Pergamon

Pattern Recognition, Vol. 28, No. 7, pp. 1033–1043, 1995
Elsevier Science Ltd
Pattern Recognition Society
Printed in Great Britain.
0031-3203/95 $9.50 + .00

0031-3203(94)00183-9

INTEGRATION OF SHAPE FROM SHADING AND STEREO*

JAMES EDWIN CRYER, PING-SING TSAI and MUBARAK SHAH†
Computer Science Department, University of Central Florida, Orlando, FL 32816, U.S.A.

(*Received* 4 *August* 1994; *in revised form* 13 *December* 1994; *received for publication* 3 *January* 1995)

Abstract—Stereo algorithms suffer from the lack of local surface texture due to smoothness of depth constraint, or local miss-matches in disparity estimates. Thus, most stereo methods only provide a coarse depth map which can be associated with a low pass image of the depth map. On the other hand, shape from shading algorithms generally produce better estimates of local surface areas, but some of them have problems with variable albedo and spherical surfaces. Thus, shape from shading methods produce better detailed depth information, and can be associated with the high pass image of the depth map. In order to compute a better depth map, we present a method for integrating the high frequncy information from the shape from shading and the low frequency information from stereo. The proposed algorithm is very simple, takes about 0.7 s for a 128 × 128 image on a Sun SparcStation-1, is non-iterative, and requires very little adjustment of parameters. The results obtained with a variety of synthetic and real images are discussed. The quality of depth obtained by integrating shading and stereo is compared with the ground truth (range image) using height error measure, and improvement ranging from 30 to 50% over stereo, and from 65 to 98% over shading is demonstrated.

Shape from shading Shape from stereo Integration of visual modules
Human Visual System

1. INTRODUCTION

Modern Computer Vision research follows the Marr paradigm[1] that treats vision as a large, complex information processing systems. Individual perceptual modules can be identified in the system which are responsible for the computation of shape from shading, stereo, motion, texture and contour, as well as processes for determining the location and nature of illumination sources, three dimensional motion, and other methods. During the last two decades there has been significant interest in these individual modules, which are termed *shape from* X. Interesting results have been reported, in particular, in motion, stereo and shading. Marr envisioned that the output from the individual modules will ultimately be integrated into a single representation called 2.5 *D sketch*. However, this integration was never accomplished by Marr. Vision inherently is an ill-posed problem and the solutions for *shape from X* obtained by considering each module individually may not necessarily *exist*, may not be *unique* and may not be *stable*. Therefore, in order to tackle these problems we need more information. In particular, if we combine information from different image cues like stereo, shading and motion, the solution may be significantly improved. Surprisingly, it is only recently that researchers in Computer Vision have started realizing the benefits of integrating information from separate modules. These are the work of Horn[2] on combining shading with contour, Grimson's[3] use of shading in determining the surface orientation of feature-point contours obtained from stereo, Aloimonos's methods[4] for combining shading and motion, texture and motion, and motion and contour, and Waxman's approach for combining stereo and motion.[5]

The objective of this research is to work on the integration of *shape from* X modules. In particular, we are interested in combining the depth information (3D shape) from two very important cues, stereo and shading. Shape from shading is the estimation of 3D shape from a 2D image given the light source and surface reflectance information. Stereo is the estimation of 3D shape from two images taken by cameras which are slightly shifted along the axis which the two cameras are aligned on (usually the x axis). The image of the object in the left stereo image is shifted with respect to the image in the right stereo image. This shift, which is also called the disparity, is inversely proportional to the 3D distance (depth) from the camera to the object. Frankot and Chellappa[6] pointed out that correspondence between stereo image pairs provides low frequency information not available in shading alone, and shading provides information not available from either sparse or low resolution stereo correspondences. Pentland[7] also suggested that linear shape from shading would be most useful in conjunction with some other depth cue, such as stereo, which could reliably provide the coarse, low frequency structure of the

* The research reported here was supported by the National Science Foundation under grant number CDA 9200369.
† Author to whom correspondence should be addressed.

scene. He further suggested that we could produce a better shape estimate by "blending" together the shading and stereo information in the frequency domain, giving the most weighted to the stereo information in the low spatial frequencies and the most weight to the shading information in the high spatial frequencies. So our criterion for combining shape from shading and stereo is very simple. We want to keep and amplify the low frequency information from the stereo, and add it with the amplified high frequency information from the shape from shading results.

The organization of the rest of the paper is as follows. The next section deals with the problems in the shape from stereo and shading. We briefly describe stereo and shading and summarize Barnard's stereo algorithm and Pentland's shading algorithm which will be used in this work. In section three, we will discuss Hall and Hall's filter, which will be used for integration of stereo and shading. Section four is the main thrust of the paper, where we describe our method for integration of stereo and shading. Section five deals with related work, and the comparison of our proposed method with the previous methods. Finally, the results for synthetic and real images are presented in Section 6.

2. SHAPE FROM X

The human visual system responds to light reflecting from objects. The visual system is able to determine depth from monocular scenes such as 2D images from smooth changes along the surface of the object, from former knowledge of size relationships of objects in the image, from occlusion, from size and from texture gradients. In this section we will focus on the shape from shading and stereo.

2.1. *Shape from shading*

Shape from shading deals with the recovery of 3D shape from a *single* monocular image which is only one cue that humans use to determine shape. There are two main classes of algorithms for computing shape from a shaded image: global methods and local methods. The global methods, in general, are very complex and slow. Sometimes they (e.g., variational calculus methods) require more than thousands of iterations to converge. A representative method of global approaches is used by Horn.[8] The local methods,[9,10] on the other hand, are simple, fast and give accurate local details within each homogeneous area, but are not accurate enough globally.

In shape from shading algorithms it is assumed that the surface reflectance map is given, or its form is known. However, surfaces of most objects in the real world have mixed reflectance forms, such as Lambertian with Specular, or Lambertian with various albedo values. We will not be able to recover the accurate 3D shape information with the shape from shading method alone.

