Neural network based photometric stereo using illumination planning

Yuji Iwahori \ Wataru Kato [1], Md. Shoaib Bhuiyan [1],
Robert J. Woodham [2] and Naohiro Ishii [1]
[1] Faculty of Engineering, Nagoya Institute of Technology, Nagoya 466, Japan
[2] Dept. of Computer Science, Univ. of British Columbia, Vancouver B.C. Canada V6T 1Z4

Abstract

This paper proposes a new approach of neural network based photometric stereo in which all the directions of illumination are close to and is rotationally moved around the viewing direction. Adjacent light sources cause a numerically ill-conditioned problem. Ill-conditioning is overcome in two ways. First, many more than three images are acquired. Second, principal components analysis is used as a linear preprocessing technique to determine a reduced dimensionality subspace for use as input. The approach is empirical and uses the radial basis function (RBF) neural network to do non-parametric functional approximation. One neural network maps image irradiance to surface normal. A second network maps surface normal to image irradiance. The two networks are trained using samples from a calibration sphere. Comparison between the actual input and the inversely predicted input is used as a confidence estimate. Further, we introduce the illumination planning, which disposes the information of the cast shadow region from the ill-effected light sources for each point on the object using the confidence estimate. Then, the best combination of NN units with well-effected light sources for each point are used to get the relevant results. The results on real data are also evaluated.

1 Introduction

Shape-from-shading [Horn, 1975; 1977] is the problem of determining surface shape from the smooth shading present in a single image. Photometric stereo [Woodham, 1980] determines surface orientation locally using multiple images of an object surface acquired from a single viewpoint under different conditions of illumination. Principles of optics are used as a source of radiometric constraint.

Recently, theoretical analysis of shape-from-shading and photometric stereo seem to dominate the literature at the expense of practice. Theoretical work often considers Lambertian reflectance alone. Regrettably, this convinces many potential implementors that the approach is of little practical value. In "physics-based vision," serious attention now is paid to other reflectance models [Healey *et* al., 1992] [Healey *et al,* 1994]. Clearly, this is fundamental. But, it has not yet proven effective in practice.

In [Woodham, 1994], the non-linear mapping between image irradiance and surface orientation was represented explicitly in a lookup table (LUT). This allowed near real-time implementation of three light source photometric stereo on full frame video data at near video frame rates (i.e., 15Hz).

Iwahori has pursued neural network implementations of photometric stereo [iwahori *et* a/., 1995]. Using principal components analysis (PCA) in conjunction with a neural network for the case of a moving, nearby light source was first reported in [iwahori *et* al., 1995]. Not surprisingly, it was observed that, with only small movements of the light source, standard photometric stereo becomes numerically ill-conditioned. For this reason, it seemed difficult to realize implementations based on a single moving light source. In the present version of the work, the moving light source is standardized into a rotational pattern.

While, some approaches using the illumination planning to photometric stereo are proposed in [Solomon *et* al., 1992] and [Solomon *et al.,* 1995]. These approaches are intended to remove the glossy components from multiple light sources.

Factors limiting application of photometric stereo include shadows, interreflection and the reality that the available range of illumination directions often is restricted. There are trade-offs. Greater difference in the directions of illumination leads to a better conditioned numerical estimation of surface orientation. At the same time, greater difference in the directions of illumination leads to a larger number of surface points that fail to be commonly illuminated and hence fail to support any estimation of surface orientation. This paper proposes a new NN based approach using the illumination planning to remove the cast shadow region as much as possible from the multiple rotational light sources. The illumination planning is used is to determine the best combination of

light source for each point on the object to recover the gradient. Combination of two distinct NNs are used to determine a local confidence. Experiments on real data are demonstrated for the illumination planning.

