}{\sqrt{1+p^2+q^2} \sqrt{1+p_s^2+q_s^2}} \right].$$

A particularly simple case arises when the light source lies in the same direction as the viewer, that is, when, $(p_s, q_s) = (0, 0)$ because then, $R(p, q) = (1/\sqrt{1+p^2+q^2})$. This special case can be used when a concrete example of a rotationally symmetric reflectance map is needed with brightness decreasing monotonically with surface slope s , where $s = \sqrt{p^2+q^2}$.

III. IMPOSSIBLE SHADED IMAGES

A. Circularly Symmetric Dark Blotch

Suppose we are given the image

$$E(x, y) = \begin{cases} 1/\sqrt{1+\cos^2 r}, & \text{for } r < \pi/2; \\ 1, & \text{for } r \geq \pi/2 \end{cases}$$

where $r = \sqrt{x^2+y^2}$ (shown in Fig. 1), with reflectance map

$$R(p, q) = \frac{1}{\sqrt{1+p^2+q^2}}.$$

Then, there is a "solution"

$$z(x, y) = \begin{cases} \sin r; & \text{for } r < \pi/2; \\ 1; & \text{for } r \geq \pi/2. \end{cases}$$

Although this function is smooth almost everywhere, it has a conical singularity at the origin. Does there exist a solution that has continuous first derivatives everywhere? The answer is "no," as we show next (generalizing to an even wider class of impossible images).

B. Compact Dark Blotch on Unit Brightness Background

Suppose we are told that the reflectance map has a unique isolated global maximum of one at the origin, that is

$$R(p, q) \begin{cases} = 1, & \text{for } (p, q) = (0, 0); \\ < 1, & \text{otherwise.} \end{cases}$$

In this case, a surface facing the viewer directly has brightness one, and surface patches oriented differently are always darker (as would be the case if the surface was a Lambertian reflector with the light source in the direction towards the viewer). Suppose that image brightness is less than one in some simply connected compact region D , whereas brightness equals one outside this region, that is

$$E(x, y) \begin{cases} < 1, & \text{for } (x, y) \in D; \\ = 1, & \text{otherwise.} \end{cases}$$

Note that outside the region D , the surface gradient (p, q) must be zero since that is the only gradient for which $R(p, q) = 1$. If $p = 0$ and $q = 0$ in some connected region, then surface height $z(x, y)$ must be constant in that region. Now, either the surface $z(x, y)$ is constant in the region D or it has at least one extremum there. However, the surface cannot be constant in D because that would imply that the brightness there equaled one; therefore, it must have an extremum. The first partial derivatives must vanish at that extremum since we are assuming that the surface $z(x, y)$ has continuous first derivatives. However, this implies that the brightness at the extremum must be one, which contradicts the assumption that brightness is less than one everywhere in the region D . Thus, there is no surface with continuous first derivatives that will give rise to the given image.

We note that a compact dark blotch on a unity brightness background needs to have at least one interior point where brightness equals one if it is to be the shaded image of some surface with continuous first derivatives.

C. Source Not at Viewer

It is possible to extend the result of the previous section to reflectance maps that have their unique isolated global extremum somewhere other than the origin in gradient space. Suppose the extremum in the reflectance map occurs at $(p, q) = (p_s, q_s)$, as happens with a Lambertian surface when the light source is away from the viewer. Then, the surface orientation outside the region D is fixed, and the surface there is planar with surface normal

$$\hat{\mathbf{n}} = \frac{(-p_s, -q_s, 1)^T}{\sqrt{1+p_s^2+q_s^2}}.$$

We can then use a similar argument to that in the previous section by using a new coordinate system oriented with the z axis parallel to $\hat{\mathbf{n}}$. The only potential problem arises from the possibility that the point where the surface exhibits an extremum in the light-source coordinate system may be obscured by another portion of the surface when seen from the viewing direction. However, this can only occur if there is a fold in the surface, as seen from the viewing direction, and the derivatives of surface height are discontinuous at the fold.