Pentland[9] proposed a local algorithm based on the linearity of the reflectance map in the surface gradient (p, q), which greatly simplifies the shape from shading problem, and is very suitable for our purpose. The reflectance function for the Lambertian surfaces is modeled as follows:

$$E(x, y) = R(p, q) \tag{1}$$

$$= \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}} \tag{2}$$

$$= \frac{\cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma}{\sqrt{1 + p^2 + q^2}} \tag{3}$$

where $E(x, y)$ is the gray level at pixel (x, y), Z is the depth map, $p = (\partial Z/\partial x)$, $q = (\partial Z/\partial y)$, $p_s = (\cos\tau\sin\sigma/\cos\sigma)$, $q_s = (\sin\tau\sin\sigma/\cos\sigma)$, τ is the tilt of the illuminant and σ is the slant of the illuminant. By taking the Taylor series expansion of the reflectance function, equation (1), about $p = p_0, q = q_0$, up through the first order terms, we have

$$E(x, y) = R(p_0, q_0) + (p - p_0)\frac{\partial R}{\partial p}(p_0, q_0) + (p - q_0)\frac{\partial R}{\partial q}(p_0, q_0). \tag{4}$$

For Lambertian reflectance [equation (3)], the above equation at $p_0 = q_0 = 0$, reduces to

$$E(x, y) = \cos\sigma + p\cos\tau\sin\sigma + q\sin\tau\sin\sigma.$$

Next, Pentland takes the Fourier transform of both sides of this equation. Since the first term on the right is a DC term, it can be dropped. Using the identities:

$$\frac{\partial}{\partial x} Z(x, y) \leftrightarrow F_Z(\omega_1, \omega_2)(-i\omega_1) \tag{5}$$

$$\frac{\partial}{\partial y} Z(x, y) \leftrightarrow F_Z(\omega_1, \omega_2)(i\omega_2), \tag{6}$$

where F_Z is the Fourier transform of $Z(x, z)$, we get,

$$F_E = F_Z(\omega_1, \omega_2)(-i\omega_1)\cos\tau\sin\sigma + F_Z(\omega_1, \omega_2)(-i\omega_2)\sin\tau\sin\sigma,$$

where F_E is the Fourier transform of the image $E(x, y)$. The depth map $Z(x, y)$ can be computed by rearranging the terms in the above equation, and then taking the inverse Fourier transform. Other shape from shading algorithms use more complex models such as inter-reflections,[11] changing albedo and specular reflectance.[12] These methods are more difficult and require more computational time to solve.

2.2. *Shape from stereo*

Stereo vision is present in the human visual system and does not depend on the complex cues like occlusion, shadows, texture gradients and size of objects used by a monocular visual system. Shape from stereo uses two images from which the depth map can be calculated.

There are three main classes of algorithms for computing shape from stereo pairs: feature-based approaches, area-based approaches, and miscellaneous approaches. In the feature-based methods, the depth is computed at feature locations (mostly edges); in this case it is necessary to perform interpolation between features in order to get the dense depth maps. In the area-based approaches depth is computed for each pixel. The area-based methods have difficulties with the areas of nearly homogeneous image intensity which lack spatial structure. There are several miscellaneous approaches, such as Barnard's stochastic stereo algorithm,[13] which can run very fast on a suitable parallel machine, but the computed depth map suffers from the lack of local surface texture. In general, most stereo methods only provide coarse depth maps, and the fine details are missing in these depth maps.

In Barnard's stereo approach the problem is to find an assignment of disparities, $D(i, j)$, such that two criteria, *similar intensity* and *smoothness*, are satisfied.:

$$E = \sum_{j=1}^{n} \sum_{i=1}^{n} \| I_L(i, j) - I_R(i, j + D(i, j)) \|^2 + \lambda \| \nabla D(i, j) \|^2, \quad (7)$$

where I_L and I_R are the left and right images, $D(i, j)$ is the disparity map, the ∇ operator computes the sum of the absolute differences between disparity $D(i, j)$ and its eight neighbors, and λ is a constant. For a 128×128 image, and a disparity range of 10 pixels, there are 10^{16384} possible disparity assignments, which results in combinatorial explosion. Barnard uses a simulated annealing to solve this problem. The algorithm is as follows:

(1) Select a random state S.
(2) Select high temperature T.
(3) While $T > 0$.
  (a) Select S'
    $\Delta E \leftarrow E(S') - E(S)$.
  (b) if $\Delta E \leq 0$ then $S \leftarrow S'$
  (c) else $P \leftarrow \exp -\frac{\Delta E}{T}$, $X \leftarrow rand(0, 1)$.
    if $X < P$ then $S \leftarrow S'$
  (d) if no decrease in E for several iterations then lower T.

3. HALL AND HALL'S FILTER

Hall and Hall[14] describe a model to simulate the visual properties of the human eye by combining the low and high frequency information (Fig. 1). The first box is a low pass filter which is derived from a lens of diameter 3 mm. The second box is the log operation performed by the retina, and the third box is the high pass filter derived from the neuron model. The low and high pass filters (as shown in Fig. 2) in Hall and Hall's

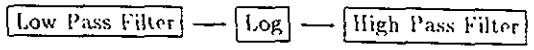
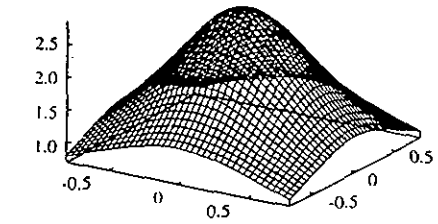
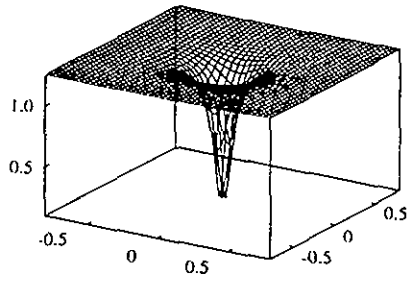


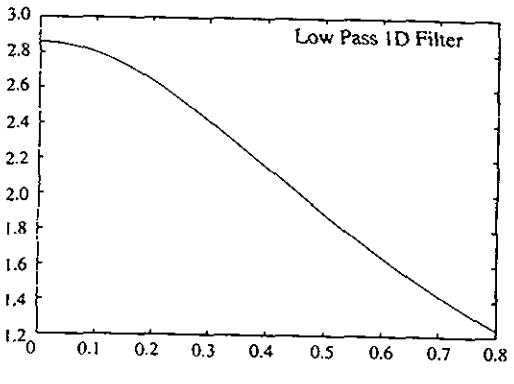
Fig. 1. Hall and Hall's model for human visual system.