2 Neural Network Based Photometric Stereo

2.1 Principle of photometric stereo

Let the object surface be given explicitly by $z = f(x, y)$ in a right-handed Euclidean 3D scene coordinate system, with the positive Z direction pointing to the viewer. Image projection is assumed to be orthographic. Let a (unit) surface normal vector at any surface point be $[n_1, n_2, n_3]$. Photometric stereo uses multiple images obtained under the identical geometry but with different conditions of illumination. With p light source directions, p images, and hence p equations, are obtained.

$$\begin{aligned} E_1(x,y) &= R_1(n_1,n_2,n_3) \\ E_2(x,y) &= R_2(n_1,n_2,n_3) \\ &\vdots \\ E_p(x,y) &= R_p(n_1,n_2,n_3) \end{aligned} \tag{1}$$

where $E(x,y)$ is the image irradiance and $R(n_1, n_2, n3)$ is the reflectance map, defined using the surface normal, $[n_1, n_2, n_3]$, to represent surface orientation. With $p > 2$, the p image irradiance measurements, $E_1, E_2, \ldots, E_p$, generally overdetermine the local surface orientation at each point, (x,y), since surface orientation only has two degrees of freedom. But, if all directions of illumination are nearly collinear, as will be the case here, the problem can be ill-conditioned.

2.2 Principal Components Analysis

Principal components analysis (PCA) is a technique from multivariate statistical data analysis. The basic idea is to describe the dispersion of an array of n points in a p-dimensional space via a new set of orthogonal linear coordinates, called the principal components, determined so that the principal components are mutually uncorrelated and so that the sample variances of the n points are ranked in decreasing order of magnitude with respect to these new coordinates. PCA is an invertible coordinate transformation that captures no more (or no less) information than was originally present.

Algebraically, PCA involves finding the eigenvalues and eigenvectors of the sample covariance matrix. (A variant of standard PCA uses the sample correlation matrix instead). A covariance matrix is symmetric and positive definite so that its eigenvectors form an orthogonal set. The (normalized) eigenvectors define the new set of coordinates axes (i.e., the principal component axes). Since PCA does remove correlation from data coordinates (in a p-dimensional space), it does have application to dimensionality reduction, given sample data with significant correlation. Dimensionality reduction typically is achieved by simple elimination of lower order principal components. With p directions of illumination that are nearly collinear, one expects a high degree of correlation between images. It is because of this that PCA is helpful.

PCA is implemented as follows. First, calculate the $p \times p$ covariance matrix, $\mathbf{K}$, using sampled object points commonly illuminated in each of the p input images. Second, determine the normalized eigenvectors, ϕ_k, and associated eigenvalues, λ_k, of the covariance matrix, $\mathbf{K}$, $k = 1, 2, \ldots, p$. The normalized eigenvectors define a new, orthonormal basis for p-space. Let Φ be the $p \times p$ matrix whose columns are ϕ_k, $k = 1, 2, \ldots, p$. Let $[E_1, E_2, E_3, \ldots, E_p]$ be the p-tuple of image irradiance measurements from a given object point. PCA defines a new p-tuple, denoted here by $[A_1, A_2, A_3, \ldots, A_p]$, where

$$\begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{bmatrix} = \Phi^T \begin{bmatrix} E_1 \\ E_2 \\ \vdots \\ E_p \end{bmatrix} \tag{2}$$

(^{T} denotes matrix transpose). That is, the matrix Φ determines the coefficients required to represent the given $[E_1, E_2, E_3, \ldots, E_p]$ in terms of the new basis, ϕ_k, $k = 1, 2, \ldots, p$. For each of n sample p-tuples, one obtains a corresponding $[A_1, A_2, A_3, \ldots, A_p]$. The kth principal component, viewed as a data set, is the set of all n values of the kth coordinate of $[A_1, A_2, A_3, \ldots, A_p]$. One chooses to retain the three most significant principal components $[A_1, A_2, A_3, \ldots, A_m]$ from the original set $[E_1, E_2, E_3, \ldots, E_p]$ for each point, where m means the dimension number of sub-space.

2.3 RBF Networks and OLS Learning

Neural networks are attractive for non-parametric functional approximation. A radial basis function (RBF) neural network is one choice suitable for many applications. In particular, it has been widely used for strict interpolation in multidimensional spaces. 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 expected to work well when many training vectors are available.

RBF networks represent non-linearity via the choice of basis functions. A Gaussian isn't the only choice of radial basis function for RBF networks but it is the choice widely used and the one used here. One common learning algorithm for RBF networks is based on first randomly choosing data points as RBF centers and then solving for the optimal weights of the network. Performance, however, critically depends on the chosen centers. In practice the centers often are chosen to be an arbitrary selected subset of the data points.

