Natural Photometric Stereo?

José R.A. Torreão

Computação Aplicada e Automação/TEE, Universidade Federal Fluminense
24.210-240 Niterói, RJ, Brasil
$jrat$@caa.uff.br

Abstract. The human brain seems to have the ability of inferring shape from the binocular fusion of some kinds of monocular images. Recently, we have observed that photometric stereo (PS) images, which are monocular images obtained under different illuminations, produce a vivid impression of depth, when viewed under a stereoscope. Lately, we have found that the same is true of pairs of images obtained in different spectral bands. Employing an optical-flow based photometric stereo algorithm on a type of "colour separated images", which have been so produced as to emulate the kinds of records generated by the photosensitive cells in the human retina, we have been able to obtain depth estimates from them. This has led us to speculate on the possibility that a process similar to PS could work on the human visual system. Here we present our claim for a natural photometric stereo process, invoking some physical and biological arguments, along with experimental results, that could support it.

1 Introduction

Photometric stereo (PS) images are monocular images obtained under different illuminations, and have been employed for the estimation of surface gradient in computer vision [8]. Recently, we have observed that this kind of images can also be matched to yield relevant depth cues, when viewed under a stereoscope [6]. Based on that, we have suggested a new approach to shape estimation from PS images, whereby the optical flow resulting from the displacement of the irradiance pattern over the imaged surface, due to the illumination change, is estimated and related to surface shape via a structure-from-motion approach. We accomplish this by modelling the displacement of the irradiance pattern over the surface as a rotation, allowing for a correction factor in terms of a position-dependent translation along the $z-$direction (optical axis direction); by estimating the values of the rotation parameters, we are thus able to obtain a depth reconstruction of the imaged surface in the absence of explicit information about its reflectance properties [5,7].

Lately, we have found that the ability of the human brain to match monocular images, as observed in the case of the photometric stereo pairs, also applies to images obtained in different spectral bands. Thus, monocular images captured through filters which pass different bands of wavelengths also seem to carry shape information apt to be extracted through binocular fusion. Testing our PS algorithm on this kind of images, we have, somewhat surprisingly, been able to obtain good shape estimates, and this has led us to consider the possibility that a mechanism similar to the artificial vision process of photometric stereo could operate in the human visual system. Underlying such natural photometric stereo, we would have the physical and biological principles of light scattering in the Earth's atmosphere, and light detection by the photosensitive receptors in the retina. In the next section, we describe both these principles, giving what would be the bases of our claim for the plausibility of a natural PS process. Then we discuss the results of some experiments with the use of our optical-flow based PS algorithm for shape estimation from "colour separated" images (images obtained in different spectral bands). Finally, as an appendix, we present the fundamentals of the PS algorithn employed.

2 The Bases for a Natural PS

The illumination of natural scenes as observed on Earth depends to a great extent on the diffusion of light in the atmosphere. If not for this phenomenon, the sun would appear as a bright spot on a completely dark background (as it looks when seen from the Moon, for instance), and the same would be true of any artificial light source. The multiple scattering of photons from the molecules and aerosols in the atmosphere is what diffuses the light, making it visible from all directions.

It has been known for sometime that the atmospheric scattering is highly dependent on the wavelength of the radiation, and that, as a result, the light falling on a given surface patch in a scene will have different spectral compositions, depending on the patch's orientation relative to the light source. It is this fact which explains, for instance, the blue colour of the sky and its reddish cast along the horizon at dusk. Figure 1, which is adapted from [1], shows plots of the computed radiance of the sky, for wavelengths at both ends of the visible spectrum ($\lambda = 0.4\mu$ and $\lambda = 0.7\mu$), as a function of the cosine of the zenith angle. Dashed curves are for the sun at zenith, and solid curves for the sun on the horizon ($\theta = 86.3^0$). The predominance of red light at low solar elevations (large zenith angles) is already apparent from the latter: for $\lambda = 0.7\mu$, the radiance is concentrated at large values of θ, being an order of magnitude stronger than that at $\lambda = 0.4\mu$, for $\theta = 90^0$. (In the figure, radiances are given for two values of the Earth albedo, $A = 0.8$ and $A = 0$, and have been averaged over all azimuth angles. The Earth's surface was assumed lambertian).