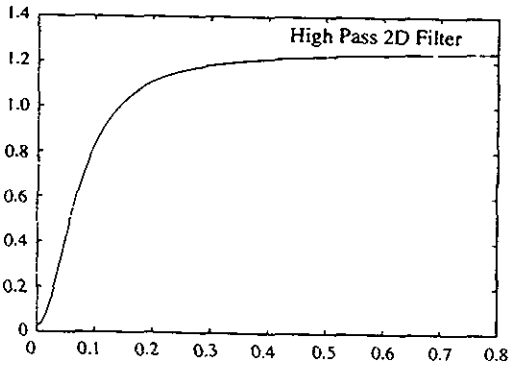
2D Low Pass



2D High Pass



1D Low Pass



1D High Pass

Fig. 2. Hall and Hall's low and high pass filters.

model of the Human Visual System which work in the frequency domain are shown as follows:

$$Low(\omega) = \frac{2\alpha}{\alpha^2 + \omega^2} \tag{8}$$

and

$$High(\omega) = \frac{a^2 + \omega^2}{2a_0 a + (1 - a_0)(a^2 + \omega^2)}, \tag{9}$$

where $\omega = \sqrt{u^2 + v^2}$, u and v represent the two dimensional frequencies in the Fourier domain. The term α is the spatial angular frequency. A typical value of α is 0.7 for a 3 mm diameter of the iris opening. The term a_0 represents the distance factor, which is the amount of change between the low and high frequencies. The other term a represents the strength factor, which is the rate of the cutoff point change between the low and high frequencies. For the human visual system the normal values for a_0 and a are, respectively 0.2 and 0.01.

Originally, Hall and Hall's filters were used in modeling the human visual system for perception using intensity images. However, these filters use the high/low frequency emphasis technique, which is very suitable for our purpose, and can still be used in filtering the stereo and shading depth maps.

4. INTEGRATION OF SHAPE FROM SHADING AND STEREO

The stereo methods provide the coarse shape information, and the shape from shading methods provide the detailed feature information. In the frequency domain the low spatial frequencies are related to the coarse shape information, and the high spatial frequencies are related to the details of the shape. Therefore, our criterion for combining shape from shading and stereo is very simple: *Keep the low frequency information from stereo, and add with the high frequency information from the shape from shading.*

Intuitively, one can use a high pass filter to separate the high frequency information from the shading result, and a low pass filter to separate the low frequency information form the stereo result. Then one can combine the low frequency information from stereo with the high frequency information from shading. However, choosing a good filter and the proper cut-off points is not an easy job. We have experimented with some filters,[15] but have not obtained significant improvement in the depth map. For example, the ideal filter produced very bad results due to the sharp cut-off point. We also applied the high pass Butterworth filter to the shading depth map and the low pass Butterworth filter to the stereo depth map, and combined these filtered depth maps. However, the resultant depth map was not much better than stereo or shading alone.

We will use Hall and Hall's filter, as described in the previous section, to combine the stereo and shading depth maps. The high pass filter designed by Hall and Hall attenuates the low frequency information, and emphasizes the high frequency information. Therefore, by inverting Hall and Hall's high pass filter (it will serve as a low pass filter now) we can get the low frequency information from an estimated depth map. This process is done to the estimated depth map from both stereo and shading algorithms. Next we compute the high frequency information of the shading results by subtracting the low frequency information from the Fourier Transform of the original depth map. Finally, the low frequency information from stereo and the high frequency information from shading are combined in the frequency domain to produce a better depth map. The flowchart of our method is shown in Fig. 3.

Mathematically, the proposed method can be summarized as follows:

$$F_{Z_C}(\omega) = F_{Z_{ST}} \times High(\omega)^{-1} + F_{Z_{SS}} \times (1 - High(\omega)^{-1}),$$

where $High(\omega)^{-1}$ is the inverse of Hall and Hall's high pass filter (equation 9), and F_{Z_C}, $F_{Z_{ST}}$ and $F_{Z_{SS}}$, respectively are the Fourier Transforms of the combined depth map, stereo depth map and shading depth map. The combined depth, Z_C, is computed by taking the inverse Fourier Transform of the above equation.

The proposed method is very simple, and computationally inexpensive. Once the stereo and shading depth maps are available, the integration takes about 0.7 CPU s

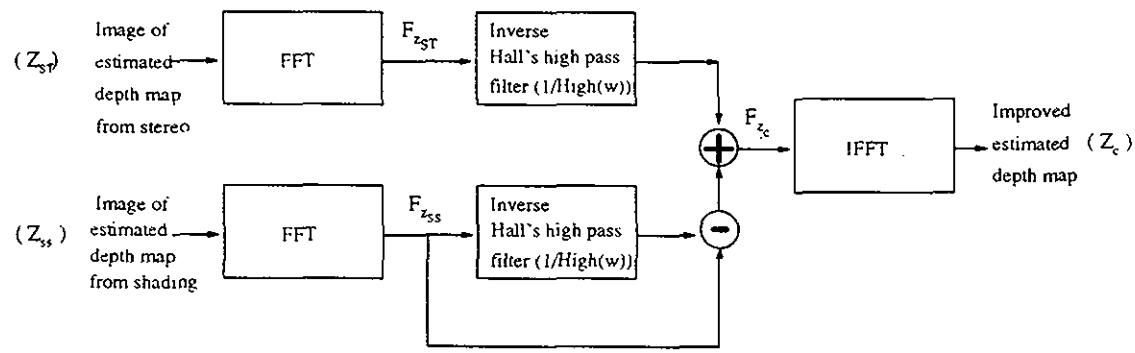


Fig. 3. Flowchart of proposed method for combining the stereo and shading.

for a 128 × 128 image on a Sun SparcStation-1. The major computation occurs with the Fast Fourier Transform (FFT), and Inverse Fourier Transform (IFFT), which is known to be of order $N \log N$ for N data points. Since we are using Pentland's shading algorithm which employs FFT, the Fourier Transform of the shading depth map is already available. Additional computation only involves the FFT of the stereo depth map. The computation for integration of stereo and shading consists only of the application of the filter, which is of order n^2 for a $n \times n$ image. The proposed algorithm requires very little adjustment of parameters. In our experiment, all the parameters of the filter are fixed. The distance factor, a_0, is set to 0.2, and the strength factor, a, is set to 0.05. One may need to change the strength factor, which is the rate of the cutoff point change between the low and high frequencies, if the input depth maps have more error in the low or high frequency information.

5. DISCUSSION

There are several other possibilities for integrating stereo and shading. The stereo depth map can be used to improve the shape from shading algorithm. For instance, in Ikeuchi-Horn's shape from shading algorithm[2] it is assumed that the surface orientation at the occluding contours is available. Their method iteratively computes the surface orientation at the remaining locations by propagating the surface orientation at the occluding contours. The contour depth map computed by the feature-based stereo can be used for this purpose. In fact, Blake *et al.*[16] have shown that if the boundary information (depth) is available at the occluding contours then shape from shading convergences to a unique solution.

