Reconstruction of shape from shading in color images

L. L. Kontsevich

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco, California 94115

A. P. Petrov

Institute of Atomic Energy, Kurchatov Square, Moscow 123182, Russian Federation

I. S. Vergelskaya

Institute for Problems of Information Transmission, Ermolovoy 19, 101447 Moscow, Russian Federation

Received April 2, 1992; revised manuscript received July 16, 1993; accepted October 21, 1993

We propose a method for shape reconstruction from color shades produced by multiple chromatic light sources. The linear relation between surface-normal vectors and three-dimensional response vectors for a uniformly colored and illuminated region of a surface can be reconstructed in two steps. In the first step a quadratic form of metric in response space induced from a natural metric in normal space is reconstructed. At this stage proper image segmentation can be obtained. In the second step an exact mapping from response space into the space of surface normals is reconstructed. The matrix for this mapping is one of the square roots of the quadratic-form matrix that satisfies the integrability constraint. The method is in all respects much simpler than existing methods for solving the depth-from-shading task for monochromatic images.

1. INTRODUCTION

It is well known that shades create a strong impression of depth, although shading, like most other depth cues, cannot be interpreted without some *a priori* knowledge about the properties of possible scenes. For instance, a static three-dimensional scene and its two-dimensional photograph can produce identical patterns of responses in the retina. In general terms the constraint that the depth-from-shading mechanism imposes onto natural scenes is usually formulated as follows: spatial variations of ambient light and surface color are generally smaller than variations of surface orientation.[1,2] Most known approaches to the problem of shape-from-shading computation use stronger and more specific constraints, such as constancy of the surface-reflection function and illumination of all elements of a surface by the same (usually restricted) set of light sources with known apertures.

A natural scene taken as a whole does not satisfy all these constraints. Therefore the first essential task must be segmentation, that is, isolation of regions that satisfy some of the constraints mentioned above. There are stringent requirements for segmentation: failure at this stage inevitably leads the shape computation to a meaningless solution.

One approach to the problem takes input for computer simulations of the shape-from-shading mechanism from monochromatic images, and some successful algorithms have been proposed.[1,3,4] However, for monochromatic images the segmentation task cannot be solved reliably.

Adding color information makes segmentation much more reliable. A theory of color-related segmentation was elaborated by Nikolaev.[5,6] This theory was paraphrased in English by Brill[7] with some modifications. The central concept of Nikolaev's theory is rank. Nikolaev's definition of rank is too detailed for our purposes, and we use the definition proposed by Brill. According to Brill, the rank of a given region on a surface is equal to the dimension of the set of response vectors from this region. Nikolaev described elegant algorithms for isolation of uniformly colored Lambertian regions that have ranks of 1 and 2. (In most cases the regions are illuminated by 1 or 2 light sources with different spectra.) However, the mathematical method used in Nikolaev's theory is not applicable to the regions of rank 3 (in most cases these regions are illuminated by three or more light sources with different spectra).

The case of numerous light sources with different spectra is not so exotic as it may seem: natural scenes usually contain many objects, which often serve as secondary light sources for one another.

In the present paper we generalize Nikolaev's theory for the case of arbitrary light-source apertures (Nikolaev dealt mainly with distant point light sources). Then we describe the algorithm for identification of regions of rank 3. Our approach is based on a method for decomposition of arbitrary illumination to three basic light sources, proposed by Petrov.[8] The advantage of this method is that it avoids some constraints and achieves solutions for light sources and reflectance with arbitrary spectra. We then describe a surprisingly simple solution of the shape-from-shading problem for regions illuminated by many light sources.[9] Results of experimentation with natural and computer-generated images show that this algorithm is precise and stable against input errors.

0740-3232/94/031047-06$06.00

2. LINEAR RELATION BETWEEN RESPONSES AND SURFACE-NORMAL VECTORS

First we consider the simple case in which a surface is illuminated by a single infinitely distant source of light. Let the direction toward the light source be given by unit vector $\hat{p}$, and let $S(\lambda)$ be the energy distribution of ambient light as a function of the wavelength λ.

Assume that the surface is Lambertian. Then reflectance properties for its elements can be described by the spectral reflectance function $\rho(\lambda)$. The local orientation of the surface is defined, according to convention, by the unit normal vector $\hat{n}$.