Here we will invoke such dependence of the atmospheric scattering on the wavelength of the incident radiation in support of our speculation about a natural photometric stereo. When light of a given spectral composition, coming from the sun or from an artificial source, falls on a surface, each spectral component will arrive mainly from a certain direction, according to its wavelength. Upon impinging on the surface, some of the spectral components will be absorbed, and some will be reflected, reaching the observer's eye and stimulating the photosensitive cells in its retina. As it is known, such cells may be grouped into four systems, each responding to a band of light frequencies: the three cone systems (see Figure 2) and the rod system.

The so-called blue cones respond to the entire lower-frequency half of the visible spectrum, with peak sensitivity at 0.44μ. The green cones and the red cones, with peak sensitivities at 0.535μ and 0.565μ, respectively, respond, each of them, to almost two-thirds of the visible spectrum. The rod system, on the other hand, with peak sensitivity at about 0.5μ, covers almost the entire spectral band, and responds to radiance levels about 1000 times weaker than those needed to activate the cone systems, thus being capable of functioning independently of them.

Each of the photosensitive systems produces, by itself, a monochromatic record of the scene - colour sensation arising from a comparison of the intensities which they report. In reference [4], in a context which will be explained in the following section, such systems have been called *retinexes*.

If we consider the images of a scene formed by the different retinex systems, we will be led to the conclusion that they can be considered as similar to a photometric stereo set: at a given point, each of them records the radiance emitted in a certain spectral band by the corresponding surface patch in the scene - such radiance arising from the illumination of the patch by light in that specific band, which, due to scattering, comes mainly from a certain spatial direction. Since the recorded spectral band is different for each retinex, so is the associated illumination, and thus, what is available to the observer's eye are a set of monocular images of the patch, obtained under different illuminations - that is to say, a set of PS images.

Given the underlying principles which would support our claim for a natural photometric stereo, we now turn to a discussion of some experiments which we have performed in order to test the validity of our hypothesis.

3 Experiments

In a series of remarkable experiments on colour vision [3,4], E. H. Land demonstrated that the perception of colour by humans is independent of the flux of radiant energy reaching the retina, which explains the fact that the human visual system is able to extract reliable and invariant colour information from scenes whose illumination may vary arbitrarily in spectral composition. Land proposed that the sensation of colour arises from the comparison of the lightness values recorded by the different systems of photoreceptors in the retina. Since he could not be certain where the transformation of lightness into colour would take place, he coined the term *retinex* (a combination of retina and cortex) to describe the overall process.

In his studies, Land produced what he called retinex records: black-and-white photographs taken with films whose spectral responses had been altered to match the spectral sensitivities of the different photoreceptors in the retina. What he did not observe then, as we recently did, is that pairs of such retinex records can be binocularly fused, when viewed under a stereoscope, to produce an unmistakable impression of depth (besides producing colour, if viewed through appropriate filters).

Such observation is what has led us to hypothesize the natural photometric stereo process. In order to test our theory, we have thus chosen to extract portions of Land's retinex images, shown in [4], applying to them our PS depth estimation algorithm. Since the sensitivity curves for the red and green cones are broadly overlapping, and both systems have approximately the same threshold response values, the corresponding retinex records seemed the most appropriate for our purposes, as we tried to avoid problems with overall intensity differences in the input images.

Figures 3 to 5, show the results which we have obtained in our experiments. Each figure depicts the pair of red and green retinexes used as input, and the depth estimates obtained from them (in arbitrary units). We have tried to extract portions of Land's original records which would carry the images of a single object at a time, but, except for Figure 5 (green pepper), this has not been possible. The retinexes in Figure 3 depict an orange and part of a strawberry, and those in Figure 4 depict a lemon and a radish. Since, in each of the last two cases, the vegetables shown appear with very different intensities, the fact that our algorithm has been able to approximate their correct shape from a single pair of images came as a bit of a surprise. Even more surprising is the fact that the algorithm seems to have been able to infer the correct relative positions of the objects: strawberry in front of orange, and radish in front of lemon.

