Journal of King Saud University – Computer and Information Sciences (2012) **24**, 129–136

King Saud University

Journal of King Saud University – Computer and Information Sciences

www.ksu.edu.sa
www.sciencedirect.com

ORIGINAL ARTICLE

Application of neural network in integration of shape from shading and stereo

Sanjeev Kumar [a,*], **Manoj Kumar** [b]

[a] *Department of Mathematics, IIT Roorkee, Roorkee 247 667, India*
[b] *Department of Computer Science, B.B.A. University, Lucknow, India*

Received 7 June 2011; revised 5 February 2012; accepted 8 February 2012
Available online 25 February 2012

KEYWORDS

Disparity;
Function approximation;
Neural network;
Shape from shading;
Stereo vision

Abstract In this paper, a simple and efficient approach is presented for the reconstruction of 3-D surfaces using the integration of shape from shading (SfS) and stereo. First, a new SfS algorithm is derived to obtain the depth-map of a 3-D surface using linear and generalized Lambertian reflectance model. Later, the accuracy of the depth-map is improved by integrating stereo depth data. The stereo sparse depth data are obtained at the points which have higher similarity score in the rectified pair of stereo images. A feed-forward neural network is used to integrate the SfS and stereo depth data due to its strong nonlinear function approximation property. The integration process is based on the correction of 3-D visible surface obtained from SfS using the stereo data. The experiments have been performed on real and synthetic images to demonstrate the usability and accuracy of the approach.

1. Introduction

Shape recovery of objects' surfaces is a special discipline in computer vision. This aims the recovery of object's shape or calculation of depth map (i.e., the distance between the camera sensor and objects in the scene). This has a wide domain of applications like 3-D reconstruction (surgery, architecture etc.), distance measurement of obstacles (robotics, vehicle control, etc.), reconstruction of surfaces of planets from photographs acquired by aircrafts and satellites, etc. Estimating depth information of a 3-D scene from its 2-D stereo images taken from different camera positions is known as stereo reconstruction. One of the drawbacks of stereo reconstruction is that it provides the depth information on sparse data points and hence cannot be used for dense reconstruction.

Shading is a unique cue for recovering the shape of 3-D objects, due to its omni-presence under all illumination conditions. However, most of the existing shape from shading (SfS) algorithms is unable to provide the correct 3-D depth information on the boundary of the surfaces. Moreover these algorithms have problems with variable albedo and spherical surfaces. Hence, the performance of 3-D vision systems can be improved when various sources of information about the 3-D scene like stereo, shading and contour etc. are incorporated.

* Corresponding author. Tel.: +91 1332 285824.
E-mail address: malikfma@iitr.ernet.in (S. Kumar).
Peer review under responsibility of King Saud University.

ELSEVIER Production and hosting by Elsevier

1319-1578
http://dx.doi.org/10.1016/j.jksuci.2012.02.001

2. Background research and contribution

There are mainly four approaches used in SfS viz. minimization, local, propagation and linear. The minimization approach is used by Frankot and Chellappa (1988) in which they have enforced integrability constraint. An efficient propagation approach has been developed by Bichsel and Pentland (1992), in which one can recover the depth directly of a continuous surface. In Lee and Roseneld (1985), local approach has been proposed in which surface is approximated by spherical patches. In Tsai and Shah (1994), an efficient SfS approach has been proposed in which the linearization of Lambertian reflectance map is given in terms of depth.

From the last two decades, attention has been given to integrate shading and stereo vision sources. In Frankot and Chellappa (1988), it has been pointed out that the correspondence between stereo images provides low frequency information which is not available in shading alone, and shading provides information not available from either sparse or low resolution stereo correspondences. Therefore, a game-theoretic approach has been proposed to integrate these vision modules (Bozma and Duncan, 1994). An edge based stereo method for integrating stereo and shading has been proposed in Bulthoff and Mallot (1988). In Ikeuchi (1987), a dual photometric stereo system has been proposed, where two sets of images with different viewing directions are used to generate a pair of surface orientation maps. In Chiradia et al. (1989), a new scheme has been proposed in which sparse depth map provides a first estimate of surface shape. A local SfS method is applied to one of the images to get an estimate of surface orientation. The result of this integrated approach is a dense depth map. In Zheng and Kishino (1992), a new method has been given in which different vision cues obtained from a rotating object into a 3-D model have classified these vision cues in such a way that different methods could be applied to different cues. A new strategy for combining shape from shading and stereo has been given in Cryer et al. (1995). They have kept and amplified the low frequency information obtained from the stereo, and added them with the amplified high frequency information resulting from the shape from shading. Recently in Haines and Wilson (2007), a framework has been developed for the integration of depth and orientation information using Gaussian–Markov random field.