The learning procedure adopted here is based on the orthogonal least squares (OLS) method. 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. It chooses radial basis function

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. Here, we use reduced dimensionality data preprocessed by PCA as described in section 2.2. Many training vectors are available since 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.

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 has 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 input values, $[A_1, A_2, A_3]$. The resulting network trained using the calibration sphere can then be used to estimate the surface orientation of other test objects.

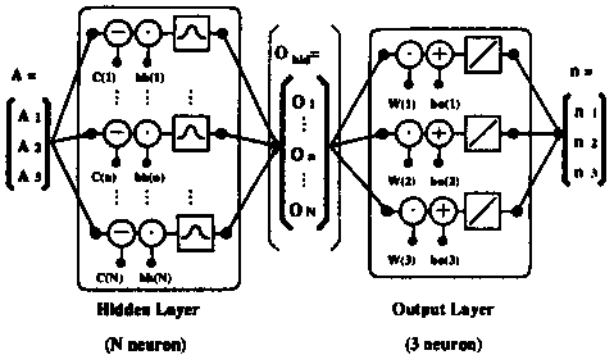


Figure 1: RBF Neural Network

The 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, $[A_1, A_2, A_3]$ from the estimated output, $[n_1, n_2, n_3]$. Comparison between the actual input and the inversely predicted input then serves as a suitable confidence estimate.

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 the 3D subspace from the calibration sphere is input to this two step network, the output of the second step, which we call the resynthesized input subspace, 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 input subspace to be quite different since the resynthesized values necessarily correspond to points on the calibration sphere.

3 Illumination Planning

3.1 Implementation

NN-based photometric stereo using more than three light sources can be implemented as follows.

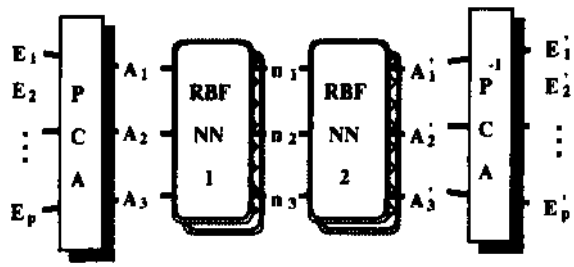


Figure 2: Process units

Process consists of two NN units. Two RBF-NNs are learned under the condition that the number of the dimension for both of inputs and outputs is three for each NN. First, PCA gets $A = (A_1, A_2, A_3)$ from $E = (E_1, E_2, \cdots, E_n)$ to make it efficient to learn the input/output mapping and to save the learning time by NN. Then, $A = (A_1, A_2, A_3)$ is input to the first RBF-NN1 and the surface normal vector $n=(n_1, n_2, n_3)$ is obtained, then the inverse RBF-NN2 is applied to estimate $A = (A_1, A_2, A_3)$ from n. By applying the inverse transformation of PCA, we can get the corresponding output $E = (E_1', E_2', \cdots, E_p')$.

Confidence estimate for each point on the test object is calculated as

$$C = \frac{\sqrt{\sum_{i=1}^{p} |E_i - E_i'|^2}}{p} \quad (i = 1, 2, \cdots, p) \qquad (3)$$

The precision for total points is evaluated as

$$\text{mean}(C) = \frac{\sum_{n=1}^{N_{obj}} C(n)}{N_{obj}} \qquad (4)$$

where N_{0bj} is the total pixel number of the test object region.

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. Details of the calibration procedure used are described elsewhere [Woodham, 1994; Iwahori *et al,* 1995].

Illumination planning is applied by picking up the data that does not give the cast shadow under the multiple light sources for each point on the test object.

3.2 Strategy of illumination planning

Three light source photometric stereo necessarily causes cast shadow problem for the complicated curved object and the image irradiance equation cannot be applied for cast shadow region. Generally, shadow region is given by two causes: one is the region where the shadows depends on the geometrical local shape, and the other is the cast shadow which is defined as that another local shape interrupts the illumination and it casts shadow. The former shadow occurs when the angle between the surface normal vector and the light source direction vector becomes greater than 90 degrees, however, actual image irradiances still hold some information to recover the local surface gradient. While, the latter shadow is defined as the cast shadow and it is impossible to use the irradiance values to recover the local surface gradient.