D. Singular Points Are Not Required

The above examples might appear to suggest that singular points are crucial to the construction of impossible shaded image examples since the exterior of the region D consists entirely of singular points, where surface orientation can, in fact, be recovered locally, but this is not so. Consider, for example, a reflectance map that has brightness one for zero slope with brightness falling off with slope in such a way that one can write

$$f(s) \leq R(p, q) \leq g(s) \text{ where } s = \sqrt{p^2+q^2}$$

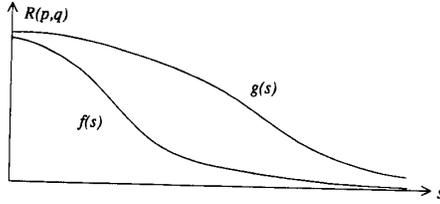


Fig. 2. In this generalization, the reflectance map need not be radially symmetric but must be bounded above and below by two radially symmetric, monotonically decreasing functions.

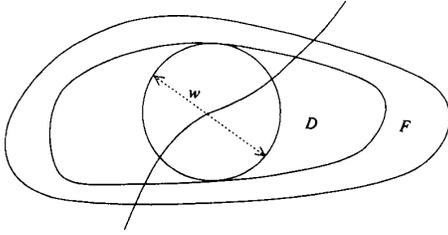


Fig. 3. For the impossible image discussed here, brightness is low inside the region D , high outside the region F , with a smooth transition in part of F that is not also in D . A curve of steepest descent that passes through the center of the largest circle that can be inscribed in D can be constructed. A contradiction is indicated if the change in height along this curve in the region D is greater than the change of height along the perimeter of region F .

for two monotonically decreasing function $f(s)$ and $q(s)$ (see Fig. 2).

A special case of this is a rotationally symmetric reflectance map with brightness dropping monotonically with slope. A particular instance of this special case is a Lambertian surface with the source at the viewer, as mentioned above.

Now, suppose that there is a compact, simply connected region D in which the brightness is low, nested inside another compact simply connected region F , outside of which brightness is high (see Fig. 3).

In the part of F that is not in D , brightness makes a smooth transition, that is

$$\begin{cases} 0 < E(x, y) < E_i & \text{for } (x, y) \in D. \\ E_i \leq E(x, y) \leq E_o & \text{for } (x, y) \in F - D. \\ E_o < E(x, y) < 1 & \text{otherwise.} \end{cases}$$

There are no singular points since $E(x, y) < 1$ everywhere; yet, as we show next, we can choose E_i and E_o such that there is no smooth surface giving rise to this shading.

We have, for the slope of the surface inside the region, $D: s \geq s_i = f^{-1}(E_i)$, whereas the slope outside F satisfies $s \leq s_o = g^{-1}(E_o)$. Now, inside the region F , the slope is guaranteed to be nonzero; therefore, the surface $z(x, y)$ has a unique nonzero gradient at every point. This gradient field can be integrated out to yield lines of steepest ascent on the surface. Such steepest ascent curves cannot cross or terminate in F , and so can therefore be followed all the way from one point on the boundary of F to another. Such a steepest ascent curve passing through a point in D will similarly cross the boundary of D in two places.

Suppose that w is the diameter of the largest inscribed circle of the region D . Then, the steepest ascent curve passing through the center of this circle must have length at least w and, hence, a change in z from one end to the other of at least ws_i . Now, suppose that p is the perimeter of the region F . Then, the shortest distance along the boundary of F between the two points where this steepest ascent curve touches the boundary is at most $p/2$, and therefore, the change in z is at most $(p/2)s_o$. Given w and p , we can now choose E_i and

E_o and, hence, s_i and s_o , such that $ws_i > (p/2)s_o$, which leads to a contradiction.

Note: The only reason that two nested regions are needed in this construction is to allow brightness to vary smoothly in the transition, as it must, since we have assumed that the first derivatives and, hence, brightness is continuous.