In the existing literature, the main emphasis for the integration of stereo and shading was given to the traditional surface generation techniques like interpolation, iterative method for surface modeling, etc. These algorithms use stereo vision for providing the proper values to the initial and the boundary conditions of the SfS problem (Bae and Behabib, 2003; Jin et al., 2000). Hence, there is a possibility for propagating the error from stereo vision to the solution of SfS problem (Banerjee et al., 1992). In this work, stereo vision and SfS are used as constraints on the depth map information simultaneously. A multilayer feed-forward neural network has been used for correcting the depth map obtained from SfS with the help of available stereo sparse data. The stereo sparse data are obtained at the points, which have higher similarity score in the rectified stereo pairs. From the trained network, a significantly improved depth map is obtained on the boundary of the surface which is erroneous on using shape from shading module alone.

In the literature, quite a few algorithms exist to integrate stereo and SfS using neural networks (Mostafa et al., 1999a,b). In Mostafa et al. (1999a), the differences between the depths obtained with stereo (on the sparse points) and SfS have been obtained and an error surface has been fitted using this difference data. Finally, this surface has been used to correct the visible surface obtained from shape from shading. Extended Kalman filter based learning has been used for the surface fitting. The weakness of this method was that any error in the surface fitting (since the surfaces are having the free form shapes) propagates in the final results (obtained by the integration of stereo and SfS). In Mostafa et al. (1999b), the range data have been used instead of sparse stereo data using the similar integration strategy as in Mostafa et al. (1999a). Our proposed approach includes the following steps:

- Training of neural network by using the SfS data on sparse points (where stereo data are presented) as input and the stereo data as the output.
- The trained network has been used to obtain the final depth map by using the dense depth-map obtained with SfS as the input of the trained network.

Moreover, a new SfS algorithm is used to obtain the depth-map of a 3-D surface using linear and generalized Lambertian reflectance model. We are using linear approach as (Tsai and Shah, 1994), together with generalized Lambertian reflectance map instead of Lambertian reflectance map given by Oren and Nayar (1993). Here, we have adopted an approach proposed by Ferrari and Stengel (2005) for the training of neural network. In this approach, the adjustable parameters or weights are determined by solving linear systems of equations for the matching of input-output and gradient information almost exactly.

3. Stereo reconstruction

3.1. Camera model and stereo setup

The camera model used in this paper is a well-known pinhole camera which describes the perspective projection model. From the mathematical point of view, the perspective projection is represented by the projection matrix P of size 3×4, which makes correspondence from 3-D point $W = [x_w\ y_w\ z_w 1]^T$ to 2-D image points $m = [X\ Y\ 1]^T$ and λ is a scale factor (an arbitrary positive scaler) that represents the homogeneous coordinate system:

$$\lambda m = PW \quad \text{with} \quad P = A[R|t] \tag{1}$$

where the projection matrix P is factorized into rotation R, translation t and the intrinsic matrix A.

The imaging setup using two cameras is shown in Fig. 1(a). Let I_1 and I_2 be the first and second image planes of the pair of cameras C_1 and C_2 respectively. The proposed stereo imaging system is designed to take care that a point W in 3-D space is viewed by both the cameras and the orientations of both of the cameras are not necessary to be parallel. Fig. 1(b) shows the rectified pair of stereo images.

3.1.1. SSD measure for disparity estimation

The sum of square differences (SSD) measure has been used to estimate the disparity between the pair of stereo images. For