4 Conclusions

We have put forth a claim for the plausibility of the operation, in the human visual system, of a shape estimation mechanism similar to the artificial vision process of photometric stereo. Our claim for such natural photometric stereo is based on the fact that human vision employs multispectral images of the observed scene, which are obtained through the four types of photosensitive cells present in the retina: the red, green and blue cones, and the rods. Functioning in an environment where the scattering of electromagnetic radiation results in light coming from different directions according to its wavelength, the multispectral systems in each eye would produce records of the scene which are monocular images obtained under different illuminations - that is to say, which are similar to PS images. In support of our claim, we have presented the results of experiments with the use of an optical-flow based photometric stereo algorithm for shape reconstruction from images which emulate the spectral records produced by the human visual system. Such results corroborate our observation that images obtained in different spectral bands can be binocularly fused to yield relevant depth information.

5 References

[1] Driscoll, W.G. and W. Vaughan, eds. *Handbook of Optics*, McGraw-Hill, New York, 1978.

[2] Horn, B.K.P. and B.G. Schunck, *Determining Optical Flow*, Artif. Intell., vol. 17, pp. 185-203, 1981.

[3] Land, E.H., *Experiments in Color Vision*, Scientific American, vol. 200(5), pp. 84-99, May 1959.

[4] Land, E.H., *The Retinex Theory of Color Vision*, Scientific American, vol. 237(6), pp. 108-128, December 1977.

[5] Torreão, J.R.A. and G.M. Mattos, *An Optical Flow Approach to Photometric Stereo*, Procs. VIII SIBGRAPI, pp. 213-218, São Carlos, SP, October 1995.

[6] Torreão, J.R.A., B.M. Carvalho, and G.M. Mattos, *Binocular Fusion of Photometric Stereo Images*, Procs. 2nd. Asian Conference on Computer Vision, vol. II, pp. 336-340, Singapore, December 1995.

[7] Torreão, J.R.A., *Estimating 3-D Shape from the Optical Flow of Photometric Stereo Images*, Technical Report 02-96, Computação Aplicada e Automação, UFF. Submitted for publication.

[8] Woodham, R.J., *Photometric Method for Determining Surface Orientation from Multiple Images*, Optical Engineering, vol. 19(1), pp. 139-144, 1980.

6 Appendix

Given a pair of photometric stereo images, it is possible to find the optical flow which results from the displacement of the irradiance pattern over the imaged surfaces, due to the change in illumination [2,5]. The PS optical flow, $D(s)$, carries information about the shape of the imaged surface, which can be recovered by a structure-from-motion scheme, if we make some assumption about the underlying movement. We model such movement as a rotation, allowing for a correction factor in the form of an arbitrary translation along the $z-$axis (optical axis direction); as it is known, under orthographic projection such a translation does not affect the optical flow.

Calling $\Theta = (A, B, C)$ the rotation vector, and $V(x, y)$ the translation along the $z-$direction, the equation of motion for the irradiance pattern becomes

$$\Delta R = \Theta \times R + V(x, y) \tag{1}$$

where $R = (x, y, z)$ is the position vector given in a coordinate system fixed with respect to the camera.

Since we are assuming that the irradiance pattern moves as if sliding over the imaged surface, we also impose the condition that the displacement vector, ΔR, should be, at each point, perpendicular to the local surface normal, i.e.,

$$\Delta R.\hat{n} = 0 \tag{2}$$

where

$$\hat{n} = \frac{(-p, -q, 1)}{\sqrt{p^2 + q^2 + 1}} \tag{3}$$

is the unit normal vector.

Now, from (1) and (2), we obtain the constraint

$$V(x, y) = D_X p + D_Y q + Bx - Ay \tag{4}$$

Returning to equation (1), it is easy to obtain the surface height function, in the form

$$z(x, y) = \frac{D_X x + D_Y y}{Bx - Ay} \tag{5}$$

which is independent of the rotational component C. (For $(x, y) = (0, 0)$, one should use $z(0, 0) = (D_X + D_Y)/(B - A)$). Also, employing (5) back in (1), we can find C as

$$C = \frac{AD_X + BD_Y}{Bx - Ay} \tag{6}$$

Thus, in principle, if we can obtain estimates of the rotational components, A and B, we will be able to recover the shape of the imaged surface. As it turns out, such approach has some shortcomings. First, the denominators in expressions (5) and (6) can be problematic for small values of x and y; second, and more basic, the inaccuracies in the estimation of the optical flow field, $D(s)$, could render the shape reconstruction through (5) unreliable. We have, therefore, found more appropriate to employ a least-squares strategy, which yields

$$z(x,y) = \frac{B(D_X + Cy) - A(D_Y - Cx)}{A^2 + B^2} \qquad (7)$$

giving the surface height function, at each point in the image plane, in terms of the measured disparity field and the three unknown rotation parameters, A, B and C, which must be estimated.

