# Neural Network Implementation of Photometric Stereo

Yuji Iwahori
Faculty of Engineering
Nagoya Institute of Technology
Nagoya 466, Japan
e-mail: iwahori@center.nitech.ac.jp

Ardeshir Bagheri
Dept. of Computer Science
Langara College
Vancouver, BC, Canada V5Y 2Z6
e-mail: bagheri@cs.ubc.ca

Robert J. Woodham
Dept. of Computer Science
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
e-mail: woodham@cs.ubc.ca

## Abstract

This paper describes a neural network implementation of photometric stereo. Photometric stereo determines the surface orientation at each visible point on an object surface using measurements of image irradiance acquired under different conditions of illumination. The implementation described here uses a radial basis function (RBF) neural network to learn the mapping between three measured input image irradiances and the corresponding surface normal on a calibration sphere. The resulting neural network then is used to estimate the surface orientation of other test objects. The approach is empirical in that no explicit assumptions are made about light source directions or surface reflectance. It is sufficient that the calibration sphere and subsequent test objects be viewed under the same conditions of illumination and be made of the same material (i.e., have the same reflectance properties). The objective is to exploit the ability of neural networks to do non-parametric functional approximation and thus to avoid modeling errors associated with assumed parametric representations of surface reflectance.

Photometric stereo can fail locally in regions of cast shadow and significant interreflection. Here, we exploit the redundancy inherent in the three image irradiance values measured at each surface point to determine a local confidence estimate. This is achieved by training two distinct neural networks. The first predicts the intended output (surface normal) from the given input (three image irradiances). The second inversely predicts the input (three image irradiances) from the estimated output (surface normal). Comparison between the actual input and the inversely predicted input is used as the confidence estimate. Experiments on real data are described.

## 1 Introduction

Shape-from-shading, the problem of determining surface shape from the smooth shading present in a single image, was originally formulated by Horn [1,2]. Photometric stereo was introduced by Woodham [3] as a method to determine surface orientation locally using multiple images of an object surface acquired from a single viewpoint under different conditions of illumination. These approaches, and others that followed, exploit principles of optics as a source of radiometric constraint. In 1992, Healey, Shafer and Wolff [4] edited a three volume collection that includes most of the key papers defining the approach now referred to as "physics-based vision."

A variety of reflectance models have been proposed for both black and white (B&W) and color imaging. In some cases, the mapping between image irradiance and surface orientation takes on a simple functional form. Examples include (ideal) Lambertian reflectance, dielectrics (such as plastic) and metals. Reflectance modeling typically combines the optical parameters of the surface material with a geometric approximation to surface roughness. In computer vision, implementations commonly assume a particular functional form for reflectance. Calibration is used to estimate the unknown parameters of that functional form. Reflectance models

typically also are based on distant point source illumination so that one of the calibration sub-tasks is to estimate the direction of the light source(s). Conventional approaches to three light source photometric stereo determine surface orientation by solving a (generally nonlinear) set of three simultaneous image irradiance equations based on an assumed functional form for reflectance [5].

Alternatively, Woodham [6] has developed implementations of photometric stereo in which the reflectance function is determined empirically using a calibration object of known shape, typically a sphere. The mapping between image irradiance and surface orientation is represented explicitly in a lookup table (LUT). No explicit assumptions need to be made either about light source directions or about the functional form of surface reflectance. It is sufficient that the calibration sphere and subsequent test objects be viewed under the same conditions of illumination and be made of the same material (i.e., have the same reflectance properties). Empirical calibration has the added benefit of automatically compensating for the transfer characteristics of the sensor.

In an early neural network approach to shape-from-shading, Sejnowski and Lehky [7] proposed a method to recover surface curvature from an image of a curved object. Their method is applicable to Lambertian (convex or concave) surfaces. It estimates curvature of a test object after training with examples from images of objects with different curvatures and lighting directions. In other previous work, a neural network based implementation of photometric stereo was described [8]. In that work, the Phong model [9] of specular reflectance was assumed and the neural network was used to estimate unknown parameters in what overall was a parametric method.