The receiving system consists of optical and receptive components. The first component projects the scene onto the receptive surface. We assume for simplicity that the receptive surface is flat and perspective distortions are negligible and therefore that the orthogonal projection is an adequate approximation. The second component of the system, the receptive surface, has in all its pixels identical triads of spectrally selective sensors. Suppose for simplicity that responses of the sensors are linear functions of light flow. The spectral sensitivity of the ith sensor is denoted $\nu_i(\lambda)$.

Thus in each pixel a three-dimensional response vector $\mathbf{r}$ is available. In a pixel corresponding to some surface element, the components of the response vector can be expressed as follows:

$$r_i = \hat{n} \cdot \hat{p} \int S(\lambda)\rho(\lambda)\nu_i(\lambda)\mathrm{d}\lambda\,, \tag{2.1}$$

or, in a brief form,

$$\mathbf{r} = \mathbf{M}\hat{n}\,, \tag{2.2}$$

where $\mathbf{M}$ is a matrix with entries

$$m_{i,j} = p_j \int S(\lambda)\rho(\lambda)\nu_i(\lambda)\mathrm{d}\lambda \tag{2.3}$$

(p_j is the jth coordinate of vector $\hat{p}$).

This linear relation is valid only when the light source really illuminates the surface. This means that normals must belong to a half-space defined by the inequality

$$\hat{n} \cdot \hat{p} \geq 0\,. \tag{2.4}$$

Now we are ready to consider the general case in which the surface element is illuminated by an arbitrary number of infinitely distant point light sources. As long as the receiving system is linear, sensor responses for the set of light sources are equal to the sum of responses produced by each light source alone. That is, relation (2.2) is still valid, but now

$$\mathbf{M} = \sum_{\hat{n}\cdot\hat{p}\geq 0} \mathbf{M}(\hat{p})\,, \tag{2.5}$$

in which all light sources satisfying condition (2.4) are included (Fig. 1). For extended light sources, continuous integration must be used instead.

If we rotate a surface element while the light sources remain stationary, the given matrix $\mathbf{M}$ will be constant as long as the surface element is illuminated by the same constant set of light sources. This means that different elements of a uniformly colored surface can have the same matrix $\mathbf{M}$.

Elements of a large piece of the surface can be illuminated by different sets of light sources and may have a different spectral reflectance. In Section 3 we describe an algorithm for the isolation of a uniformly colored region all of whose points are illuminated by the same set of light sources.

3. IDENTIFICATION OF A REGION OF RANK 3

The rank of matrix $\mathbf{M}$ is precisely the rank as defined by Brill,[7] but now this concept is applicable to the case of illumination by light sources with arbitrary apertures. For isolation of rank-1 and rank-2 regions Nikolaev's algorithms can be used without any modifications.

When some surface element is illuminated by one distant point light source or by a set of distant sources with similar spectra, the matrix $\mathbf{M}$ has rank 1. According to Eq. (2.5), $\mathbf{M}$ can have rank 3 if the surface element is illuminated by at least three light sources that are spatially separated and have different spectral functions (but this condition is not sufficient). Here we consider only the case of rank 3, that is, when matrix $\mathbf{M}$ is not singular.

There is another source that reduces the dimension of the response set. This is the dimension of the normal set. The dimension of the response set cannot be greater than the dimension of the normal set. There are two kinds of surface that have a dimension of normals that is less than 3. These are a plane (its normal set is one dimensional) and a generalized cylinder (its normal set is two dimensional). There are simple methods to detect such surfaces, which remain outside the scope of this study.

For a surface region all elements of which are uniformly colored and illuminated by the same set of light sources, there is a linear mapping $\mathbf{M}$ of normals to the response space. As long as $\mathbf{M}$ is not singular, it establishes an iso-

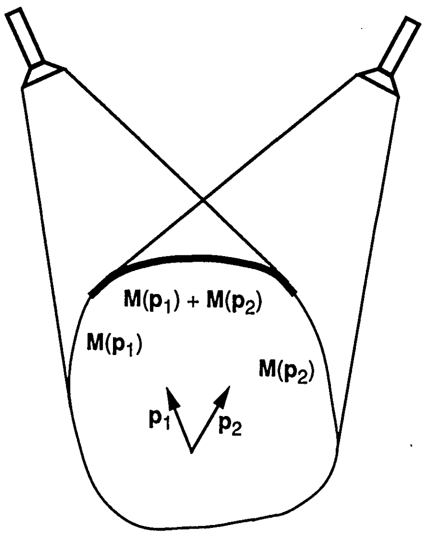


Fig. 1. Arbitrary surface illuminated by two light sources ($\hat{p}_1$ and $\hat{p}_2$). The region outlined by the heavy black line is illuminated by both sources simultaneously. In this region $\mathbf{M} = \mathbf{M}(\hat{p}_1) + \mathbf{M}(\hat{p}_2)$.

morphism between the physical space and the response space. Considering **M** as a basis transformation, we may take **r** for the coordinates of the normal vectors in a new basis.