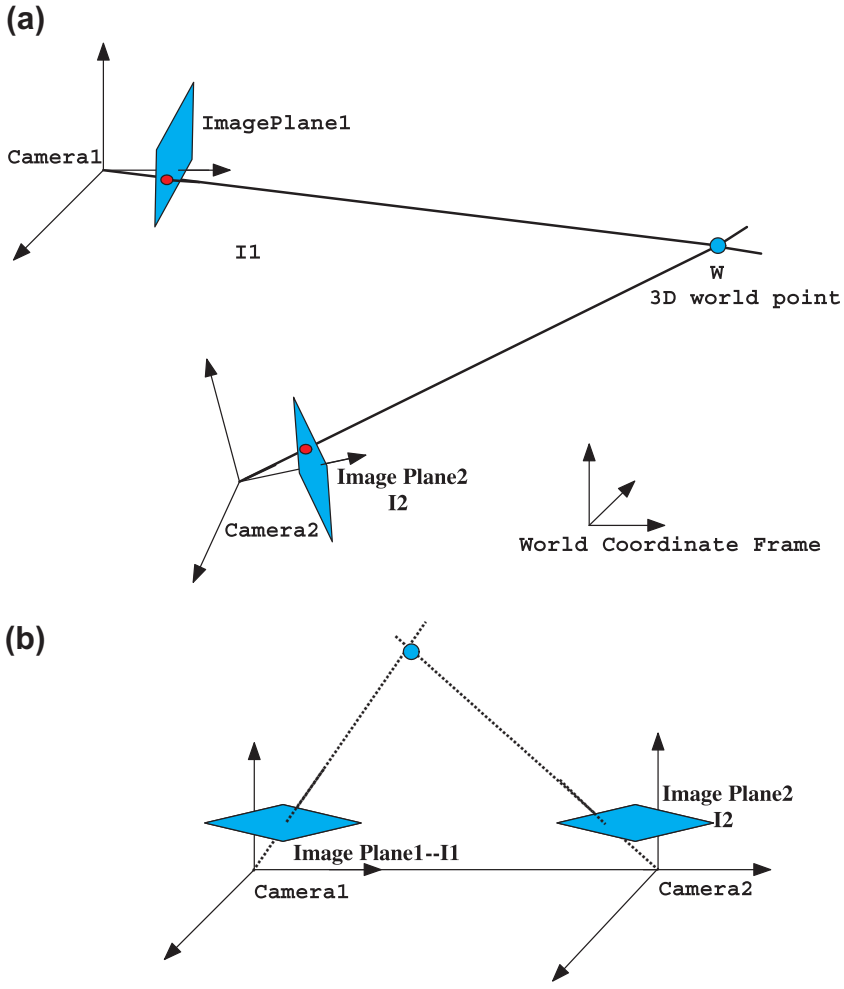

Figure 1 A rectified stereo imaging system.

each pixel in the left image (reference image I_l) from the rectified stereo pair, similarity scores are computed by comparing a fixed, small window of size 3×3 centered on the pixel to a window in the right image (I_r), shifting along the corresponding horizontal scan line. Windows are compared through the normalized SSD measure, which quantifies the difference between the intensity patterns:

$$C(x,y,d) = \frac{\sum_{(\xi,\eta)}[I_l(x+\xi,y+\eta) - I_r(x+d+\xi,y+\eta)]}{\sqrt{\sum_{(\xi,\eta)} I_l(x+\xi,y+\eta)^2 \sum_{(\xi,\eta)} I_r(x+\xi,y+\eta)^2}} \tag{2}$$

where $\xi \in [-n,n]$ and $\eta \in [-m,m]$. The disparity d estimate for pixel (x,y) is the one that minimizes the SSD error:

$$d_0(x,y) = \arg\ \min\ C(x,y,d) \tag{3}$$

However we can observe that squared differences need to be computed only once for each disparity, and the sum over the window need not be recomputed from scratch when the window moves by one pixel. A threshold value σ has been considered which chooses the pixels according to $C(x,y,d) < \sigma$. Using this condition, disparity values have been obtained on sparse points but these disparities are accurate enough.

Once we obtain the disparity d for the matching points (having high similarity measure value) in left and right images, the depth value Z_s has been computed using the following formula

$$Z_S = f\frac{b}{d} \tag{4}$$

where b is the baseline distance in stereo vision system. The depth information given by Z_S is sparse as well as more accurate if we compare this to the depth map, which is obtained by SfS process.

4. Shape from shading

Shape from Shading deals with the recovery of 3-D shape from a single shaded image by exploiting shading information contained in the image. To recover a 3-D shape from its image, it is necessary to know how the images are formed. Various models are proposed for the image formation on the basis of material property of object surface and light conditions. In SfS problems, most widely used image formation model is Lambertian reflectance model due to its simplicity in use and almost fair applicability in approximating most of the object

surfaces in the real world. In Lambertian model, the gray level in image depends on the light source direction and surface normals. Thus, the image brightness is the function of surface shape and light source direction. The recovered shape can be represented in several ways: depth $Z(x,y)$, surface normal (n_x,n_y,n_z) or surface gradient (p,q). The depth can be considered as the relative surface height above the xy plane. The surface normal is the orientation of a vector perpendicular to the tangent plane on the surface object. The surface gradient $(p,q)=\left(\frac{\partial Z}{\partial x},\frac{\partial Z}{\partial y}\right)$ is the rate of change of depth in x and y directions. The surface slant ϕ and tilt θ, are related to the surface normal as $(n_x,n_y,n_z)=(l\sin\phi,l\sin\phi\sin\theta,l\cos\phi)$, where l is the magnitude of surface normal. The unit surface normal $(\hat{n})$ and surface gradient (p,q) are related as follows:

$$\hat{\mathbf{n}}=\frac{(-p,-q,1)}{\sqrt{1+p^2+q^2}} \tag{5}$$

If we assume that the viewer and the light sources are far from the object, then we can introduce the reflectance map, as a means of specifying the dependence of brightness on surface orientation. If we elect to use the unit surface normal $\hat{n}$ as a way of specifying surface orientation, then the brightness can be computed as a function of orientation in terms of $R(\hat{n})$. If we use p and q instead of $\hat{n}$, then it can be computed in the form of $R(p,q)$. The general solution of shape from shading problem is based on the so called image irradiance equation which relates image irradiance to scene radiance:

$$E(x,y)=R(\hat{n}(x,y)) \quad \text{or} \quad E(x,y)=R(p,q) \tag{6}$$

where $E(x,y)$ is the image irradiance at the point (x,y), and $R(\hat{n}(x,y))$ is the radiance of surface patch with unit normal $\hat{n}(x,y)$, which can also be written in terms of surface gradient (p,q).

4.1. Lambertian reflectance map

The most widely used reflectance map in SfS is Lambertian reflectance map. Let $\hat{n}$ and $\hat{s}$ be the unit surface normal to the object surface and the unit illuminates vectors respectively and given as follows

$$\hat{n}=\frac{(-p,-q,1)}{\sqrt{1+p^2+q^2}}, \quad \hat{s}=\frac{(-p_s,-q_s,1)}{\sqrt{1+p_s^2+q_s^2}} \tag{7}$$

then the scene radiance of the surface patch is given by their scalar product, in the following form.

$$R(p,q)=\rho\frac{1+pp_s+qq_s}{\sqrt{1+p^2+q^2}\sqrt{1+p_s^2+q_s^2}} \tag{8}$$

where ρ is the albedo of the surface and $0\leqslant\rho\leqslant 1$.

4.2. General Lambertian reflectance map

There are several real world objects, for which the Lambertian model can prove to be an inaccurate approximation to the diffuse components. The brightness of a Lambertian surface is independent of viewing direction, however the brightness of a rough diffuse surface increases as the viewer approaches the source direction. To deal with this problem, (Oren and Nayar, 1993) have developed a comprehensive model which is the generalization of Lambertian reflectance map. The scene radiance in general Lambertian reflectance is given as follows:

$$L_r(\theta_r,\theta_i,\phi_r-\phi_i;\sigma)=\frac{\rho}{\pi}\cos\theta_i\{A+B\max[0,\cos(\phi_r-\phi_i)]\sin\alpha\sin\beta\} \tag{9}$$

where

$$A=1.0-0.5\frac{\sigma^2}{\sigma^2+0.33} \quad \text{and} \quad B=0.45\frac{\sigma^2}{\sigma^2+0.09}$$

The angles θ_i and θ_r are the tilt angles of incidence and reflection, while ϕ_i and ϕ_r are the slant angles for illumination and viewer. σ is used for roughness which is the standard deviation of normal distribution. However, roughness is supposed to be normally distributed with zero mean. The value of σ is small for less rough surface and more for more rough surfaces. In experiments, we have taken the value of σ as 0.1.

Here, the product of $\sin\alpha$ and $\sin\beta$ will be equivalent to the product of $\sin\theta_i$ and $\sin\theta_r$, since the angles α and β are given as

$$\alpha=\max(\theta_r,\theta_i) \quad \text{and} \quad \beta=\min(\theta_r,\theta_i)$$

if surface normal and light source both are specified in the viewer oriented system then we specify the incident and excitant direction in the local system such that the excitant ray lies along the z-axis in the direction toward to the viewer. Therefore

$$\sin\theta_r=\sqrt{\frac{p^2+q^2}{1+p^2+q^2}}, \quad \sin\theta_i=\sqrt{1-\frac{(1+pp_s+qq_s)^2}{(1+p^2+q^2)(1+p_s^2+q_s^2)}}$$

$$\cos\theta_i=\frac{(1+pp_s+qq_s)}{\sqrt{(1+p^2+q^2)}\sqrt{(1+p_s^2+q_s^2)}}$$

and

$$\cos(\phi_r-\phi_i)=\frac{p^2+q^2-pp_s-qq_s}{\sqrt{(p^2+q^2)(1+p^2+q^2)(1+p_s^2+q_s^2)-(1+pp_s+qq_s)^2}}$$