In a recent shape from shading algorithm reported by Leclerc and Bobick,[17] which uses the conjugate gradient method for minimizing the cost function, the depth is iteratively refined. Leclerc and Bobick assume that a good initial guess for depth at each pixel is available. In their case, the dense depth map computed by the area-based (correlation-based) stereo method was used to obtain a good initial estimate. The difference between their and our method is that they use shape from shading to improve the result of stereo matching. They do not directly combine the results of shape from shading and stereo matching. A drawback of their method is that the two modules, stereo and shading, are not independent. Their shape from shading result is highly dependent on the stereo matching result.

The shape from shading can also be used to improve the depth map computed by stereo. For instance, Grimson[3] uses shading in determining the surface orientation of feature-point contours obtained from stereo.

Frankot and Chellappa[6] presented an elegant approach for enforcing the integrability constraint in shape from shading. They compute the orthogonal projection onto a vector subspace spanning the set of integrable slopes. This projection maps closed convex sets into closed convex sets, and hence, is attractive as a constraint in iterative algorithms. The authors noted that the low frequency information is lost in the process of image formation and due to regularization penalty and periodic boundary conditions. They showed improvements of their shape from shading results by incorporating the low frequency information obtained from another source [like the Digital Terrain Model (DTM)].

Our approach for integrating stereo and shading is very different from the previous approaches. We assume that each module is working independently and in parallel, and can therefore be treated as a black box. In fact, our method for combining shading and stereo is not dependent upon any particular method for shading or stereo. In principle, any method can be used. We have used Pentland's shading method and Barnard's stereo method, because working implementations of those two methods are available in our lab.

6. EXPERIMENTS

6.1. *Results for synthetic data*

The proposed method was first tested on two synthetic images, Tomato and Mozart. The stereo images were created using the following formulas:

$$x' = \frac{(x-b)f}{f-z}$$

and

$$x'' = \frac{(x+b)f}{f-z} \tag{11}$$

where x' is the position in the left stereo pair, x'' is the position in the right stereo pair, $2b$ is the distance between two cameras, z is the depth value at the original x position in the image, and f is the focal length of the camera. This can be seen in Fig. 4. We first generate a gray level image using the Lambertian model, equation (2), based on the range data and the given light source direction. Then, for each pixel (x, y), we compute the corresponding positions in the left and right stereo pairs, (x', y) and (x'', y), using the above formulas, and set the intensity value $I(x, y) = I_L(x', y') = I_R(x'', y'')$.

In order to evaluate the performance of our algorithm, we need to choose some error measures. Horn[8] proposed a number of quality measures for displaying the algorithm's progress. Szeliski[18] chose four different measures to study their shape from shading algorithm. The first two measures, the cost (or energy) error and the magnitude of the residual error, are not suitable for us. The other two measures are the magnitude of the gradient error and the magnitude of the height error. Since we are fusing two depth maps, which are heights, we will use the height error instead of gradient error. The square root of the mean-squared error (RMS) with respect to the ground truth depth is com-

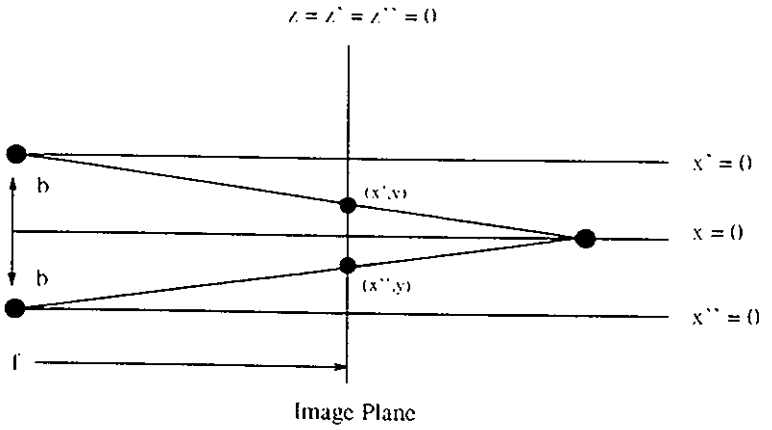


Fig. 4. The stereo camera imaging system.

puted using the following formula:

$$E = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}((Z(i,j) - \hat{Z}(i,j))^2}{n^2}} \qquad (12)$$

where $Z(i,j)$ is the actual ground truth depth, and $\hat{Z}(i,j)$ is the estimated depth, adn n^2 is the number of pixels on the object surface.

The results for the Tomato images are shown in Fig. 5. The gray level stereo pairs generated from the true depth map [shown in Fig. 5(a)] are shown in Fig. 5(b) and (c). The focal length and the distance between the two cameras were respectively assumed to be 400 and 60. The estimated depth map computed by Barnard's stereo algorithm is shown in Fig. 5(e). The average height square error is 0.46. This depth map is good. However, there are some noticeable errors in the depth map. For instance, the depth around outer portions of the tomato do not seem to be correct, and the surface patch around the upper left part is almost flat. The estimated depth map computed by Pentland's linear shape from shading algorithm (applied to the right stereo image) is shown in Fig. 5(f). Since the tomato is similar to a spherical object, and it is well known that the linear shape from shading method proposed by Pentland does not compute a good depth map for spherical surfaces,[19] the average height square error is about 1.85. The result obtained by integrating stereo and shading using our method is shown in Fig. 5(d), the average height square error reduces to 0.24. This is appproximately a 48% improvement over the stereo, and a 98% improvement over shading. We feel that achieving this great improvement by using a very simple algorithm is remarkable. Figure 5(g)–(i) shows the reconstructed gray level images using the estimated depth maps in (d)–(f) with the light source direction (0.01, 0.01, 1).