Initially having been in the Euclidean basis, in the new basis the quadratic form of the metric will be determined by the matrix **Q**:

$$\mathbf{Q} = (\mathbf{M}^{-1})^T\mathbf{M}^{-1}. \quad (3.1)$$

A change of the basis does not affect length, so

$$|\mathbf{n}| = |\mathbf{r}|_{\mathbf{Q}} = \mathbf{r}^T\mathbf{Q}\mathbf{r} = 1. \quad (3.2)$$

Thus all response vectors **r** in the region must have a unit length in the metric that is defined by the quadratic form **Q**.

The quadratic form **Q** specifies the mapping **M** up to an orthogonal transformation **O** (rotation and symmetry), because

$$[(\mathbf{MO})^{-1}]^T(\mathbf{MO})^{-1} = (\mathbf{M}^{-1})^T\mathbf{O}\mathbf{O}^T\mathbf{M}^{-1} = (\mathbf{M}^{-1})^T\mathbf{M}^{-1}. \quad (3.3)$$

Q can be recovered from the image by means of Eq. (3.2). This relation provides a linear equation for elements of the symmetrical matrix **Q**:

$$r_1^2q_{1,1} + r_2^2q_{2,2} + r_3^2q_{3,3} + 2r_1r_2q_{1,2} + 2r_1r_3q_{1,3} + 2r_2r_3q_{2,3} = 1. \quad (3.4)$$

Since **Q** is determined by six parameters, one needs to know responses in at least six pixels of the region to obtain **Q**.

Suppose that we have calculated the quadratic form **Q** for some set of pixels in the region. As noted above, the response vectors from the region to be isolated have the unit length in metric that is defined by **Q**. But on the region's border either the surface color or the set of light sources changes, and Eq. (3.2) is violated.

This simple scheme cannot be applied directly to real-image processing. Surface color fluctuates, its reflectance is not perfectly Lambertian, and light sources are not infinitely distant. Uniformly colored and illuminated regions can be determined approximately, and the solution must fulfill two requirements: first, that the region be as large as possible and second, that the matrix **M** be approximately constant within the region. These requirements contradict each other.

We propose an algorithm that achieves a reasonable balance of these requirements. At the beginning it is necessary to choose six compactly located pixels (e.g., a 2 × 3 rectangle) in that part of the image where the region is to be identified. Let these points be an initial approximation of the region sought. Then we can calculate the quadratic form **Q** for this region and obtain norms of response vectors for all the image points. Let the new approximation of the region consist of all pixels that satisfy the following condition:

$$1 - \varepsilon_1 < |\mathbf{r}|_Q < 1 + \varepsilon_2. \quad (3.5)$$

(We empirically determined the values for $\varepsilon_1 = 1/3$ and $\varepsilon_2 = 1/2$ and used these in all the following examples.) Iterations continue as long as the region keeps growing.

Tests of this algorithm have shown that the region grows for several steps (not more than ten) and then stabilizes. The solution depends only slightly on the initial approximation of the region. If by chance the initial approximation falls upon the border of two adjacent regions, the algorithm converges to one of them.

4. INTEGRABILITY CONSTRAINTS FOR SURFACE NORMALS

Now we describe the conditions under which a vector field is normal to a surface. A vector field of surface normals defines a first-order differential form, that is, a differential of the coordinate z as a function of coordinates x and y on the surface:

$$\mathrm{d}z = -\frac{n_1}{n_3}\mathrm{d}x - \frac{n_2}{n_3}\mathrm{d}y. \quad (4.1)$$

It is well known that the first-order form defined on a simply connected region is a differential of some function if and only if the corresponding second-order form (the differential of the first-order form) is equal to zero. When a region has holes, it is necessary for contour integrals around the holes to be equal to zero, too. Thus the differential of the first-order form must be equal to zero.