The novel idea in this paper is to exploit the ability of a neural network to do non-parametric functional approximation in order to "learn" the required mapping between image irradiance and surface orientation. Data obtained from a calibration sphere are used to train the network. As with the previous non-parametric implementation [6], no explicit assumptions need to be made either about light source directions or about the functional form of surface reflectance. Here, however, the explicit lookup tables are replaced by the learning operation in a neural network. The two methods are directly comparable. At the same time, the expectation is that the neural network approach will generalize to higher dimensional problems where the space required to explicitly represent lookup tables would become prohibitive.

The particular approach described uses a radial basis function (RBF) neural network and the orthogonal least squares (OLS) learning method described in [10]. Further, we exploit the redundancy inherent in the three image irradiance values measured at each surface point to determine a local confidence estimate. This is achieved by training a two distinct neural networks. The first predicts the intended output (surface normal) from the given input (three image irradiances). The second inversely predicts the input (three image irradiances) from the estimated output (surface normal). The difference between the actual input and the inversely predicted input is used as the confidence estimate. This helps to detect local regions where photometric stereo fails due to cast shadow or significant interreflection. Experiments on real data are described.

## 2 Problem Formulation

A standard geometry of shape from shading is assumed. That is, let the object surface be given explicitly by $z = f(x, y)$ in a right-handed Euclidean world coordinate system, with the positive Z direction pointing to the viewer. Image projection is assumed orthographic with the image XY axes coincident with the world XY axes. Let a (unit) surface normal vector at any surface point be $(n_1,n_2,n_3)$. Then an image irradiance equation is

$$E(x, y) = R(n_1, n_2, n_3) \tag{1}$$

where $E(x, y)$ is the image irradiance and $R(n_1, n_2, n_3)$ is the reflectance map, here using the surface normal, $(n_1,n_2,n_3)$, to represent surface orientation. A reflectance map combines information about surface material, scene illumination and viewing geometry into a single representation that determines image brightness as a function of surface orientation.

Photometric stereo uses multiple images obtained under the identical geometry but under different conditions of illumination. In three light source photometric stereo, three image irradiance equations,

$$\begin{aligned} E_1(x, y) &= R_1(n_1, n_2, n_3) \\ E_2(x, y) &= R_2(n_1, n_2, n_3) \\ E_3(x, y) &= R_3(n_1, n_2, n_3) \end{aligned} \tag{2}$$

generally overdetermine the local surface orientation at each point, $(x, y)$, since three intensity measurements, $E_1$, $E_2$, $E_3$, are available to estimate the two unknowns of surface orientation. When using the surface normal to represent surface orientation,

it may appear that there are three unknowns, $n_1$, $n_2$, $n_3$, to determine. There are only two since we also have the *a priori* constraint that the length of the normal vector is one.

Some formulations of photometric stereo, including Woodham's original paper [3], allow the material's bidirectional reflectance factor (i.e., albedo) also to vary spatially. With varying albedo, there would indeed be three unknowns, two for surface orientation and one for albedo. Here, constant albedo is assumed so that the problem, as described, is overdetermined locally. We can then exploit the redundancy inherent in the three image irradiance values measured at each surface point to determine a local confidence estimate.

# 3 Neural Network Learning

## 3.1 RBF Networks and OLS

It is argued that RBF neural networks often can be designed in a fraction of the time it takes to train standard feed-forward networks. They are claimed to work well when many training vectors are available. An RBF network is an attractive choice in many applications. In particular, they have been widely used for strict interpolation in multidimensional spaces [11,12].