Next, we tested our method for the Mozart image. The results are shown in Fig. 6. The true depth map is shown in Fig. 6(a). The gray level stereo images generated from the true depth map are shown in Fig. 6(b) and (c). In this case also, the focal length and the distance between the two cameras were, respectively assumed to be 400 and 60. The estimated depth map computed by Barnard's stereo algorithm is shown in Fig. 6(e). The average height square error is 0.77. In this case, it is also obvious that the stereo does a poor job on details. The surface is not that smooth, and the surface patches around the nose and eyes are not correct. The estimated depth map computed by Pentland's linear shape from shading algorithm (applied to the right stereo image) is shown in Fig. 6(f). The average height square error is 1.5. Pentland's method does a very poor job on this image. It is almost impossible to perceive a face from this depth map. For instance, the areas corresponding to the center of the face have incorrect dips in the surface. The results obtained by integrating stereo and shading using our method are shown in Fig. 6(d). The average height square error in this case reduces to 0.55. There is about a 30% improvement over stereo, and a 63% improvement over shading. This depth map is much closer to the original range image. The detailed surface patches around the nose and eyes are noticeable. Figure 6(g)–(i) shows the reconstructed gray level images using the estimated depth maps in (d)–(f) with the light source direction (0.01, 0.01, 1). It is very interesting to note that, even though the shading depth map in Fig. 6(f) appears to be very poor, the reconstructed gray level image in Fig. 6(i) looks much better than the reconstructed gray level image from the stereo depth map [shown in Fig. 6(h)]. Some obvious problems around the nose (e.g., a line throughout the image) are noticeable in Fig. 6(i). The reconstructed gray level image [shown in Fig. 6(g)] is much closer to the gray level image [shown in Fig. 6(b)] generated using the true depth map.

6.2. *Results for real data*

The proposed method was also tested on two real stereo pairs. The results are shown in Figs 7 and 8.

The results for the Renault images are shown in Fig. 7. The stereo images are shown in Fig. 7(a)–(b), and

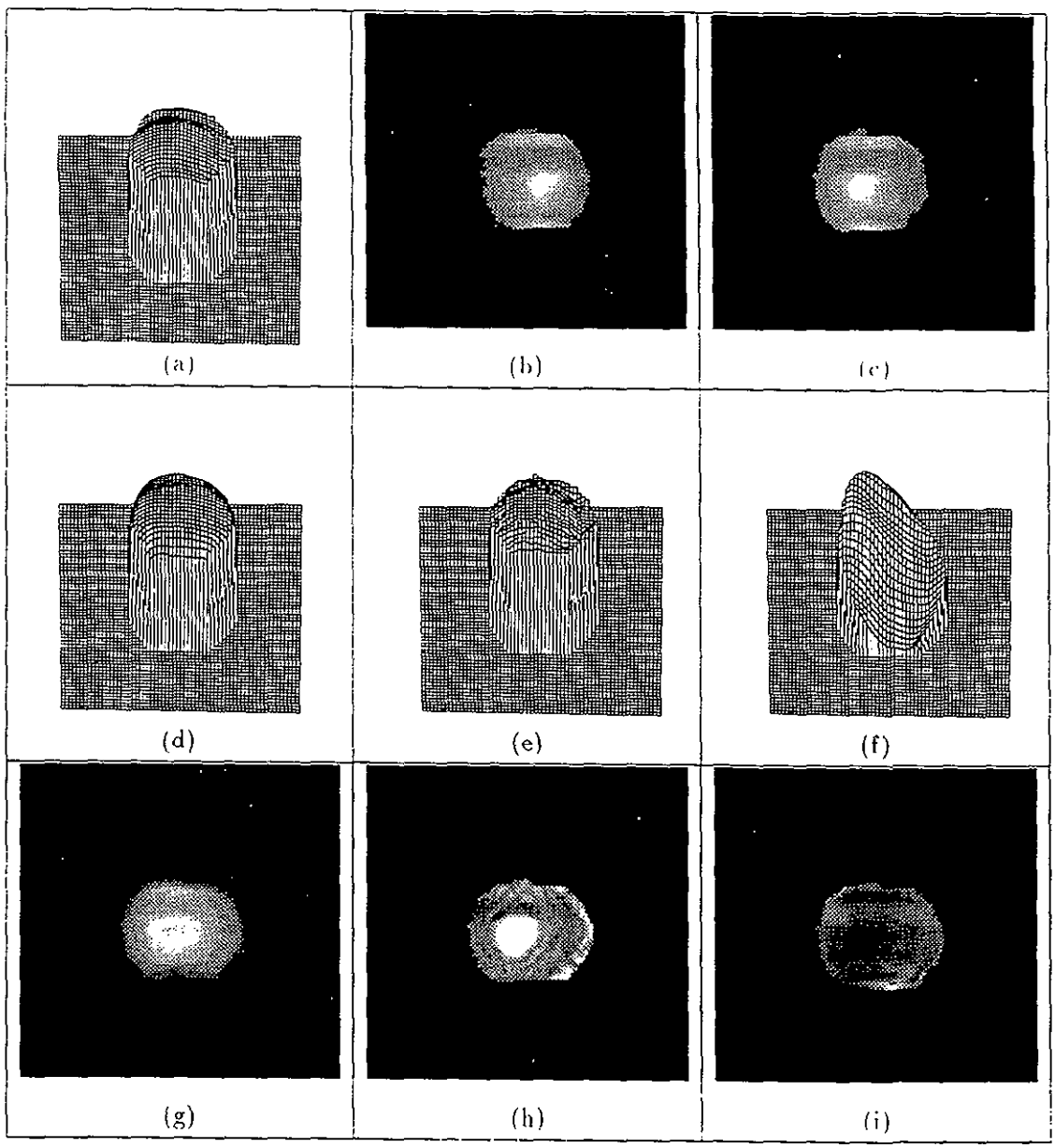


Fig. 5. Results for the Tomato images. (a) A 3D plot of the range data. (b) Left stereo gray level image. (c) Right stereo gray level image. (d) A 3D plot of the estimated depth map by our method. (e) A 3D plot of the estimated depth map by stereo algorithm. (f) A 3D plot of the estimated depth map by Pentland's shape from shading algorithm. (g) A reconstructed gray level image using the estimated depth map in (d). (h) A reconstructed gray level image using the estimated depth map in (e). (i) A reconstructed gray level image using the estimated depth map in (f).