Let us formulate this condition for a square grid of pixels. Partial derivatives at pixel positions can be obtained easily from normals in accordance with Eq. (4.1). We denote partial derivatives at the point halfway between adjacent pixels by p and q:

$$p_{i,j} = [(\partial z/\partial x)_{i,j} + (\partial z/\partial x)_{i+1,j}]/2,$$
$$q_{i,j} = [(\partial z/\partial y)_{i,j} + (\partial z/\partial y)_{i,j+1}]/2. \quad (4.2)$$

Equality to zero for the differential of the form 1 at a point with coordinates $(i + 1/2, j + 1/2)$ in terms of the calculus of finite differences can be written as follows:

$$p_{i,j} + q_{i+1,j} - p_{i,j+1} - p_{i,j} = 0. \quad (4.3)$$

Thus, in terms of the least-squares method, the expression

$$\sum_{i,j}(p_{i,j} + q_{i+1,j} - p_{i,j+1} - q_{i,j})^2 \quad (4.4)$$

vanishes for the correct normal field. This condition will be used extensively in the next section.

5. RECONSTRUCTION OF BASIS

Now we are ready to describe the final part of shape reconstruction. Suppose that on the image of a uniformly colored surface a region illuminated by the same set of light sources has been isolated. Then the scalar product for any pair of response vectors from this region can be computed. This means that we are able to find an angle between any two normals within the region. In other words, we know the relative directions of normals within the region.

Metric **Q** determines an orthonormal basis in space up to arbitrary rotation and symmetry [see Eq. (3.3)]. We shall try to find in the response space a basis corresponding to the Euclidean basis of the physical space whose **X** and **Y** axes are coincident with the **X** and **Y** axes of the

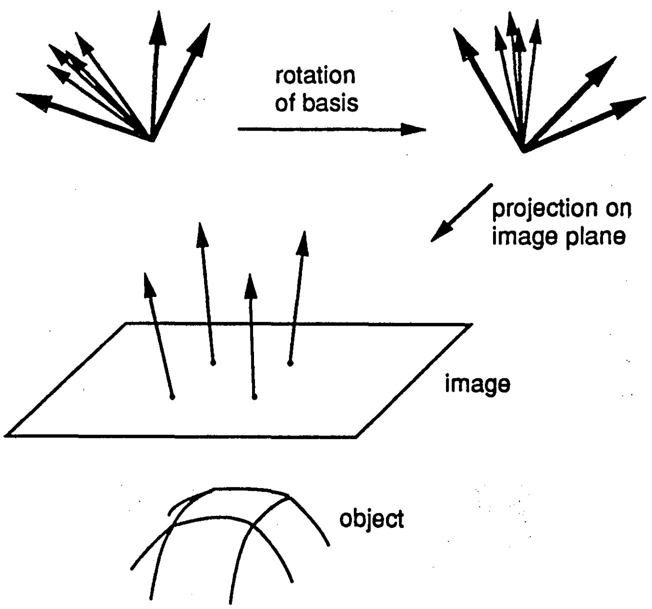


Fig. 2. Rotation of the basis in response space causes a change of normal field. The relative orientations of the normals remain invariable.

image. This basis defines the absolute orientation of normals relative to the image plane (Fig. 2).

To solve this problem we use the results given in Section 4. For correct orientation of normals, consistency condition (4.3) must hold in the region. This reduces the problem of finding the basis in response space to a search for the minimum of expression (4.4).

Thus it is a simple problem to find the correct orientation for the basis. Initially it is necessary to calculate an arbitrary orthonormal basis of response space in metrics **Q** (e.g., by means of the Gram–Schmidt algorithm). Then this basis should be transformed by rotations and central symmetries until the minimum of expression (4.4) is found. If the minimum value is close to a zero value, the solution is correct. Otherwise there is no solution for the entire region.

It is important to note that we seek the minimum, rather than the basis that precisely satisfies the consistency condition expression (4.4). Processing of images of real objects has shown that the value of expression (4.4) is as a rule close to zero. Nevertheless, if one is to obtain the surface, it is necessary to make a final correction for the normal field. The approximation of an inconsistent vector field by a consistent one is obtained by minimization of

$$\iint [(z_x - p)^2 + (z_y - q)^2]\mathrm{d}x\mathrm{d}y, \qquad (5.1)$$

where p and q are estimations of partial derivatives and z_x and z_y are partial derivatives of the approximating surface.