On substituting the values of these trigonometrical identities in terms of p and q, the image irradiance equation for the general Lambertian reflectance map can be written as:

$$E(x,y)=L_r(p,q) \tag{10}$$

In the above qualitative model, the inter-reflection factor is ignored. This model can be viewed as a generalization of the Lambertian model, which becomes a Lambertian model in the case of $\sigma=0$. Here, $E(x,y)$ is the image irradiance at the point (x,y), while $L_r(p,q)$ is the radiance of a surface patch with unit normal $\hat{n}$ at the point (x,y). The image irradiance equation is a nonlinear first order partial differential equation. Without the loss of generality, we are assuming $\rho=1$ in further derivation. As given in the survey (Zhang et al., 2000), the linear approach (Tsai and Shah, 1994) gives better results with less time complexity. In order to linearize the reflectance map, the linear approximations of p and q in terms of Z are given as

$$p=\frac{\partial Z}{\partial x}=Z(x,y)-Z(x-1,y) \tag{11}$$

$$q=\frac{\partial Z}{\partial y}=Z(x,y)-Z(x,y-1) \tag{12}$$

Now (10) can be written as:

$$E(x,y)-L_r(Z(x,y)-Z(x-1,y),Z(x,y)-Z(x,y-1))=0 \tag{13}$$

which is equivalent to

$$f(E(x,y), Z(x,y), Z(x-1,y), Z(x,y-1)) = 0 \tag{14}$$

Now, for a fixed point (x,y) of a given image E of size $N \times N$, by taking the Taylor series expansion up to the first order of the function f about the given depth $Z^{(n-1)}$,we get a linear system of N^2 equations. Again, using Jacobi iterative method for solving this system of equations, we get the following iterative formula for computing the depth value Z at each point (x,y) of the given image.

$$Z^{(n)}(x,y) = Z^{(n-1)}(x,y) + \frac{-f(Z^{(n-1)}(x,y))}{\frac{\partial}{\partial Z(x,y)} f(Z^{(n-1)}(x,y))} \tag{15}$$

and the derivative $\frac{df(Z(x,y)}{dZ(x,y)}$ can be calculated as follows:

$$\frac{df(Z(x,y)}{dZ(x,y)} = -\frac{1}{\pi}[AU' + BU'V + BUV'] \tag{16}$$

Let us assume

$$a = \frac{1 + pp_s + qq_s}{(1+p^2+q^2)^{(\frac{3}{2})}}$$

$$b = \sqrt{(p^2+q^2)(1+p^2+q^2)(1+p_s^2+q_s^2) - (1+pp_s+qq_s)^2}$$

$$a' = \frac{(1+p^2+q^2)^{(\frac{3}{2})}(p_s+q_s) - 3(p+q)(1+p^2+q^2)^{(\frac{1}{2})}(1+pp_s+qq_s)}{(1+p^2+q^2)^3}$$

$$b' = \frac{(p+q)(1+p_s^2+q_s^2)(1+2p^2+2q^2) - (p_s+q_s)(1+pp_s+qq_s)}{\sqrt{(1+p_s^2+q_s^2)(p^2+q^2)(1+p^2+q^2) - (1+pp_s+qq_s)^2}}$$

$$c = p^2+q^2-pp_s-qq_s \text{ and } c' = 2p+2q-p_s-q_s$$

then the terms U, U', V and V' involved in (16) are given as

$$U = \frac{1}{(1+p_s^2+q_s^2)}[ab], \quad U' = \frac{1}{(1+p_s^2+q_s^2)}[ab'+ba'],$$

$$V = \frac{c}{b} \quad \text{and} \quad V' = \frac{bc'-cb'}{b^2}$$

By assuming the initial value of $Z^{(0)}(x,y) = 0$ for all pixels, the depth can be iteratively found using (15). We calculate $f(Z^{(n-1)}(x,y))$ and $f'(Z^{(n-1)}(x,y))$ at each iteration. The iterative equation (15) will not work when $f'(Z^{(n-1)}(x,y))$ is zero, hence to avoid this difficulty, we have introduced a constant C which is approximately equal to $f'(Z^{(n-1)}(x,y))$ but not zero.

5. Integration of shape from shading from stereo data

Integration of stereo and SfS data is performed by using a feed-forward neural network. The main aim of integration process is correction of the depth map obtained from SfS on the boundary of the surface and the resolution of ambiguity of 3-D visible surface up to some extent. The integration is considered as an accuracy improvement process or a highly nonlinear function approximation process so that the function improves the accuracy of depth map data. Consider a nonlinear input output mapping defined by the function relationship $v = h(u)$, where the vector u is the input and the vector v is the output. The mapping function $h(\cdot)$ is unknown and highly nonlinear. Now for a known set of input-output values (u_i, v_i); $i = 1,2,\ldots,n$, the problem is to find the function $H(\cdot)$ that approximates $h(\cdot)$ over all input. That is,

$$\|H(u) - h(u)\| < \epsilon \quad \text{for all} \quad u \tag{17}$$

where ϵ is a small threshold value. This function approximation problem can be solved by using neural network with u_i playing the role of input vector and v_i play the role as desired output.