RBF networks represent nonlinearity via the choice of basis functions. A Gaussian isn't the only choice of basis function but it is the choice widely used and the one used here. One common learning algorithm for RBF networks first randomly chooses data points as RBF centers and then solves for the optimal weights of the network. Performance critically depends on the chosen centers. Centers often are chosen as an arbitrary selected subset of the data points. Typically, this is unsatisfactory. The resulting network may perform poorly, because the centers do not suitably sample the input data, or it may be very large, because a large number of centers is used.

An alternative learning procedure is based on an orthogonal least squares method [10,12]. The performance of the network still depends critically on the chosen centers. Because a fixed center corresponds to a given regressor in a given regression model, the selection of RBF centers can be regarded as a problem of subset model selection. The OLS method can be employed as a forward regression procedure to select a suitable set of centers (regressors) from a large set of candidates. At each step of the regression, the increment to the explained variance of the desired output is maximized.

The learning procedure adopted in this paper is based on the above OLS learning method. It chooses RBF centers one by one, in a systematic way, until an adequate network has been constructed. The algorithm has the property that each selected center maximizes the increment to the explained variance of the desired output while remaining numerically well-conditioned.

With this learning procedure, two RBF networks are trained using input/output data from a calibration sphere. Many training vectors are available since a data from the calibration sphere are dense and include all possible visible surface normals, $(n_1,n_2,n_3)$. Each RBF network used consists of two layers, (i.e., a hidden layer of P neurons and an output layer of 3 neurons), as shown in Figure 1.

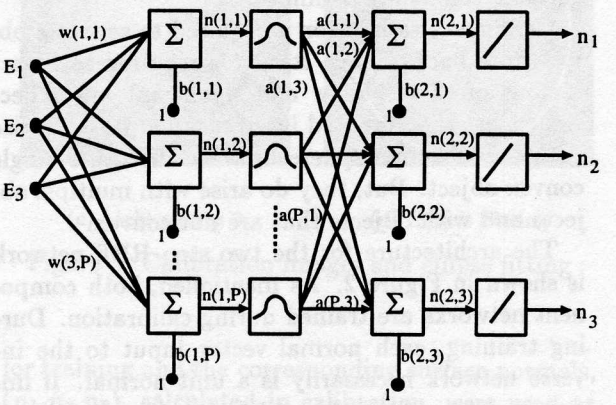



Figure 1: The RBF Neural Network

The learning procedure builds an RBF neural network one neuron at a time. Neurons are added to the network until the sum-squared error falls beneath an error goal (or a maximum number of neurons have been used). In learning, it is important that the so-called spread constant of the radial basis function be large enough that the neurons respond to overlapping regions of the input space, but not so large that all the neurons respond in essentially the same manner. Once learning is complete, that which has been learned is represented by the weights connecting each RBF neural network unit. The resulting network generalizes in that it predicts a surface normal, $(n_1,n_2,n_3)$, given any triple of irradiances values, $(E_1,E_2,E_3)$. The resulting network thus can be used to estimate the surface orientation of other test objects.

## 3.2 A Confidence Estimate

The RBF network trained as described in section 3.1 necessarily predicts a surface normal, $(n_1,n_2,n_3)$, for any input triple, $(E_1,E_2,E_3)$, whether or not that

triple comes from a point on the calibration sphere (or other test object with similar reflectance properties). Thus, it is valuable to have some measure to help segment input values for which the estimated surface normal can be believed from input values for which the estimated surface normal is suspect. Nothing in the training of the network in section 3.1 enforces the constraint that the surface normal is a unit surface normal. One idea is to use the length of the surface normal, $(n_1,n_2,n_3)$, to define a confidence measure. We have found a better idea to be the simultaneous training of a second network, during calibration, to inversely predict the input (three image irradiances) from the estimated output (surface normal). Comparison between the actual input and the inversely predicted input then serves as a suitable confidence estimate.

Confidence estimation is required to separate object from background. It also is required to detect regions of cast shadow and significant interreflection. Cast shadows and interreflection do not arise with a calibration sphere or with any other single convex object. But, they do arise with multiple objects and with objects that are not convex.