Now, for changes in the illumination direction along the xz or yz planes, it is reasonable to expect the rotation component C to be small, compared to A or B. We then proceed with the least-squares strategy, using $C = 0$, to find

$$A = \gamma B \qquad (8)$$

with

$$\gamma = \frac{\alpha \pm \sqrt{\alpha^2 + 4\beta^2}}{2\beta} \qquad (9)$$

where

$$\alpha = \int\int [D_X^2 - D_Y^2] dx dy \qquad (10)$$

and

$$\beta = \int\int [D_X D_Y] dx dy \qquad (11)$$

Since an independent estimate for B cannot be obtained via the least-squares approach, this parameter has to be properly chosen. For illumination directions on the xz plane, we have found that a value for B equal to the angle between the two illumination vectors is a plausible choice. (In the experiments with the retinexes, reported here, we always started our estimation with the arbitrary value $B \approx 0.17$ - i.e., B approximately equal to 10^0). For such B, and with A estimated from (8), we can then update our value for C through equation (6), appropriate care being taken when $Bx \approx Ay$.

The obtained estimates can be further refined: taking into account the calculated value for C, we can go back to equation (7) and repeat the least-squares estimation procedure for A. As for B, equation (4) allows for an update of its value, once an initial estimate for $z(x,y)$ has been obtained: through a least-squares approach we look for the B value which minimizes the translational displacement $V(x,y)$.

In our experiments with PS images, good reconstructions have been obtained for surfaces with lambertian plus quasi-specular reflectance. Also, with the optical flow constraint equation [2] slightly modified for taking into account the fact that there may not be conservation of intensities when the reflecting properties of the surfaces are non-uniform, our approach has yielded good shape estimates for surfaces with varying albedo.