5.1. Network's training

According to (Ferrari and Stengel, 2005), the computational neural network matches the input–output training set, exactly if, the matrix **S** of sigmoid functions evaluated at input elements is full rank. The matrix **S** forces to be square by taking the number of nodes in the hidden layer equal to the input-output training pair and a system of linear algebraic equations can be obtained by training data together with weights. If the system is full rank, then a unique solution always exists. The input parameters affect the solution of the output weight equations only through the input-to-node values determining the nature of **S**. Thus, the required weights are not unique and need to be chosen only to assure that **S** is full rank. In this context, the following algorithm is adopted for the network training which provides an almost exact matching of input-output data.

(1) Specify the training set (u, v) of input-output pairs.
(2) Set $w_{ij} = \gamma r_{ij}$, where r_{ij} is chosen from a normal distribution with zero mean and unit variance, i.e., obtained using a random number generator, γ is a user define scalar that can be adjusted to obtain input-to-node that does not saturate the sigmoid.
(3) Calculate

$$d = -\text{diag}(UW^T) \quad \text{and} \quad p^k = Wu^k + b$$

where U is a matrix composed of all the input vectors in the training set and b is a bias.
(4) Check whether the matrix S of sigmoid functions which are evaluated at input-to-node values p_i^k is singular or not.
(5) If it is singular go to step 2 otherwise compute the network output $q = S^{-1}(u)$.
(6) Check whether gradient tolerance has met. **If not** $\rightarrow$ compute

$$W_i = \frac{2}{q_i}(c^k - \sum_{l \neq i} q_l \sigma'(n_l^k) w_l),\ d_i = -u^k w_i^T \quad \text{and} \quad p_i = d_i + U w_i^T$$

where $\sigma'(\cdot)$ denotes the derivative of the sigmoid function with respect to its scaler input, and $c^k = W^T(q \otimes \sigma'(p^k))$. Again go to step 4.
If yes $\rightarrow$ network is trained. In this way, we get a good convergence of the network training.

5.2. Integration process

For the integration process, the network has been trained using the shading data available on the points where stereo depth map exists as input data and the stereo depth values as output data, i.e., a set of input-output values (Z_T, Z_S). Once the neural network is trained upto a given threshold error between desired and network output, the complete set of depth map obtained from SfS is given as network input and the final depth map is obtained Z_f as network output.

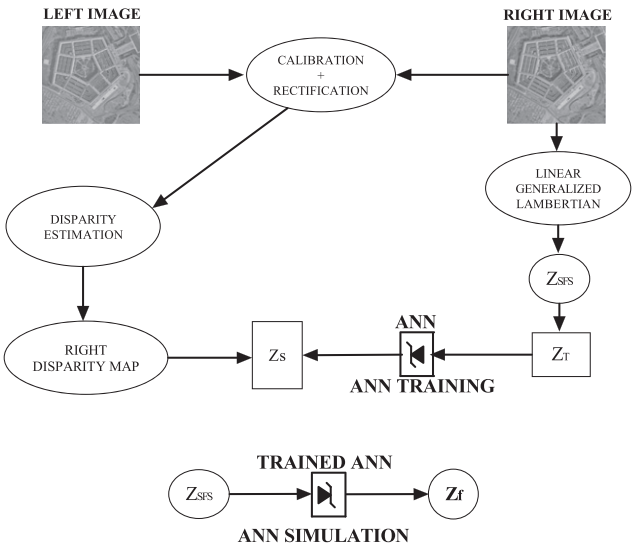

Figure 2 Block diagram of overall integration process. Z_T denotes the set of depth information obtained with SfS at sparse points where the stereo depth Z_s is calculated. Z_f is the output of the trained neural network (the final depth obtained with the integration framework).

In our network, each output in a layer is connected to each input in the next layer. In this case, all the neurons in the hidden layer have the same transfer function, with a sigmoidal nonlinearity. Also, there is no feedback between layers, so the effect of the feed-forward neural net topology is to produce a nonlinear mapping between the input nodes and the output nodes. The model that we have used consists of two input neurons (one depth map data obtained from SfS process and the other is a bias), one hidden layer and an output neuron corresponding to the depth map obtained from stereo process. We train the network on the range of input and output, such that the network could train and give a more accurate depth map for any depth map obtained using SfS. Fig. 2 gives a block diagram of the proposed integration of stereo and SfS.