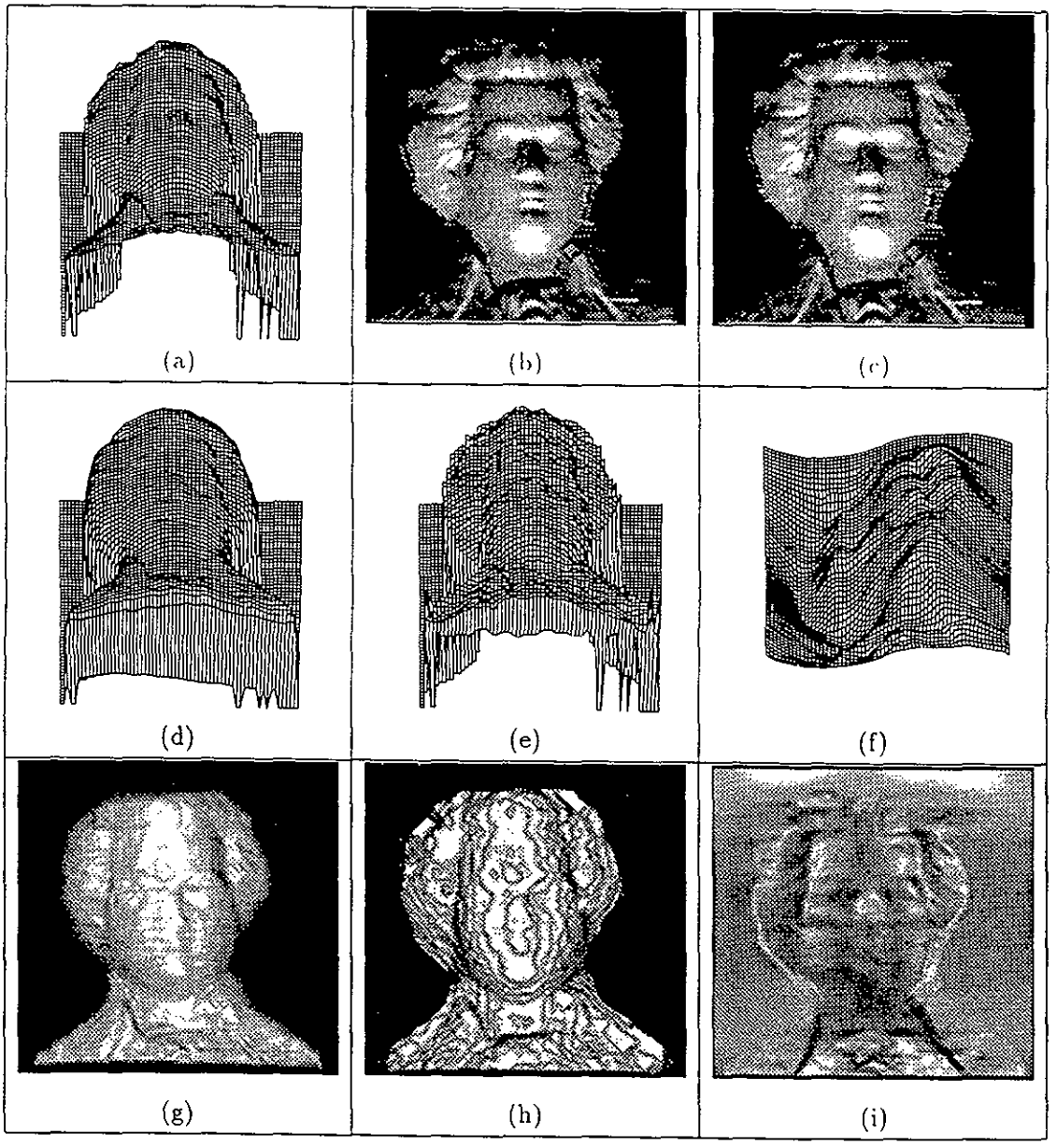


Fig. 6. Results for the Mozart images. (a) A 3D plot of the range data. (b) Left stereo gray level image. (c) Right stereo gray level image. (d) A 3D plot of the estimated depth map by our method. (e) A 3D plot of the estimated depth map by stereo algorithm. (f) A 3D plot of the estimated depth map by Pentland's shape from shading algorithm. (g) A reconstructed gray level image using the estimated depth map in (d). (h) A reconstructed gray level image using the estimated depth map in (e). (i) A reconstructed gray level image using the estimated depth map in (f).

the right stereo image is used for the shape from shading algorithm. The estimated depth map computed by the stereo algorithm is shown in Fig. 7(d), and the estimated depth map computed by the shape from shading algorithm is shown in Fig. 7(e). The obvious errors in details in the stereo results are noticeable. The object in this image does not have constant albedo, therefore Pentland's algorithm, which assumes constant albedo, encounters some problems. The results obtained by integrating stereo and shading using our method are shown in Fig. 7(c). This depth map is much better than the other two. The problems in surface details and problems due to variable albedo are almost eliminated. Figure 7(f)-(h) shows the reconstructed gray level images using the estimated depth maps in (c)-(e) with the estimated light source direction $(-0.62, 0.50, 0.60)$. (The light source direction was estimated using Pentland's improved method[7] for all the real images in this section.) The reconstructed gray level images using the depth map obtained by integrating