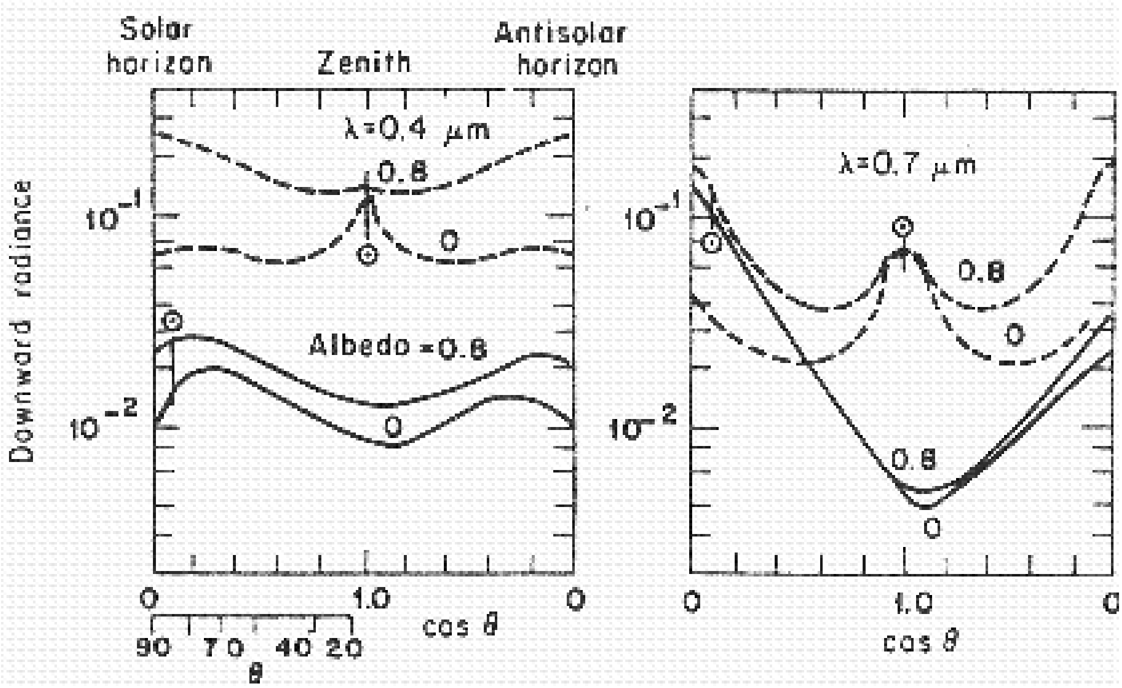


Figure 1

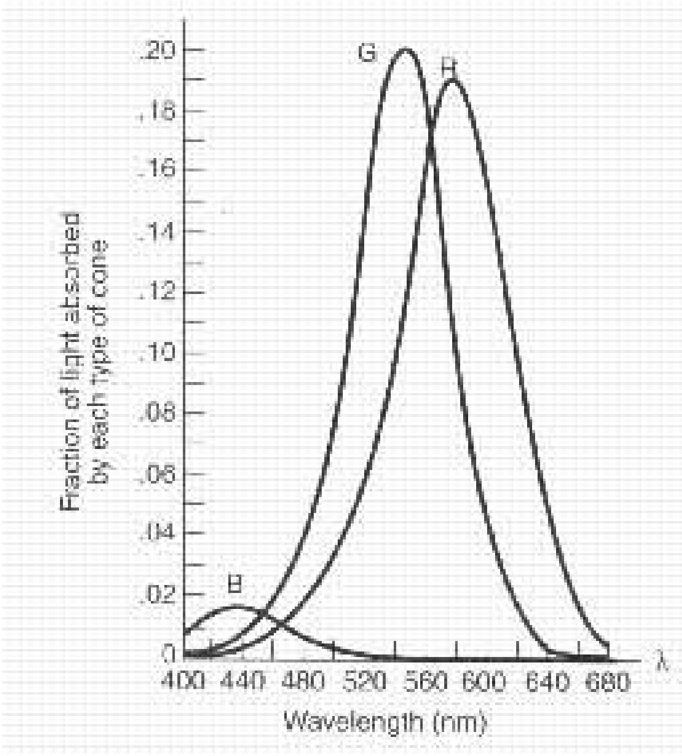