6. Results and discussions

The experiments have been conducted on real as well as synthetic pairs of stereo images. A feed forward neural network is used to integrate the data obtained from stereo and SfS. It can be seen from presented results that the quality of 3-D reconstruction of visible surfaces have been improved basically due to the integration of stereo and SfS. Fig. 3 show the 3-D shape recovery using the synthetic stereo images of Mozart and Vase surfaces. The results shown in this figure indicate the importance of integrating SfS and stereo. The integration process has been performed using three different methods: linear interpolation, nonlinear interpolation and neural network function approximation (sigmoidal interpolation). To obtain the results for linear and nonlinear interpolations, we have

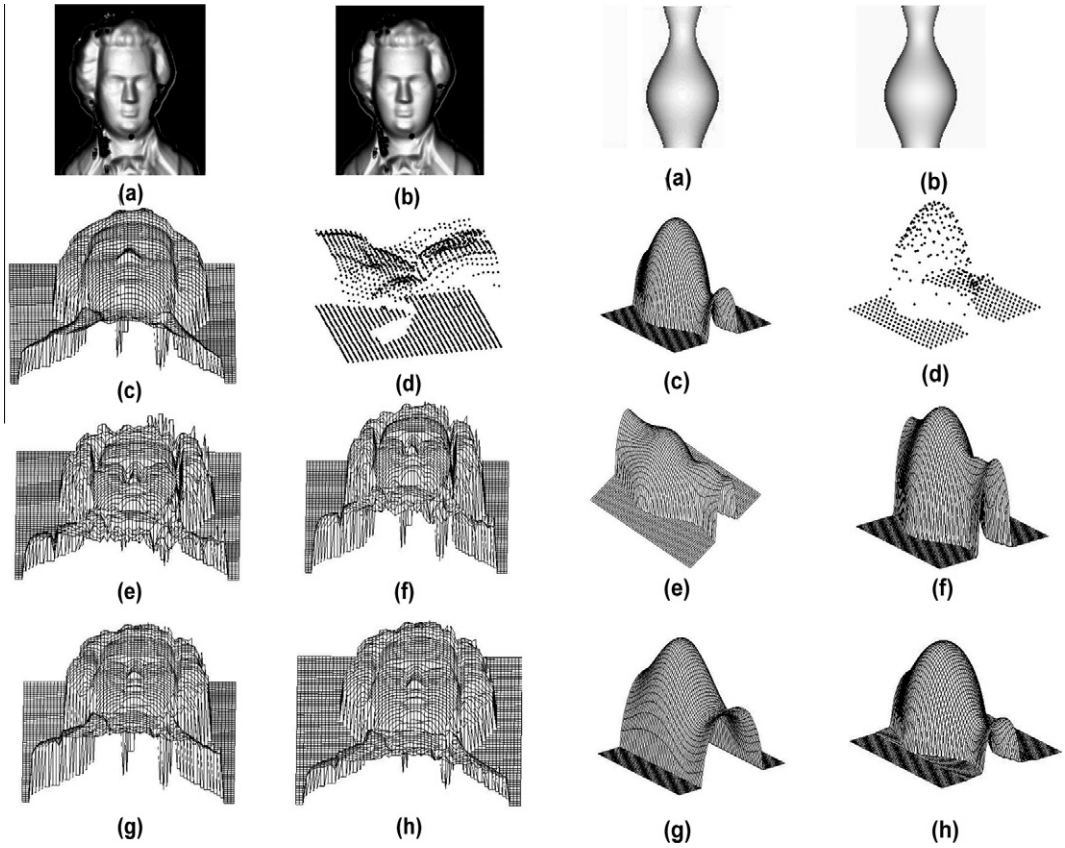

Figure 3 Results for depth using synthetic images of Mozart (left) and Vase (right). Stereo images (a)–(b), (c) Ground truth depth, (d) Sparse stereo depth, (e) depth using SfS, (f) depth using linear interpolation, (g) depth using nonlinear interpolation, (h) depth using neural network based integration.