The architecture for the two step RBF network is shown in Figure 2. As mentioned, both component networks are trained during calibration. During training, each normal vector input to the inverse network necessarily is a unit normal. If image irradiances from the calibration sphere are input to this two step network, the output of the second step, which we call the resynthesized image irradiance, should be very similar to the original input. However, if an impossible triple (i.e., one that could not have arisen from the calibration sphere) is input, we expect the resynthesised image irradiances to be quite different since the resynthesized values necessarily correspond to points on the calibration sphere.

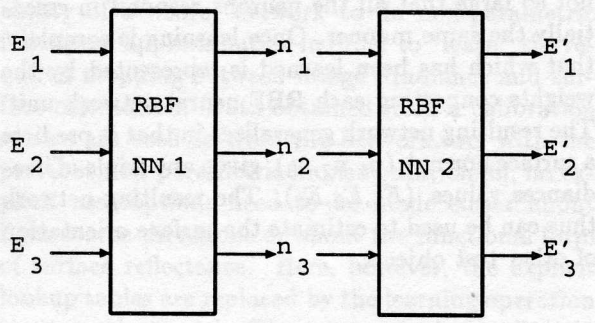



Figure 2: Two-Step RBF Neural Network

When applied to any point on a test object, the three image irradiances, $(E_1, E_2, E_3)$, are input to the first network. The length of the estimated output, $(n_1, n_2, n_3)$, is normalized to one prior to using it as input to the second network. Let $(E'_1, E'_2, E'_3)$ be the resynthesized image irradiances estimated by the second network. Let

$$d = \sqrt{(E_1 - E'_1)^2 + (E_2 - E'_2)^2 + (E_3 - E'_3)^2} \quad (3)$$

Clearly, the larger the value of $d$, the larger is the deviation of our test point from a point that could have arisen on the calibration sphere.

# 4 Experimental Results

## 4.1 Experimental Setting

A calibrated imaging facility (CIF) has been developed at the UBC Laboratory for Computational Intelligence (LCI) to control both scene parameters and conditions of imaging. It is based on an optical bench with mounting hardware for controlled positioning and motion of cameras, light sources and test objects. Specialized equipment includes a Sony DXC-755 3 CCD 24-bit RGB camera with Fujinon 10-120mm (manual) zoom lens and three Newport MP-1000 Moire (white light) projectors with associated Nikon lenses. The facility is integrated with other vision and robotics equipment in the LCI with direct access to specialized image processing hardware including DataCube modules and networks of Texas Instruments C40's and transputers. Work on photometric stereo requires multiple images of a scene acquired simultaneously under different conditions of illumination. This is most easily achieved by multiplexing the spectral dimension. With appropriate filtration, the three projectors become spectrally distinct lights sources, one red, one green, and one blue. The three color separation filters used are the Newport FS-225 set. The projectors illuminate the work space from different directions. In three light source photometric stereo, the Sony 3 CCD RGB camera is used to simultaneously acquire three separate B&W images, one corresponding to each condition of illumination.

Two objects are used in the experiments reported. One is a pottery sphere, used for calibration purposes, and the other is a pottery boy face. In this case, both objects are made of the same material with the same reflectance properties. No particular assumptions for the surface reflectance or light source directions are used or needed for the experiments. The boy face was mounted on a rotational motion stage. Different test images were acquired from the same viewpoint and illumination but with different rotations.