Figure 2

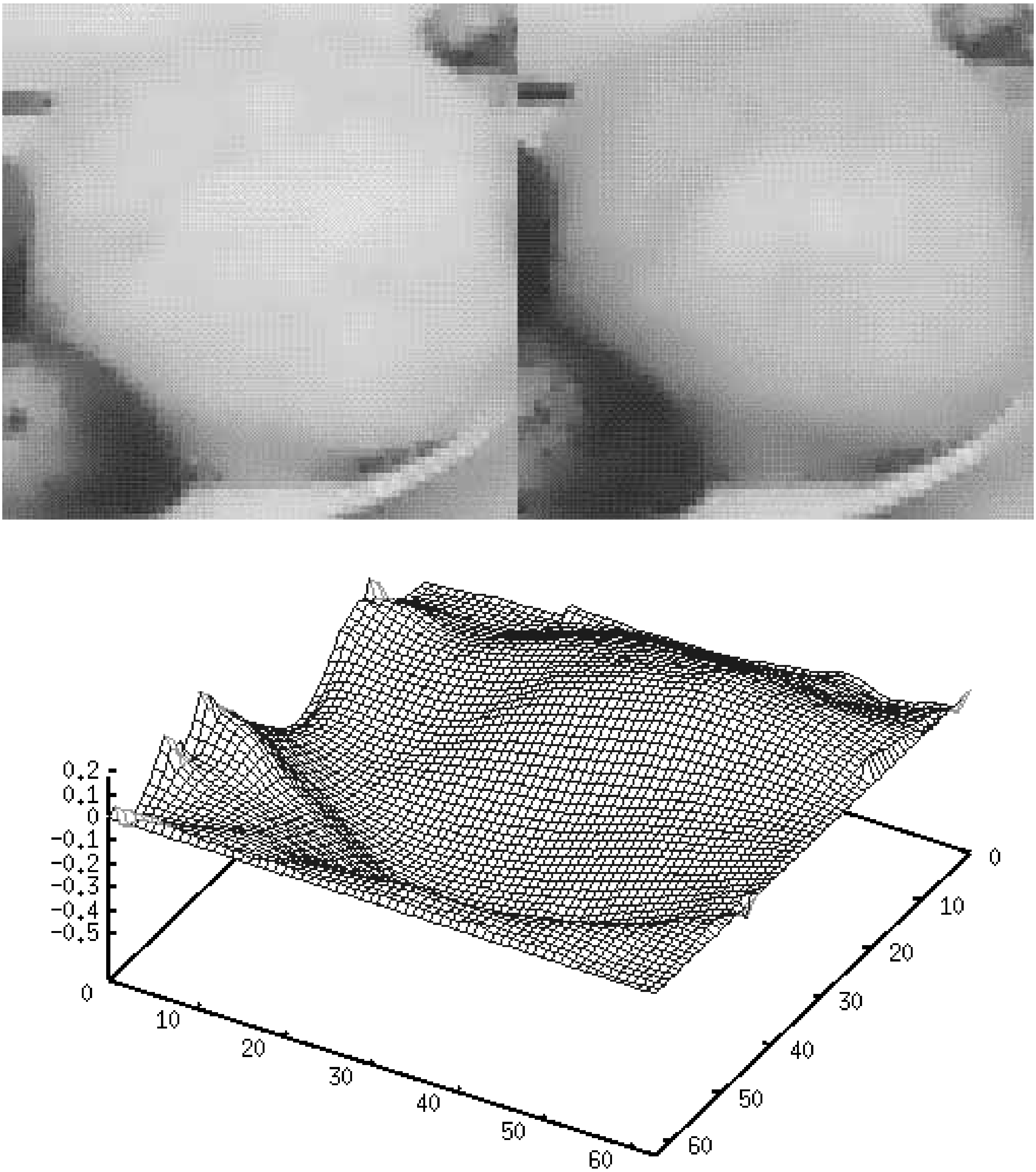
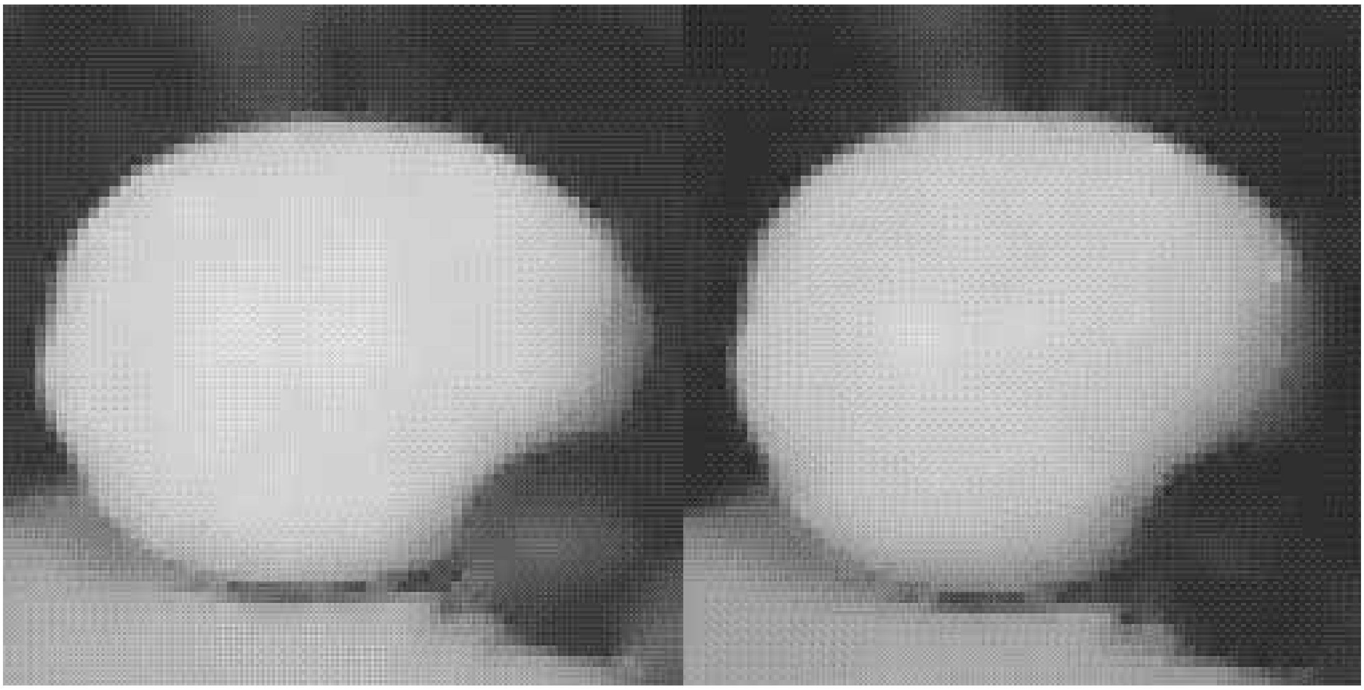