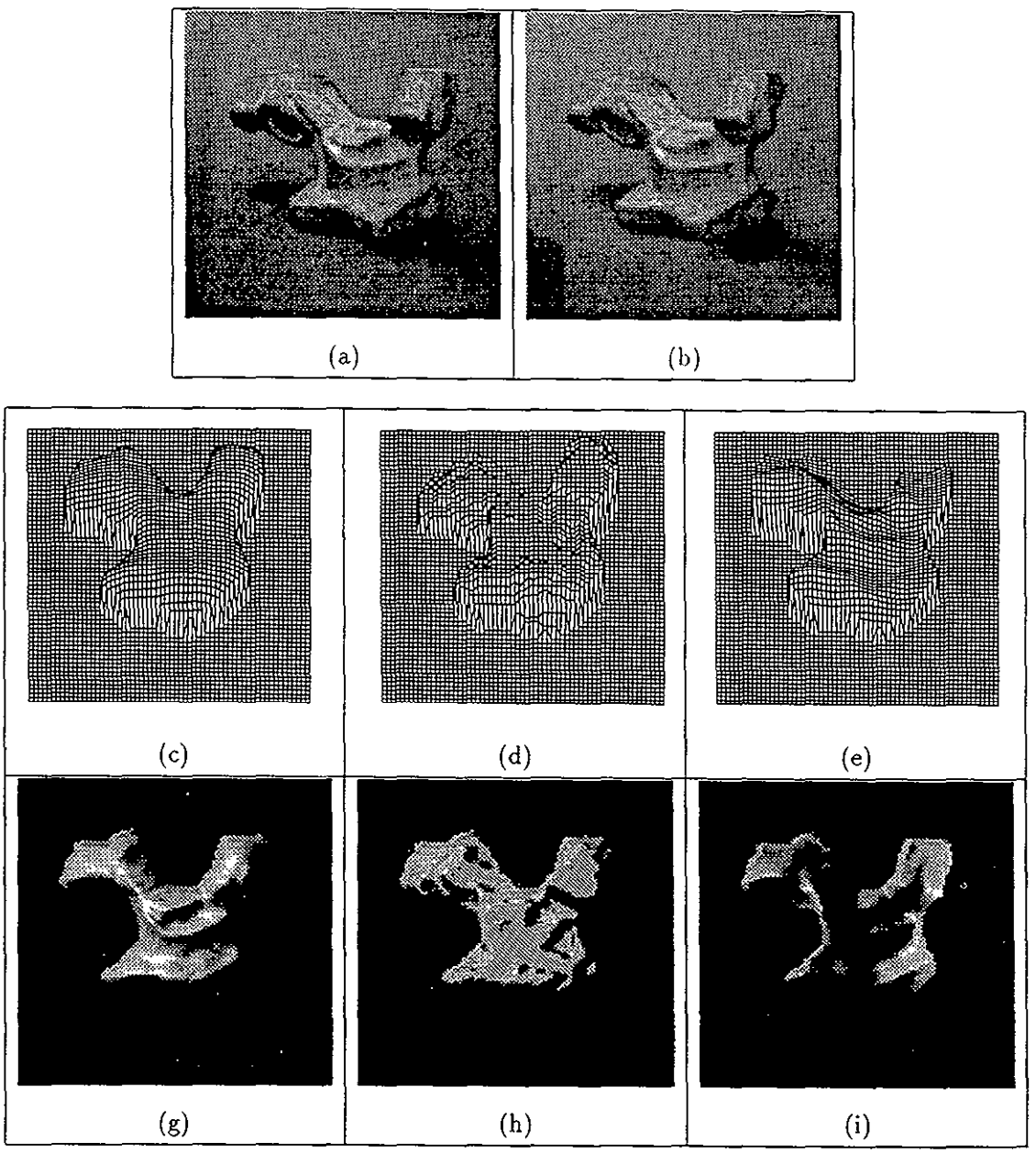


Fig. 7. Results for the Renault images. (a) Let stereo image. (b) Right stereo image. (c) A 3D plot of the estimated depth map by our method. (d) A 3D plot of the estimated depth map by stereo algorithm. (e) A 3D plot of the estimated depth map by Pentland's shape from shading algorithm. (f) A reconstructed gray level image using the estimated depth map in (c). (g) A reconstructed gray level image using the estimated depth map in (d). (h) A reconstructed gray level image using the estimated depth map in (e).

shape and stereo are much closer to the original gray level images.

Next, the results for Sandwich images are shown in Fig. 8. The stereo images are shown in Fig. 8(a)–(b), and the right stereo images is used for the shape from shading algorithm. The estimated depth map computed by the stereo algorithm is shown in Fig. 8(d), and the estimated depth map computed by the shading algorithm is shown in Fig. 8(e). The results obtained by integrating stereo and shading using our method are shown in Fig. 8(c). The stereo depth map has many problems with surface details in the Sandwich surface: instead of showing a flat planar surface it appears curved. The shading results are better than results for the stereo, but overall the sandwich surface does not appear planar due to changes in albedo. The integrated depth map is almost perfect, clearly showing one flat surface of the sandwich at a constant depth. Figure 8(f)–(h) shows the reconstructed gray level images using the estimated depth maps in (c)–(e) with the estimated