## 4.2 Calibration

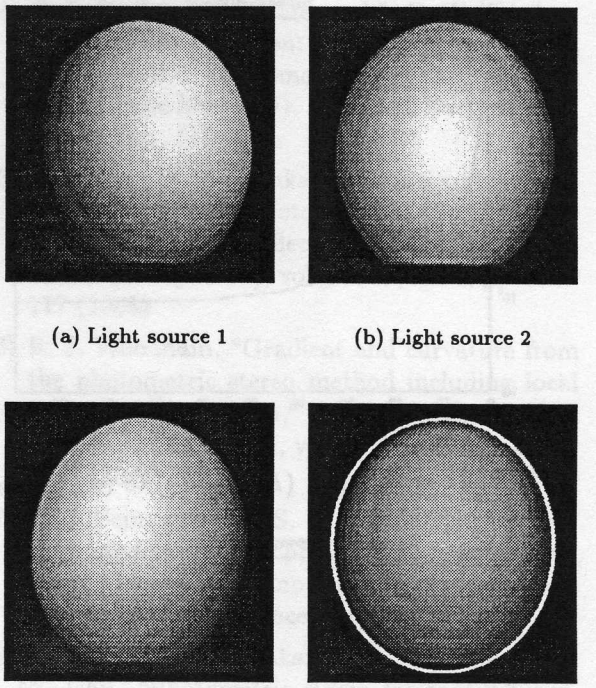
Calibration measures reflectance data using an object of known shape. A sphere is a good choice because it is convex, thus eliminating interreflection, because it is easy to dead reckon local surface orientation geometrically from the object silhouette and because it spans all possible visible surface normal vectors. In what follows, the silhouette of the calibration sphere appears as an ellipse because camera aspect ratio correction has not been applied. Indeed, aspect ratio correction can be included in the calibration. Let $\lambda$ be the scale of $y$ compared to $x$. Then, the equation of the sphere centered at the origin with radius $r$ is

$$x^2 + (\lambda y)^2 + z^2 = r^2$$

The outward (unit) surface normal vector at image point $(x,y)$ is

$$(n_1, n_2, n_3) = (\frac{x}{r}, \frac{\lambda y}{r}, \frac{z}{r})$$

Processing the calibration images involves three steps. First, the three images are summed. This is done to make sure that no part of the object's boundary is missed because it lies in shadow. Second, a threshold is selected to separate object from background. Simple thresholding is sufficient since, by design, the object is distinct from the black background. Third, a simple least squares method is used to estimate the equation of the ellipse which best fits the object boundary. The equation of the ellipse then establishes the relation between $(x,y)$ and $(n_1,n_2,n_3)$ for all points on the calibration sphere. The three original images and the fitted ellipse are shown in Figure 3. A small portion near the base of the sphere is obscured by the mount used to hold the sphere in place. Only visible points are used in network training.



(a) Light source 1 (b) Light source 2

(c) Light source 3 (d) Ellipse Fitting

Figure 3: Calibration images and ellipse fitting

## 4.3 Learning and Generalization

The two RBF networks are trained using data from the calibration sphere. The first network is trained to learn the mapping of image irradiances, $(E_1,E_2,E_3)$, to surface normals, $(n_1,n_2,n_3)$. The second network is trained to learn the inverse mapping of surface normals, $(n_1,n_2,n_3)$, to image irradiances, $(E_1,E_2,E_3)$, using the same training data. In general, both mappings are non-linear.

Thus, the number of input and output units for each network is three. Three image irradiances, $(E_1,E_2,E_3)$, spaced 10 pixels apart on the original 512×512 input images, were used as the inputs for training and the corresponding surface normals, $(n_1,n_2,n_3)$, calculated in calibration, were used as the outputs. Overall, training sets of inputs and outputs were taken from 600 points on the calibration sphere. The RBF neural network software used is from Matlab as described in section 6 of the Matlab neural network toolbox user's guide [13]. For each network, iterative learning proceeded for 100 epochs (i.e., until the number of hidden units became 100). With 8–bit-per-pixel input images, $0 \leq E_i \leq 255$, $i = 1,2,3$. For visible surface points, $-1 \leq n_i \leq 1$, $i = 1,2$, and $0 \leq n_3 \leq 1$. In learning, the spread constant of the radial basis function should be much larger than the minimum distance and much smaller than the maximum distance between input vectors. Given the range of values for $E_i$ and $n_i$, $i = 1,2,3$, cited above, the spread constant for the first network was set to 200 and the spread constant for the second network was set to 1. The learning states for the two networks are shown in Figures 4 and 5, respectively. Learning was effectively achieved using the 600 points on the calibration sphere.