Figure 3



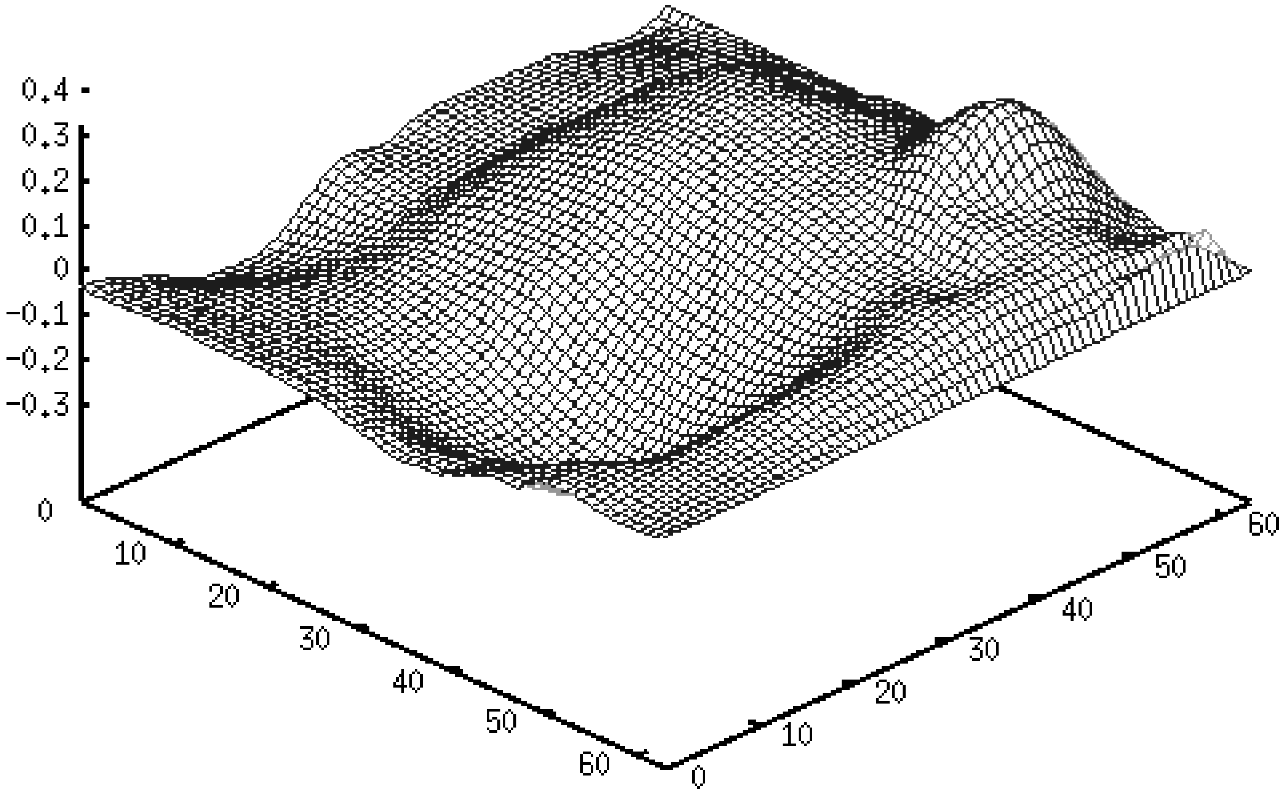