6. SIMULATION

The algorithms described here were tested extensively with the use of images of real and computer-generated scenes. We present three of them below.

In the first example the image was generated by a computer. The aim was to evaluate the degree of robustness of the proposed method. The computer-generated image corresponded to a spherical surface against a completely black background. The matrix **M** [see Eq. (2.2)] that specifies interference among illumination, surface color, and spectral sensors was unit 1. In this case coordinates of normals are equal to coordinates of corresponding response vectors, and the matrix of quadratic form **Q** is unit 1 also. We calculated response vectors without testing condition (2.4), which made the whole visible hemisphere a single segment. In some pixels, response vectors had negative components, which of course is not possible in real images. Then we added to the image random noise (normally distributed three-dimensional vectors). The response vectors of this image at the nodes of a square grid (16 × 16) were processed by the segmentation algorithm, and, after the region was identified, the shape of the surface within it was reconstructed.

The results of the reconstruction are shown in Fig. 3. Figure 3(a) corresponds to zero noise level, and Figs. 3(b) and 3(c) illustrate those changes in the computed shape that appeared with increasing noise level (the unit for the quadratic mean of the noise is one side of a grid cell). It can be seen that results of the shape reconstruction are satisfactory even for a rather high level of noise, as shown in Table 1. Surprisingly, we found that the major limitation is imposed by the segmentation part of the algorithm. It cannot identify a large region satisfactorily when the noise level is very high. Nevertheless, the segmentation seems to be good enough up to a noise level of $\sigma = 0.4$.

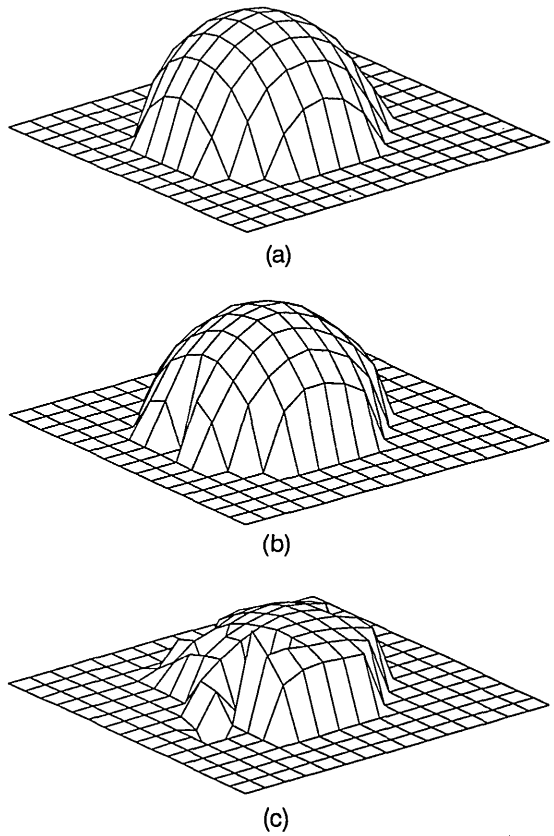
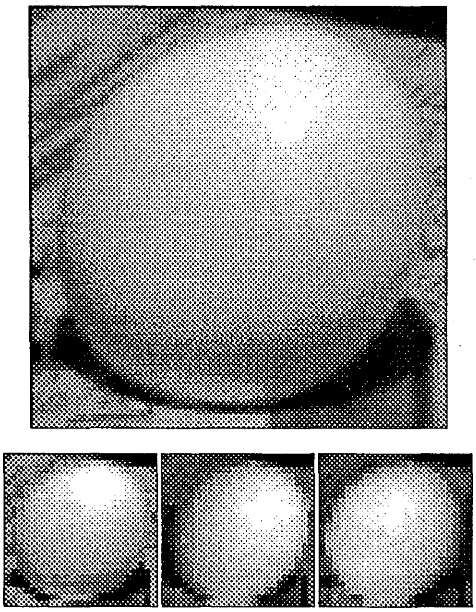