Once neural network learning is complete, the resulting networks can be applied to the test object images. Two examples are shown. Three images, called boy-1, are shown in Figure 6. For boy-1, the

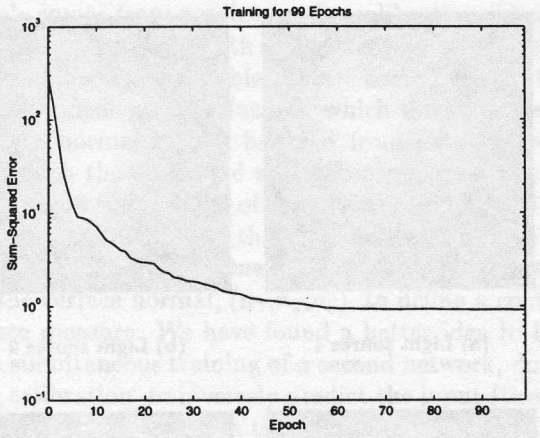



Figure 4: Learning 1: $(E_1, E_2, E_3)$ to $(n_1, n_2, n_3)$

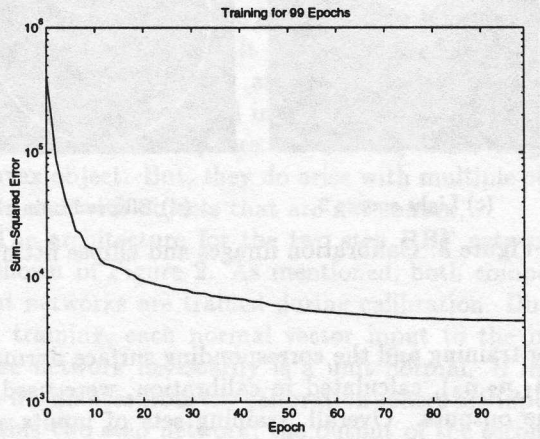
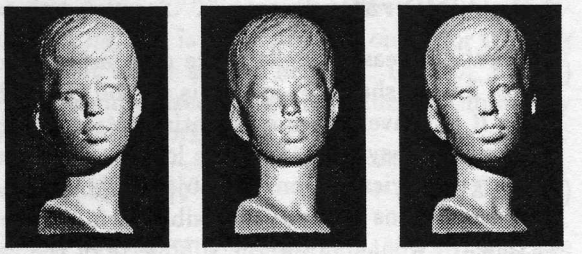



Figure 5: Learning 2: $(n_1, n_2, n_3)$ to $(E_1, E_2, E_3)$

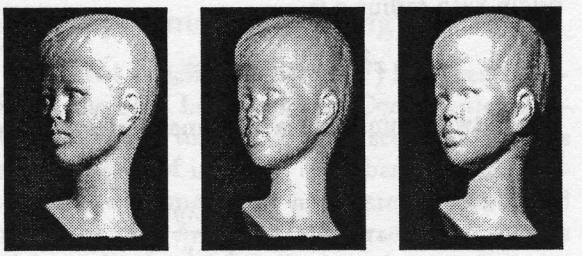
face is oriented directly towards the viewer. Figures 8-(a) and 8-(b) show the estimated surface orientation encoded as slope and aspect. Figure 8-(a) linearly encodes the slope angle, $e$, (i.e., the angle between the surface normal and the vector pointing to the viewer) as a gray value in the range black ($e = 0$) to white ($e = \pi/2$). Figure 8-(b) plots the aspect (i.e., the projection of the surface normal onto the XY plane) as a short line segment. (To avoid clutter, aspect is plotted for every fourth point in $x$ and $y$.) Figure 8-(c) shows the value of the confidence estimate, $d$, also encoded as a gray level. Bright values corresponds to points in the image where confidence in the estimated surface orientation is lowest. As can be seen, they correspond to areas effected by cast shadow or interreflection.