The approach uses the rotational multiple light sources more than three, and once recovers the distribution of surface orientation for the test object using all light sources. After this process, one can determine the confidence for each point of the test object from the difference between $E = (E_1, E_2, \cdots, E_p)$ and $E' = (E'_1, E'_2, \cdots, E'_p)$. From the confidence values, the method disposes the ill-effected light sources among all light sources and recovers the shape again from only the well-effected light sources to get the relevant results. This is proposed as the strategy of the illumination planning based approach. The procedures are given as follows.

1. Recovery from all p-light sources

   All p-light sources, i.e., $E_{all} = (E_1, E_2, \cdots, E_p)$ are input into the NN-units and obtain n_{all} and E'_{all}.

2. Extraction of cast shadow region

   The object surface regions can be defined as one of "normal reflection region", "intersection region", "shadows depending on local shape region" and "cast shadow region". This method finds the "cast shadow region" from the threshold process with Th_σ to the standard derivation σ_{all} of the difference of $(E_{all} - E'_{all})$ for each pixel.

$$D_l = E_l - E'_l \quad (l = 1, 2, \cdots, p) \tag{5}$$

$$\sigma_{all} = \sqrt{\sum_{l=1}^{p} (D_l - \frac{\sum_{l=1}^{p} D_l}{p})^2} \tag{6}$$

This process can find "normal reflection region" and "shadows depending on local shape region" under the condition that $E_{all} \approx E'_{all}$ and $\sigma_{all} \approx 0$. While, "interreflection region" increases the observed image irradiance and affects to E'_{all}, but E'_{all} are stable for all light sources and the variance of C_{all} becomes small for this interreflection region.

Only the "cast shadow region" takes the various values for the variance of the difference between E_{all} and E'_{all}. Then, the values of C_{all} and σ_{all} become high for this "cast shadow region".

From the above reasons, the appropriate threshold process to a_{all} can find the "cast shadow region".

3. Disposing light source which casts shadow

   For each point of the extracted cast shadow region, new combination group *sub* of light sources is determined by disposing the light source which casts shadow from the all combination of group *all* using the difference value Th_D of each light source.

Let D_{all} be the difference between E_{all} and E'_{all}, then the light source which has the larger value of $|D_l| = |E_l - E'_l|$ has the worse one and should be *a prtort* disposed as light source which casts shadow .

Since any E'_l actually has the positive value for the cast shadow and the range of E_l is less than it and greater than 0 for the same point, only the light sources for which D_l are negative and less than Th_D, are disposed.

4. Final recovery for each point without light sources which cast shadow

   Image irradiance vector $E_{,Aub}$ from the illumination of the extracted light source group *sub* is input into the corresponding NN-unit and the shape of cast shadow regions are recovered again. The vectors (n_{sub}, E'_{nub}) obtained here are used as the final results for the cast shadow region.

4 Experiments

4.1 Configuration

The light source configurations are shown in Figure 3. The distance between object and camera was set to be 200 (cm). Fifteen light source directions were used, each differing from the z-axis by 8.58 degrees (i.e., azimuth angle) for the illumination of eight light sources with smaller radius, while by 11.85 degrees for that of seven light sources with larger one.

Two objects are used. One is a pottery sphere, used for calibration purposes, and the other is a pottery rnoon face. 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.

4.2 Construction of NN-Units

This experiments disposes the worst light sources of one to five in the total fifteen light sources. Therefore, the number of NN-units prepared is $\sum_{r=10}^{15} {}_{15}Cr = 4944$ for the combination of $({}_{15}C_{15}$ to $ {}_{15}C_{10})$.