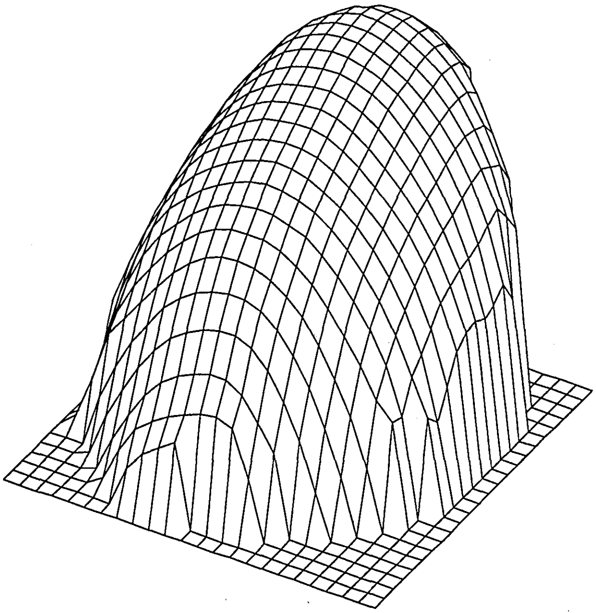
Fig. 3. Reconstructed shape of a spherical surface from the computer-generated image for different levels of noise: (a) $\sigma = 0$, (b) $\sigma = 0.2$, (c) $\sigma = 0.4$.

Table 1. Results of Shape Reconstruction

	Level of Noise			
Mean error of unit color signal	0	0.1	0.2	0.4
Mean error for depth estimate	0.12	0.11	0.14	0.25



(a)

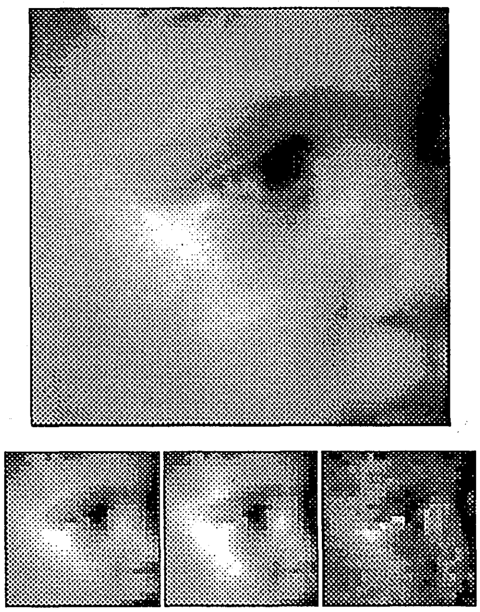


(b)

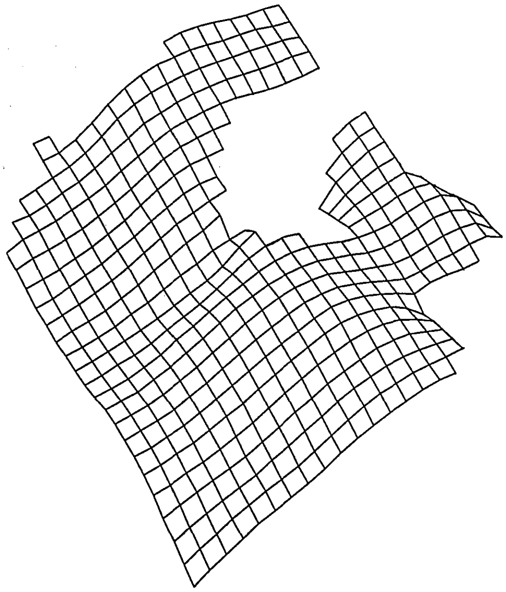
Fig. 4. (a) Top, high-resolution image of an egg in the red channel and bottom, three images for all channels with the low resolution used in the processing; (b) reconstructed shape of the egg (viewed from the side).

In the second example we processed the image of a real egg. The image was obtained by a camcorder (128 levels in each chromatic channel). The egg was illuminated by three light sources with different chromaticity (orange, green, and blue) that were placed close to the camcorder. In Fig. 4(a) we show (top) the high-resolution image in the red channel to give the reader an impression of the object and (bottom) the images from each channel with the low resolution (24 × 24) that we used in our computations. The result of the shape reconstruction is presented in Fig. 4(b), which shows the narrow end of a slightly tilted egg. The orientation of the reconstruction is close to the actual orientation.