Another three images, called boy-2, are shown in Figure 7. For boy-2, the face is rotated to the right relative to boy-1. Figures 9-(a), 9-(b) and 9-(c) show the estimated slope, aspect and confidence estimate, $d$, encoded as in Figure 8.



(a) Image 1 (b) Image 2 (c) Image 3

Figure 6: The three boy-1 images



(a) Image 1 (b) Image 2 (c) Image 3

Figure 7: The three boy-2 images

No quantitative analysis of the results has been performed since the true shape of the boy face is not known. Both the boy-1 and boy-2 results are qualitatively correct except in areas where the value of $d$ is large. Conversely, where $d$ is small, there is reason to be confident in the estimate of surface orientation. The first neural network cannot possibly be correct in regions of cast shadow and significant interreflection because there the input values fall well outside the space of inputs covered in training. The second network thus provides an essential check on the confidence to attach to the estimated surface normal.

# 5 Conclusions

A new method for implementing photometric stereo has been described. It exploits the ability of neural networks to do non-parametric functional approximation. The implementation described uses a radial basis function (RBF) neural network and an orthogonal least squares (OLS) learning method. This was sufficient both to learn and to represent the mapping between image irradiances and surface normal for three light source photometric stereo. The approach is empirical in that no explicit assumptions

are made about light source directions or surface reflectance. It is sufficient that the calibration sphere used in training and the subsequent test objects be viewed under the same conditions of illumination and have the same reflectance properties.

The implementation exploits the redundancy inherent in the three image irradiance values measured at each surface point. This is achieved by training a two step neural network. The first step predicts the intended output (surface normal) from the given input (three image irradiances). The second step inversely predicts the input (three image irradiances) from the estimated output (surface normal). Comparison between the real input and the inversely predicted input is used as the confidence estimate. The local confidence estimate helps to detect regions of cast shadow and interreflection.

Overall, the representation in terms of neural networks is both compact and efficient. It is hoped that this will allow extension to higher dimensional problems involving more than three light sources and varying reflectance properties (eg., albedo).

## Acknowledgement

The research described in this paper was performed while the first author, Iwahori, was visiting the University of British Columbia with the support of Japanese Ministry of Education. Major support for Woodham's research was provided by the Institute for Robotics and Intelligent Systems (IRIS), one of the Canadian Networks of Centres of Excellence, by the Natural Sciences and Engineering Research Council of Canada (NSERC) and by the Canadian Institute for Advanced Research (CIAR). The authors thank members of the UBC Laboratory for Computational Intelligence and Prof. N. Ishii of Nagoya Institute of Technology for useful discussions.

## References

[1] B. K. P. Horn, "Obtaining shape from shading information," in The Psychology of Computer Vision, P. H. Winston (ed.), pp. 115–155, McGraw-Hill (1975).

[2] B. K. P. Horn, "Understanding image intensities," Artificial Intelligence, vol. 8, no. 2, pp. 201–231 (1977).

[3] R. J. Woodham, "Photometric method for determining surface orientation from multiple images," Optical Engineering, vol. 19, no. 1, pp. 139–144 (1980).

[4] G. E. Healey, S. A. Shafer and L. B. Wolff (eds.), "Physics-Based Vision: Principles and Practice," (Vol. 1 Radiometry, Vol. 2 Color and Vol. 3 Shape Recovery), Jones and Bartlett Publishers, Inc., (1992).

[5] Y. Iwahori, N. Hiratsuka, H. Kamei and S. Yamaguchi, "Extended photometric stereo for an object with unknown reflectance property," Trans. IEICE (in Japanese), vol. J71-D, no. 1, pp. 110–117 (1988).