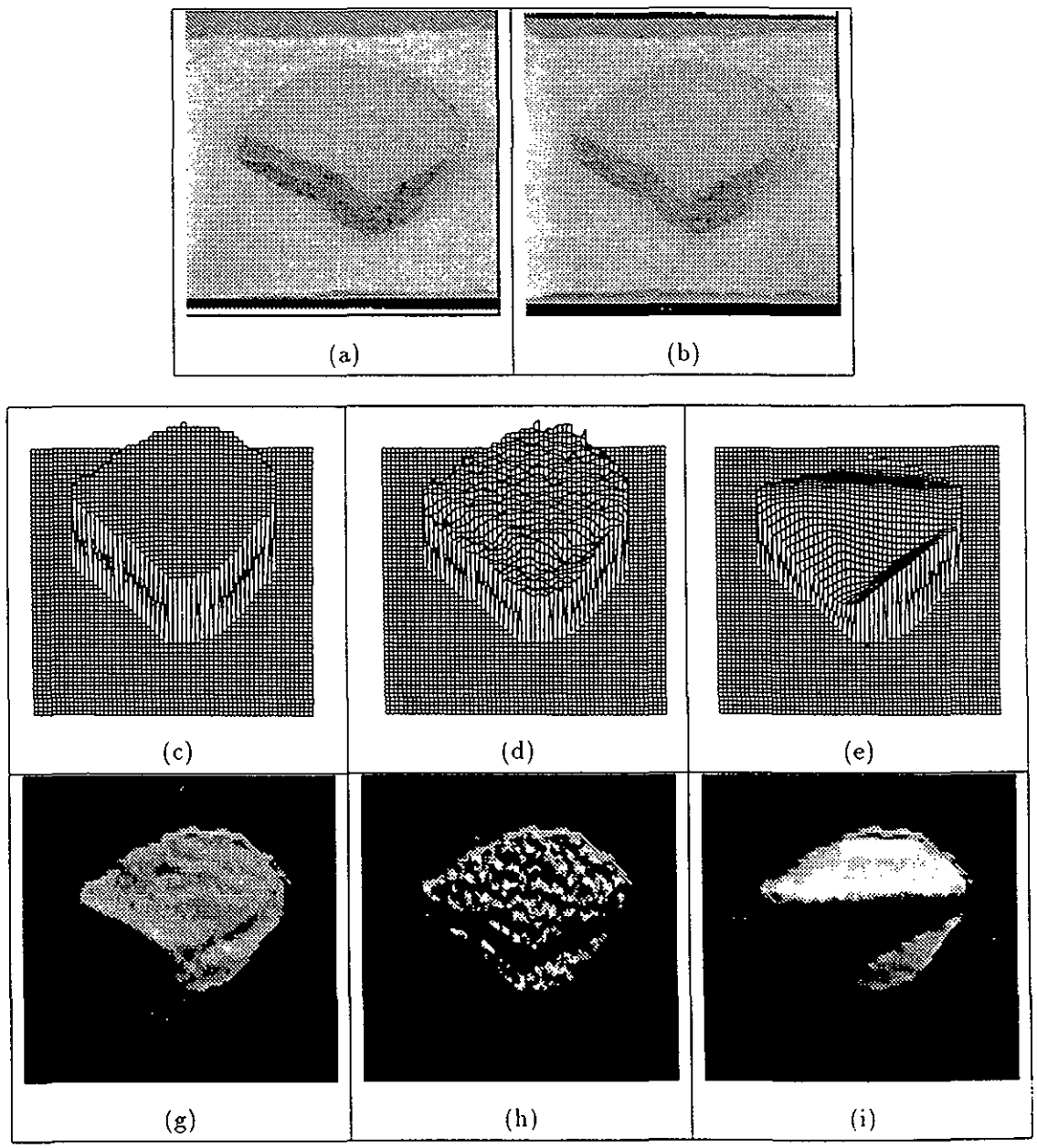


Fig. 8. Results for the Sandwich images. (a) Left stereo image. (b) Right stereo image. (c). A 3D plot of the estimated depth map by our method. (d) A 3D plot of the estimated depth map by stereo algorithm. (e) A 3D plot of the estimated depth map by Pentland's shape from shading algorithm. (f) A reconstructed gray level image using the estimated depth map in (c). (g) A reconstructed gray level image using the estimated depth map in (d). (h) A reconstructed gray level image using the estimated depth map in (e).

light source direction (0.15, 0.78, 0.61). The problems in shading and stereo depth maps are highlighted in the reconstructed gray level images. However, the reconstructed gray level image obtained by the integrated depth map is reasonable. Without the ground truth data for these real images, we cannot compute the error. However, we can clearly see the improvement from the 3D plots and the reconstructed gray level images.

7. CONCLUSIONS

In this paper we have addressed the problem of integration of *shape from X* modules. In particular, we have focused on the combination of shape from shading and stereo. Our approach is very simple. Our criteria for integration shading and stereo is: *keep the low frequency information from stereo, and add with the high frequency information from the shape from shading.*

Future work includes the integration of other shape from X modules, for example, motion, texture, and contour.

Acknowledgements—The authors are thankful to Professor Ikeuchi of Carnegie Mellon University for providing the *tomato* image, Dr Stein of University of Southern California for providing the *Mozart* image, and Professor Ramesh Jain of University of California, San Diego for providing *Sandwich, Pentagon and Renault* stereo pairs.

REFERENCES

1. D. Marr, *Vision*, Freeman, California (1982).
2. K. Ikeuchi and B. K. P. Horn, Numerical shape from shading and occluding boundaries, *Artificial Intell.* **17**, 141–184 (1981).
3. E. Grimson, Bincular shading and visual surface reconstruction, *CVGIP* 18–44, (1984).
4. J. Aloimonos and D. Shulman, *Integration of Visual Modules*, Academic Press (1989).
5. A. Waxman and J. Duncan, Binocular image flows, *IEEE Workshop Visual Motion* (1987).
6. R. T. Frankot and R. Chellappa, A method for enforcing integrability in shape from shading, *IEEE Trans. Pattern Analy. Mach. Intell.* **10**, 439–451 (1988).
7. A. Pentland, Shape information from shading: a theory about human perception, *2nd Int. Conf. Comput. Vision* 404–413, Tampa, Florida (December 1988).
8. B. K. P. Horn, Height and gradient from shading, *Int. J. Comput. Vision* **5**, 37–75 (1990).
9. A. Pentland, Linear shape from shading, *IJCV* **4**, 153–162 (1990).
10. C. H. Lee and A. Rosenfeld, Improved methods of estimating shape from shading using the light source coordinate system, *Artif. Intell.* **26**, 125–143 (1985).
11. K. Ikeuchi, S. K. Nayar and T. Kanade, Shape from interreflections, *2nd Int. Conf. Comput. Vision* 1–11 (1990).
12. G. Healey and T. O. Binford, Local shape from specularity, *CVGIP* **42**, 62–86 (1988).
13. S. T. Barnard, A stochastic approach to stereo vision, *Proc. 5th National Conf. AI* 676–680, Philadelphia, Pennsylvania (August 1986).
14. C. F. Halla and E. L. Hall, A nonlinear model for the spatial characteristics of the human visual system, *IEEE Trans. System, Man Cybern.* **6**, 161–170 (1977).
15. R. C. Gonzalez and P. Wintz, *Digital image processing*, Addison Wesley (1987).
16. A. Blake, A Zisserman and G. Knowles, Surface descriptions from stereo and shading, *Image Vision Comput.* **3**, 183–191 (1985).
17. Y. G. Leclerc and A. F. Bobick, The direct computation of height from shading, *Proc. Comput. Vision Pattern Recognition* 552–558 (1991).
18. R. Szelski, Fast shape from shading, *CVGIP Image Understanding* **53**, 129–153 (1991).
19. P. S. Tsai and M. Shah, A fast linear shape from shading, *Proc. Comput. Vision Pattern Recognition* 734–736 (1992).

About the Author—JAMES E. CRYER was born in Trenton, NJ. He received a Masters degree in computer science from the University of Central Florida in 1933 and is presently working as a program/research engineer for UCF civil engineering department.

About the Author—PING-SING TSAI was born in Taipei, Taiwan, R.O.C. He received the B.S. degree in information and computer engineering from Chung-Yuan Christian University in 1985. He is pursuing a Ph.D. in computer science at the University of Central Florida.

About the Author—MUBARAK SHAH received his B.E. degree in 1979 in Electronics from Dawood College of Engineering and Technology, Karachi, Pakistan with the highest grades in the whole University, and was awarded a five year Quad-e-Azam (Father of Nation) scholarship for his Ph.D. He spent 1980 at Philips International Institute of Technology, Eindhoven, The Netherlands, where he completed E.D.E. diploma. Dr Shah received his M.S. and Ph.D. degrees both in Computer Engineering from Wayne State University, Detroit, Michigan, respectively in 1982 and 1986. Since 1986 he has been with the University of Central Florida, where he is currently an Associate Professor of Computer Science. Prof. Shah has published his research in motion, edge and contour detection, multisensor fusion, and shape from shading and stereo. He is an associate editor of *Pattern Recognition* journal. He has served on the program committees of the *Machine Vision and Robotics Conference*, and the *ACM Computer Science Conference*. He has chaired sessions at several conferences, including the *International Conference on Pattern Recognition*, and the *SPIE Applications of AI Conference*.