E. Nested Iso-Brightness Contours

The above construction can be extended to nested isobrightness contours of monotonically increasing brightness from the inside to the outside.

Suppose that we have a dark blotch in the image that increases monotonically in brightness from the inside outward so that one can construct a set of nested isobrightness contours for brightness $0 < E_0 < E_1 < E_2 \dots < E_n < 1$, where E_0 is the brightness of the darkest point in the image. Suppose that the minimum distance between the two isobrightness contours for $E = E_i$ and $E = E_{i+1}$ is w_i . Note that the slope of the surface on points lying between these two isobrightness contours is constrained by $s > s_i = f^{-1}(E_{i+1})$. Consider the curve of steepest ascent passing through the point where $E = E_0$. The change in height along this contour between the points where it crosses the isobrightness contour $E = E_n$ is bounded below by $\delta z_n > 2 \sum_{i=0}^{n-1} s_i w_i$ while at the same time bounded above by $\delta z_n \leq (p_n/2) \bar{s}_n$, where p_n is the length of the isobrightness contour $E = E_n$ and $\bar{s}_n = g^{-1}(E_n)$. We have an impossible image unless $2 \sum_{i=0}^{n-1} s_i w_i \leq (p_n/2) \bar{s}_n$ for all n . This provides a way of constructing a variety of impossible images. It also provides a limit on how dark a blotch on a bright background can be before it can no longer be interpreted as shading on an inclined portion of the surface. These topics are explored further in [15].

F. Fold in Riemann Sheets on Gaussian Sphere

Imagine that we have a rotationally symmetric reflectance map that drops to zero at infinity in gradient space, such as the Lambertian surface with the source at the viewer. For what we will do next, it is convenient to think of the reflectance map as a function of position on the Gaussian sphere rather than as a function of the components of the gradient. The reflectance map plotted on the Gaussian sphere here has a peak of one at the "pole," corresponding to the viewing direction, and drops off to zero at the "equator," corresponding to points on the occluding boundary of the object being viewed.³

Now, suppose that we are given an image that has nonzero brightness in the interior of some compact simply connected region D with zero brightness on the boundary ∂D of this region. Then, the boundary ∂D is a *silhouette*, that is, the projection of an occluding contour on the object being viewed. If we assume that surface orientation varies continuously, there is a mapping from the object's surface to the surface of the Gaussian sphere that covers every point in one hemisphere (at least once). We can see this by noting that for any orientation in the hemisphere, there must be a point on the surface with that orientation since a plane with that orientation as its normal approaching from infinity will touch the surface somewhere [6] (if the object is convex, the *Gauss map* is invertible). Note that the occluding boundary maps onto the "equator" of the hemisphere.

It should, first of all, be clear that there must be at least one point in the image where the brightness equals the maximum brightness in the reflectance map since the pole of the hemisphere must be covered. How many such extrema can there be in the image of a single object with continuous first derivatives? We can have more than one if the object is not convex since the mapping from the surface onto the

³The Gaussian sphere approach to the analysis of shading was introduced in [16].

sphere then folds over itself. However, every time we fold it over in order to cover the pole more than once, we add *two* new places where the surface is oriented for maximal reflection of light. This suggests that there must be an odd number of bright spots in the image. There is one exception to this rule: If the fold on the Gaussian sphere happens to cross the pole, it will yield only one maximum instead of two. Of course, in this special case, any slight change in the orientation of the object with respect to the viewer will change this. Therefore, this does not apply if we assume that the viewer is in "general position."

The other possibility is that the part of the surface carrying one of the two points happens to be obscured by another part of the object, but in this case, the surface $z(x, y)$ is not a continuous function of x and y within in the region D . Furthermore, the occluding boundary is not a simply closed curve, and parts of the occluding boundary lie within D , but this is impossible since the brightness is nonzero inside D .