[6] R. J. Woodham, "Gradient and curvature from the photometric stereo method including local confidence estimation," Journal of the Optical Society of America, A, vol. 11, no. 11, pp. 3050–3068 (1994).

[7] T. J. Sejnowski and S. R. Lehky, "Neural network models of visual processing," text for 1987 Short Course on Computational Neuroscience, Society for Neuroscience, pp. 1–19 (1987).

[8] Y. Iwahori, H. Tanaka, R. J. Woodham and N. Ishii, "Photometric stereo for specular surface shape based on neural network," IEICE Trans. on Information and Systems, Special Issue on Neurocomputing vol. E77-D, no. 4, pp. 498–506 (1994).

[9] B. T. Phong, "Illumination for computer generated pictures," Comm. of ACM, vol. 18, no. 6, pp. 311–317 (1975).

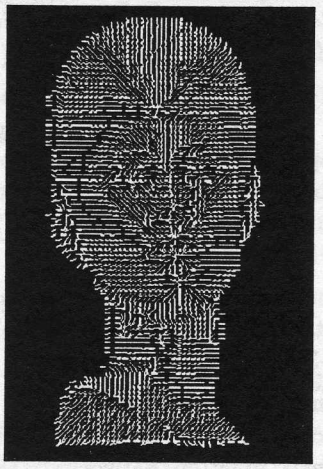
[10] S. Chen, C. F. N. Cowan and P. M. Grant, "Orthogonal least squares learning algorithm for radial basis function networks," IEEE Trans. on Neural Networks, vol. 2, no. 2, pp. 302–309 (1991).

[11] M. J. D. Powell, "Radial basis functions for multivariable interpolation: a review," in Algorithms for Approximation, J. C. Mason and M. G. Cox (Eds.), Oxford, pp. 143–167 (1987).

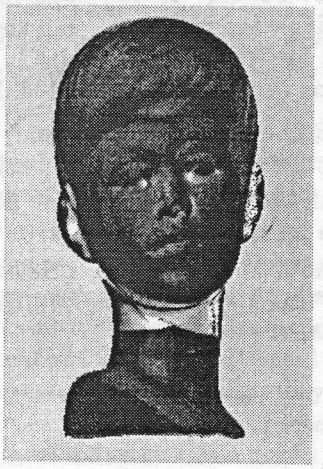
[12] D. S. Broomhead and D. Lowe, "Multivariable functional interpolation and adaptive networks," Complex Syst., vol. 2, pp. 321–355 (1988).

[13] H. Demuth and M. Beale, "Neural Network Toolbox User's Guide," The MathWorks Inc., Natick, Mass. (1994).

(a) Boy-1 slope



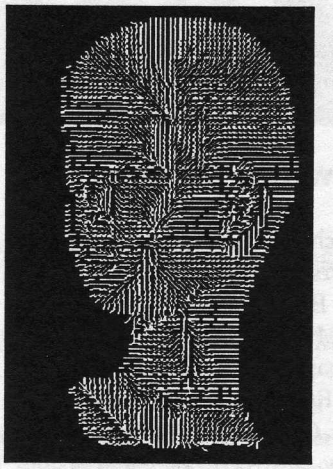
(b) Boy-1 aspect



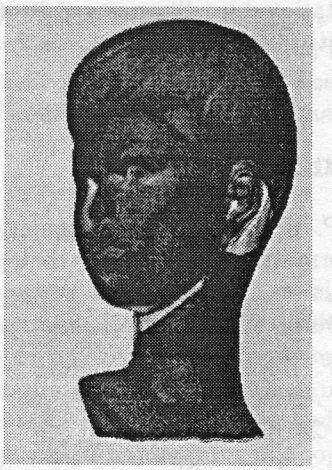
(c) Boy-1 confidence

Figure 8: Boy-1 slope, aspect and confidence

(a) Boy-2 slope



(b) Boy-2 aspect



(c) Boy-2 confidence

Figure 9: Boy-2 slope, aspect and confidence