The covariance matrix size used to recover the gradient for each point on the test object depends on the the

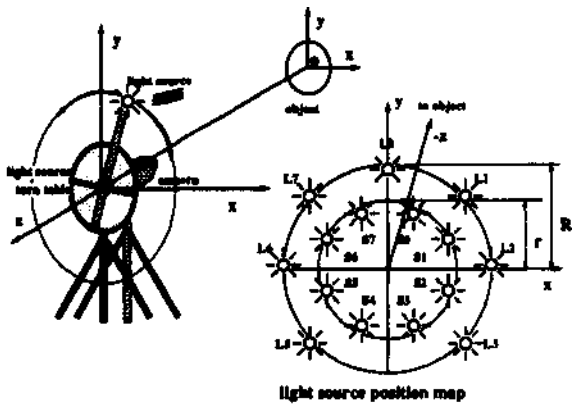
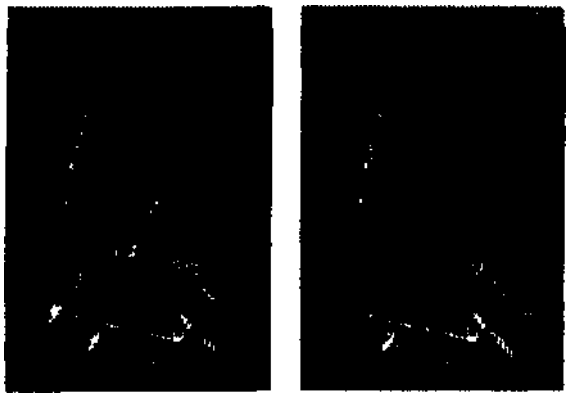


Figure 3: Configuration of Rotational Moving Light Source

number of light sources used for each point. The learning count for learning each NN-unit was set to be 40. This value is sufficient for the scale and generalization of NN used in the experiment. The values of th_σ and th_D was set to be 15 and -20 (mean of D), respectively.

4.3 Results

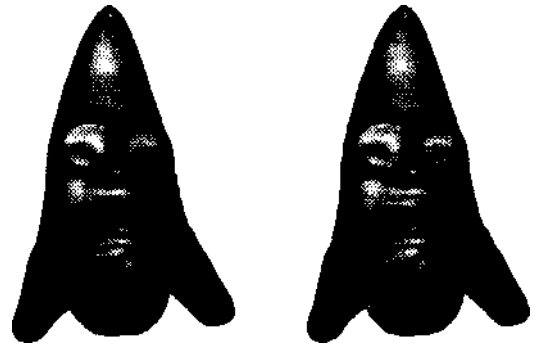
The results of this approach are shown in the Figures 4 to 7. The mean value of the estimated confidence, mean(C), was 4.11619 after the illumination planning, while mean(C) was 4.73629 before applying the illimination planning.



(a) before illumination planning (b) after illumination planning

Figure 4: Needle maps

Figure 4 shows the needle map obtained before/after the illumination planning, respectively, while Figure 5 shows the slope map which 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 6 shows



(a) before illumination planning (b) after illumination planning

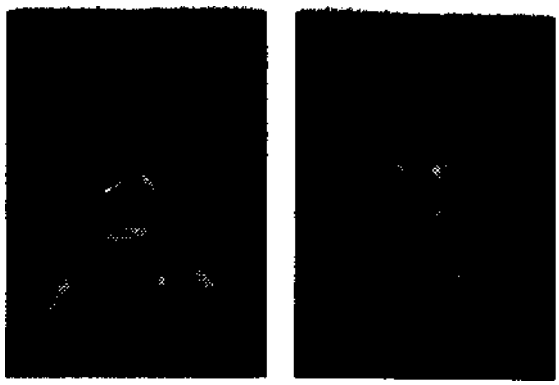
Figure 5: Slope maps

(a) before illumination planning (b) after illumination planning

Figure 6: Confidence maps

the values of the confidence estimate C, encoded as a gray level. Bright values correspond to points in the image where confidence in the estimated surface orientation is lowest. As can be seen, after the illumination planning, the confidence estimate is apparently improves for the region affected by cast shadows. Figure 7-(a) shows how the region was improved by this illumination planning as a gray level. Bright values corresponds to the well improved regions. Figure 7-(b) shows an example of the original input images from the light source L1.