G. Multiple Viewpoints or Multiple Lighting Conditions

Thus far, we have been trying to determine the "impossibility" of surfaces from a *single* image. This problem (naturally) becomes a lot easier if we have several images of the surface corresponding to different lighting conditions or different viewpoints. We can, for example, make use of certain photometric invariants [16], [20] that relate properties of the surface geometry to properties of the image brightness, assuming only a generic form for the reflectance map $R(p, q)$ and constant albedo. One specific result [20], extending a result of [16], is that the directions of the isophotes (the lines of constant image brightness) must always lie along the directions of principal curvature at parabolic lines (lines of zero Gaussian curvature); hence, these isophote directions will be invariant as we alter the viewing conditions—this has been exploited by [2]. Moreover, for generic surfaces, these are the only lines with this property. Thus, given several images, we can use these results to determine the parabolic lines of the surface. For regular (that is, not "impossible") surfaces, the parabolic lines will either be closed contours or will terminate at the boundaries of the viewed object. If we find parabolic lines that terminate inside the object, or have other undesirable behavior, then we have an "impossible" surface.

H. Iterative Solution Applied to Impossible Image

It is of interest to see what the iterative algorithm [12] will do when presented with an impossible shaded image. It is shown [15] that it finds the "solution"

$$z(x, y) = \begin{cases} \sin r: & \text{for } r < \pi/2; \\ 1: & \text{for } r \geq \pi/2. \end{cases}$$

This function is smooth everywhere except at the origin, where it has a conical singularity.

IV. DETECTING SPATIAL VARIATIONS OF ALBEDO

If a surface has a spatially varying reflecting properties, or if the illumination has spatial variations, then the normal shading rules for the image are altered. An extreme example of this is a photographic print, where all the brightness variations are due to spatially varying reflectance, and there is no shading resulting from spatial variations in surface orientation. Given the limited information in a single image, it often is not possible to separate the contributions to the brightness pattern that come from spatially varying surface orientation and those that come from spatial variations in reflectance or illumination.

As we have demonstrated in this paper, however, it is sometimes possible to show that the given image could not have arisen from a uniformly illuminated smooth surface with uniform reflecting

properties (that is, it is an impossible shaded image). The way this manifests itself when iterative algorithms are used for recovering the surface shape is that the functional cannot be reduced to zero and that discontinuities and cone-shaped singularities in surface orientation remain in the estimated solution.

When we look at images taken of the surface of rocky planets like Mars, we can get a clear impression of the shapes of surface features such as impact craters, yet we can also be aware of the fact that surface albedo varies from place to place. Our ability to separate shading and albedo variations suggest that there is some way of distinguishing the two. Quite often, what distinguishes shading from surface markings is that the latter have sharp transitions between regions of relatively constant reflectance, whereas shading typically varies smoothly. What is needed, then, is a simultaneous solution of the shading and the lightness problems [8], [1].⁴ At least there now is a diagnostic test that tells us when the assumptions of uniform albedo and uniform illumination are being violated.

Similarly, if the assumed light source is in the wrong position, there will typically not be a solution to the shape-from-shading problem. This again manifests itself as a residual error in the iterative scheme. It has been found possible, for example, to refine an estimate of the light source position by searching for the position that minimizes the residual errors [12].

V. CONCLUSIONS

We have shown in this paper that shaded images that cannot have originated from a uniformly illuminated, smooth continuous surface with uniform albedo exist. The typical condition where this occurs is when we have a dark area (corresponding to a region of high gradient) surrounded by a lighter region (with low gradient). For this to correspond to a real surface, we can establish that there must be a local extremum or area of lower gradient inside the dark region. This, in turn, will show up as either a light area in the image or an orientation discontinuity in the surface (thus violating either our intensity or smoothness constraints). We can also sometimes establish the impossibility of a shaded image by counting the number of extrema inside a region corresponding to an isolated surface patch.