Figure 4

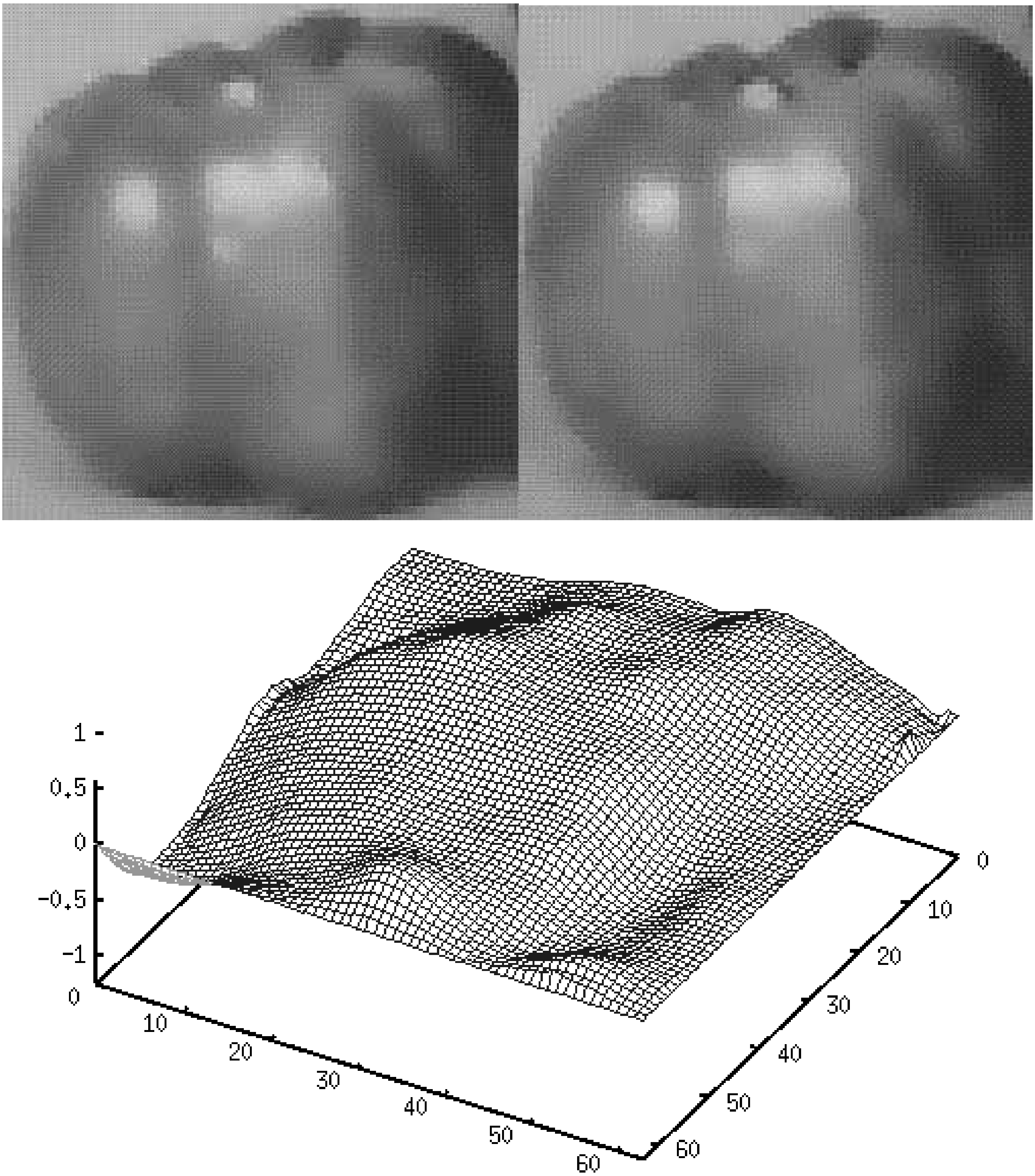


Figure 5