In the last example we demonstrate that the constraints on mattedness and uniformity of color are really not too restrictive. We processed a photograph of a child's face in a way similar to that of the previous example. Figure 5(a) (top) represents the real high-resolution image in the red



(a)



(b)

Fig. 5. Partial image of a child's face. The details are the same as for Fig. 4.

channel, and (bottom) the three images correspond to the actual images in each channel that were used as an input. The reconstructed shape of the face is shown in Fig. 5(b). The face appears tilted by approximately 30 deg: the camcorder was tilted downward at this angle. The large hole in the center of the surface corresponds to the brow and the eye: these points were dismissed by the segmentation algorithm. Close examination of the reconstructed surface shows that the reconstructed shape is close to the prototype.

7. DISCUSSION

The shape-from-shading reconstruction algorithms currently are not widely used in computer-vision applications. There are several reasons for this:

1. These algorithms are complex for analysis and computations;
2. These algorithms are applicable to a restricted class of surfaces, and it is not easy to decide whether some part of the image fits the algorithm requirements;
3. There are other algorithms based on motion and stereo parallax that allow one to reconstruct the three-dimensional structure faster and more precisely.

Our method makes the utilization of shading information more feasible, because the method is conceptually and computationally simple and because the regions of an image where it is applicable can be computed. There is an important limitation, however, to its applicability: the rank of the region must be equal to 3; whereas in reality, regions of ranks 1 and 2 are also typical. To be competitive with motion and stereo methodology, the methods for all possible ranks should be integrated. This is impossible at the moment, because there is no solution for the rank-2 case.

Nevertheless, a simple modification of our method may make it competitive with other methods for practical applications. The combination of three chromatic light sources and a color camera can be simulated by a one-channel camera with illumination that varies in time. Assume that there are three spatially separated light sources that sequentially (one by one) illuminate a surface to be analyzed. The light sources can be stationary or attached to the camera and can have the same or different spectra. The one-channel camera sequentially collects three images with different illumination. These images comprise three-dimensional response vectors that can be treated in the same way as response vectors in trichromatic images.

8. CONCLUSION

In this study we presented theoretical solutions for two related problems. The first problem was a segmentation for rank-3 regions that was left unsolved in previous studies. Now the shape-independent segmentation of color images is possible for an arbitrary number of light sources. The second problem that we solved was depth reconstruction for rank-3 regions. We showed that the algorithmic solution for this case is much simpler than traditional algorithms for a single light source. In computer simulations we showed that our method has good precision and tolerance of noise.

ACKNOWLEDGMENTS

The authors thank A. M. Bonch-Osmolovsky, V. V. Maximov, G. I. Rozhkova, and L. Hardage for helpful comments on the draft and J. Malik and J. Weber for assistance with image input.

REFERENCES

1. B. K. P. Horn, "Obtaining shape from shading information," in *The Psychology of Computer Vision*, P. H. Winston, ed. (McGraw-Hill, New York, 1975), pp. 115–155.
2. B. A. Wandell, "The synthesis and analysis of color images," IEEE Pattern Anal. Mach. Intell. **PAMI-A9,** 2–13 (1987).
3. B. K. P. Horn and M. J. Brooks, "The variational approach to shape from shading," Comput. Vis. Graphics Image Process. **33,** 174–208 (1986).
4. K. Ikeuchi and B. K. P. Horn, "Numerical shape from shading and occluding boundaries," Artif. Intell. **17,** 141–184 (1981).
5. P. P. Nikolaev, "Some algorithms for surface color recognition," in *Simulation of Learning and Behavior*, M. S. Smirnov, ed. (Nauka, Moscow, 1975), pp. 121–151 (in Russian).
6. P. P. Nikolaev, "Monocular color discrimination of nonplanar objects under various illumination conditions," Biofizika **33,** 140–144 (1988) (in Russian).
7. M. H. Brill, "Image segmentation by object color: a unifying framework and connection to color constancy," J. Opt. Soc. Am. A **7,** 2041–2047 (1990).
8. A. P. Petrov, "Light, color, and shape," in *Intellectual Processes and Their Simulation*, E. P. Velikhov, ed. (Nauka, Moscow, 1987), pp. 350–358 (in Russian).
9. L. L. Kontsevich, "Computation of surface shape for uniformly colored and illuminated-by-chromatic-light-sources convex object from its two-dimensional projection," in *Data Processing in Information Systems* (Academy of Sciences of the USSR, Moscow, 1986), pp. 16–19 (in Russian).