The theoretical arguments we have presented are in agreement with the effects observed with numerical shape-from-shading algorithms. When presented with an "impossible shaded image," the algorithm will find a solution that is smooth almost everywhere but has isolated orientation singularities ("peaks" or "cusps"). We thus have two methods for detecting when the assumptions behind our shape-from-shading algorithm are being violated. First, we can examine the intensity image to check if any of the theoretical conditions for impossible shading exist. Second, we can monitor the output of our numerical shape-from-shading algorithm to see if isolated or connected singularities exist in the final solution. Detecting these violations will hopefully lead us to more robust and more general shape-from-shading algorithms, which can detect albedo variation and discontinuities in the reconstructed surface.

ACKNOWLEDGMENT

The idea for this paper grew out of a competition organized by one of the authors (BKPH), who had a conjecture that the image presented at the beginning of Section III was indeed an impossible shaded image. Proofs of impossibility were received first from one of the coauthors (RSS) and then the other (ALY), followed by J. Smith, R. Kozera, M. Brooks, J. Oliensis, and D. Lee. We would like to thank the others, as well as M. Gennert and B. Saxberg, for their interest in this problem.

⁴Simultaneous estimation of shape and albedo has been described in a recent paper [21].

REFERENCES

- [1] A. Blake, "Boundary conditions for lightness computation in Mondrian world," *Comput. Vision Graphics Image Processing*, vol. 32, no. 3, pp. 314-327, Dec. 1985.
- [2] A. Blake, A. Zisserman, and G. Knowles, "Surface descriptions from stereo and shading," *Image Vision Comput.*, vol. 3, no. 4, pp. 183-191 1985; also in *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [3] M. J. Brooks, "Two results concerning ambiguity in shape from shading," in *Proc. Nat. Conf. Artificial Intell.* (Washington, DC), Aug. 22-26, 1983, pp. 36-39.
- [4] A. R. Bruss, "The Eikonal equation: Some results applicable to computer vision," *J. Math. Phys.*, vol. 23, no. 5, pp. 890-896, May, 1982; also in *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [5] P. Deift and J. Sylvester, "Some remarks on the shape-from-shading problem in computer vision," *J. Math. Anal. Applications*, vol. 84, no. 1, pp. 235-248, Nov. 1981.
- [6] D. Hilbert and S. Cohn-Vossen, *Geometry and the Imagination*. New York: Chelsea, 1952.
- [7] B. K. P. Horn, "Shape from shading: A method for obtaining the shape of a smooth opaque object from one view," Ph.D. thesis, Dept. Elect. Eng., Mass. Inst. Technol., 1970; also AD717336 available from Nat. Tech. Inform. Service.
- [8] —, "Determining lightness from an image," *Comput. Graphics Image Processing*, vol.3, no. 1, pp. 277-299, Dec. 1974.
- [9] —, "Obtaining shape from shading information," in *The Psychology of Computer Vision* (P. H. Winston, Ed.). New York: McGraw-Hill, 1975, pp. 115-155, ch. 4; also in *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [10] —, "Understanding image intensities (sic)," *Artificial Intell.*, vol. 8, no. 2, pp. 201-231, Apr. 1977; also in *Readings in Computer Vision* (M. A. Fischler and O. Firschein, Eds.). Los Altos, CA: Morgan Kaufmann, 1987, pp. 45-60.
- [11] —, *Robot Vision*. Cambridge, MA: MIT Press, 1986, pp. 202-277, ch. 10, 11.
- [12] —, "Height and gradient from shading," *Int. J. of Comput. Vision*, vol. 5, no. 1, pp. 37-75, Aug. 1990; also Memo 1105, MIT Artificial Intell. Lab.
- [13] B. K. P. Horn and M. J. Brooks, Eds. *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [14] B. K. P. Horn and R. W. Sjoberg, "Calculating the reflectance map," *Applied Opt.*, vol. 18, no. 11, pp. 1770-1779, June 1979; also in *Proc. DARPA Image Understanding Workshop* (Pittsburgh, PA), Nov. 14-15, 1978, pp. 115-126; also in *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [15] B. K. P. Horn, R. Szeliski, and A. L. Yuille, (1991) "Impossible shaded images," Harvard Robotics Lab. Tech. Rep. 91-15, Nov. 1991.
- [16] J. J. Koenderink and A. J. van Doorn, "Photometric invariants related to solid shape," *Optica Acta*, vol. 27, no. 7, pp. 981-996, 1980; also in *Shape from Shading*. Cambridge, MA: MIT Press, 1989.
- [17] J. Oliensis, "Existence and uniqueness in shape from shading," Unpublished Rep., Comput. Inform. Sci., Univ. Mass. Amherst, 1989.
- [18] T. Rindfleisch, "Photometric method for lunar topography," *Photogram. Eng.*, vol. 32, no. 2, pp. 262-277, Mar. 1966.
- [19] B. V. H. Saxberg "A modern differential geometric approach to shape from shading," Ph.D. thesis, Dept. Elec. Eng. Comput. Sci., Mass. Inst. Technol., 1988.
- [20] A. L. Yuille, "Zero crossings on lines of curvature," *Comput. Vision Graphics Image Processing*, vol. 45, pp. 68-87, 1989.
- [21] Q. Zheng and R. Chellappa, "Estimation of illuminant direction, albedo, and shape from shading," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 13, no. 7, pp. 680-702, May 1991.