Figure 4-(a) and Figure 5-(a) shows some local random results around the two eyes, mouth, and arms, but Figure 4-(b) and Figure 5-(b) improved the results and removed the disconnected vectors around there. Figure 6-(b) is also improved in comparison with Figure 6-(a) for the corresponding regions. Numerical evaluation shows the mean value of the confidence estimate, mean(C), was improved with the rate of 13.1% and it is shown that the the regions effected by the cast shadow



(a) Comparison map (b) An *original* image of Moon-face : LI

Figure 7: Reference images

is extracted and sufficiently recovered in addition to the normal regions, and the approach is entirely robust and practical by introducing PCA and neural networks.

5 Conclusion

This paper proposes a new neural network based photometric stereo using the illumination planning, as one of the physics based vision approach. The illumination planning determines the cast shadow region and disposes the information of the ill-effected light sources for each point on the object from the output of the neural network.

Shadows depending on local shape has the information to be used for shape recovery but cast shadow cannot be used for photometric stereo. For the implementation, PCA and RBF NN-units are effectively used for the environment. In the illumination planning, the ill-effected region for each point is found from the output of NN-units learned for the calibration sphere images from all light sources. One can know which combination of light source is used to get the relevant result for each point. Finally, each combination of NN-units can be applied to recover the surface normal for each point on the test object effected by cast shadow. It is valuable to stress that the entire approach is non-parametric, 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 pattern of illumination and be made of the same material (i.e, have the same reflectance properties.) The experiment shows the approach is useful and can improve the shape recovery.

As the future subjects, the application of NN to non-uniform test object, and to the problem of correcting interreflection are remained.

Acknowledgements

Support for the work described in this paper was provided by the Artificial Intelligence Research Promotion Foundation, Japan. Support for Woodham's research is provided by the Institute for Robotics and Intelligent Systems (IRIS) and by the Natural Sciences and Engineering Research Council of Canada (NSERC).

References

[Horn, 1975] B.K.P. Horn. Obtaining shape from shading information. *The Psychology of Computer Vision.* (P.H. Winston,ed.), pp.115-155, McGrawHill, 1975.

[Horn, 1977] B.K.P. Horn. Understanding image intensities. *Artificial Intelligence,* vol.8, pp.201-231, 1977.

[Woodham, 1980] R.J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering,* vol.19, pp. 139-144, 1980.

[Woodham, 1994] 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, pp. 3050-3068,1994.

[Healey *et ai,* 1992] G. Healey, S. Shafer, and L. Wolff, eds., *Physics-Based Vision: Principles and Practice (Vol 1 Radiometry, Vol. 2 Color and Vol. 3 Shape Recovery).* Boston, MA: Jones and Bartlett Publishers, Inc., 1992.

[Healey *et ai,* 1994] G. Healey and R. Jain. Physics-based machine vision. *Journal of the Optical Society of America, A,* vol. 11, p. 2922, 1994. (Introduction to special issue).

[Iwahori *et ai,* 1995] Y. Iwahori, A. Bagheri and R.J. Woodham. Neural Network Implementation of Photometric Stereo. In *Proceedings of Vision Interface '95,* pp.81-89, (Quebec, Canada) May 1995.

[Iwahori *et ai,* 1995] Y. Iwahori, R.J. Woodham and A. Bagheri: "Principal Components Analysis and Neural Network Implementation of Photometric Stereo", In *Proceedings of IEEE Workshop on Physics-based Modeling in Computer Vision,* pp.117-125, June 1995.

[Solomon *et ai,* 1992] F. Solomon and K. Ikeuchi. Extracting the Shape and Roughness of Specular Lobe Objects Using Four Light Photometric Stereo. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* pp.466-471,June 15-18, 1992.

[Solomon *et ai,* 1995] F. Solomon and K. Ikeuchi. An illumination planner for convex and concave 1ambertian polyhedral objects. In *Proceedings of IEEE Workshop on Physics-Based Modeling in Computer Vision,* pp.100-107, June 18-19, 1995.

[Chen *et ai,* 1991] S.Chen, C.F.N.Cowan and P.M.Grant. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on Neural Networks,* vol.2, no.2, pp.302-309, 1991.