Cluster Analysis by Binary Morphology

J. -G. Postaire, R. D. Zhang, and C. Lecocq-Botte

Abstract—A new approach to unsupervised pattern classification, which is based on the use of mathematical morphology operations, is developed. The way a set of multidimensional observations can be represented as a mathematical discrete binary set is shown. Clusters are then detected as well separated subsets by means of binary morphological transformations.

Index Terms—Binary morphology, cluster analysis, dilation, erosion, unsupervised classification.

I. INTRODUCTION

The aim of cluster analysis is to divide a set of objects into subsets according to their similarities and dissimilarities. The objects are generally represented by N -dimensional vectors of observed features. Among the various types of procedures for clustering such sets, the most popular are the hierarchical, the nonhierarchical, and the statistical approaches [1].

Independent from this research in pattern classification, mathematical morphology has been developed as an algebra of set-theoretic operations in Euclidean space for quantitative description of geometrical structures. As introduced by Matheron and Serra [2], [3], this approach has been mainly concerned with image analysis [4].

The intent of the present paper is to introduce the use of morphological filters in the field of cluster analysis. Since the language of mathematical morphology is that of set theory, it is first shown how any finite set of multivariate observations can be considered in terms of a mathematical set in a multidimensional Euclidean space. In Section II, we propose a fast preprocessing technique that yields a data representation suitable for binary morphological transformations. Morphologically filtering a data set corresponds to a sequence of dilations and erosions, which are the basic morphological operations. These operations are iteratively applied to eliminate irrelevant details in the shape of the clusters without geometric distortion (Section III). A new clustering procedure based on the detection of cluster cores by means of this nonlinear filtering technique is proposed (Section IV). The performance of this procedure is demonstrated in Section V using real and artificially generated data sets.

II. REPRESENTATION OF A SET OF MULTIDIMENSIONAL OBSERVATIONS AS A DISCRETE BINARY SET

Let us consider a set of Q N -dimensional observations

$$Y^* = \{Y_1, Y_2 \cdots Y_q, \cdots Y_Q\}$$

such that

$$Y_q = [y_{q,1}, y_{q,2}, \cdots, y_{q,n}, \cdots, y_{q,N}]^T.$$

In order to extend the theory of mathematical morphology to cluster analysis, this set of available observations must be represented as a mathematical discrete set in a Euclidean space. The procedure here

Manuscript received April 16, 1990; revised October 11, 1991. Recommended for acceptance by Associate Editor C. Dyer.

The authors are with the Centre d'Automatique, Université des Sciences et Techniques de Lille Flandres Artois, Villeneuve d'Ascq Cedex, France.
IEEE Log Number 9202387.