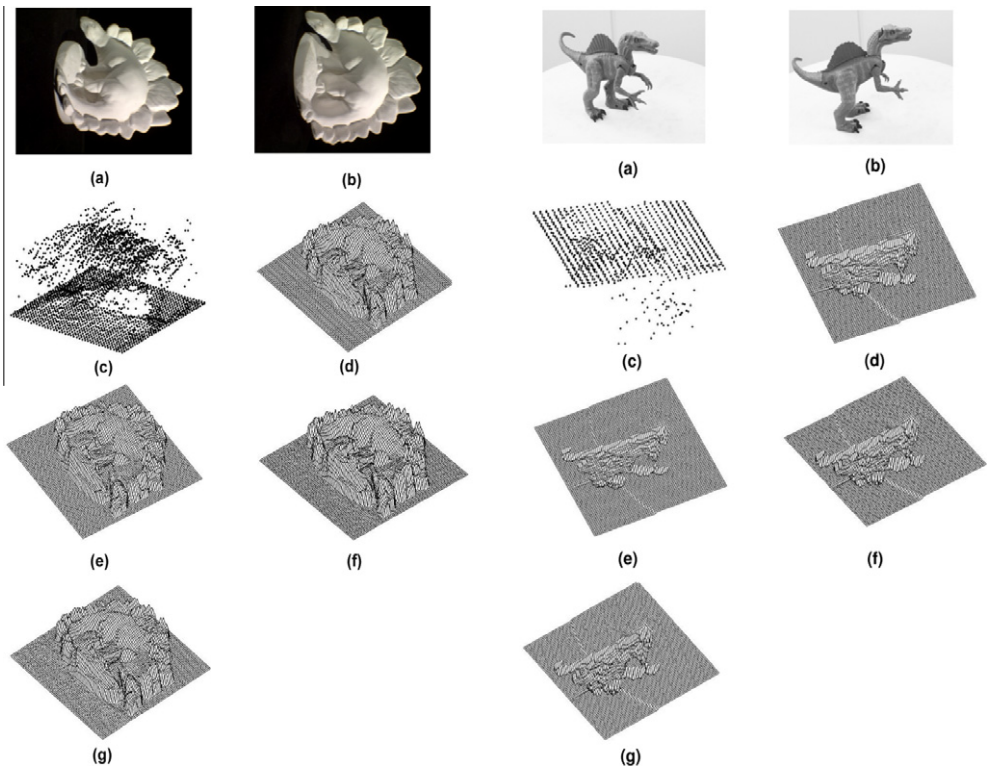

Figure 4 Results for depth using real images of Deno (left) and Dinosaur (right). Stereo images (a)–(b), (c) Sparse stereo depth, (d) depth using SfS, (e) depth using linear interpolation, (f) depth using nonlinear interpolation, (g) depth using neural network based integration.

used the *interp* function in MATLAB with linear and spline type interpolation methods. The experimental results using real images are shown in Fig. 4. The experiments have been performed on Deno and Dinosaur images pairs. The results shown in these figures indicate that the results which are obtained from integrating the stereo and SfS are better than the results obtained from SfS alone.

In Figs. 3 and 4, We have analyzed the results qualitatively by considering the 3-D plots of the depth map. We have analyzed also the error in results for the synthetic images. There are several ways to report the error behavior. We have reported the error in the following two ways:

- **Mean of depth error**: The obtained depth values have been normalized in the range of ground truth. The mean of absolute difference between the ground truth and obtained depth values has been calculated.
- **Standard deviation of depth error**: The obtained depth values have been normalized in the range of ground truth. The standard deviation of absolute difference between the ground truth and obtained depth values has been calculated.

From the Table 1, it is concluded that the proposed algorithm is accurate enough for the reconstruction of 3-D surfaces and thus the integration of SfS and stereo provides more accurate results. Furthermore, the integration based on the neural network is more accurate when compared to the linear and nonlinear interpolations based integration in the proposed framework. A comparison between the two neural network based approaches (proposed and the other is using the extended Kalman filter (EKF) based training as in Mostafa et al., 1999a) has been given. We observe that the proposed technique gives marginally better results in terms of the mean error, while a noticeable variation can be seen in the standard deviation. The reason behind this variation may be the propagation of the error of the surface fitting into the final reconstruction results which were obtained by the integration.

Further, we have analyzed the effect of changing the network parameters on final depth results. For this, we have considered two parameters, learning rate and the number of epoch. The effect of these network parameters in the case of Mozart reconstruction is given in Tables 2 and 3. It can be seen

Table 1 Errors in the final reconstruction using different methods for synthetic images.

Methods	Mean error		Stand. deviation error	
	Mozart	Vase	Mozart	Vase
SfS alone	7.7	2.1	4.5	2.8
Integration using Linear Interpolation	4.9	1.5	3.0	1.0
Integration using Nonlinear Interpolation	3.7	1.3	2.6	0.9
Integration using Neural Network	2.5	0.7	1.8	0.8
Integration using EKF based Neural Network	2.6	0.85	1.8	.95

Table 2 Analysis for different value of learning rate in the integration based reconstruction with fixed 5000 epoch limit.

Learning rate	Mean error
0.30	3.54
0.20	3.53
0.10	2.78
0.05	2.51
0.01	4.70

Table 3 Analysis for different Epoch limit in the integration based reconstruction with a constant leaning rate 0.05.

Epoch limit	Mean error
100	9.87
500	7.23
1000	5.44
5000	2.51
10000	2.51

from the Table 3 that after the 5000 epoch limit there is no improvement in the integration based reconstruction results. The network architecture has been chosen as per the method given in Ferrari and Stengel (2005) in which network contains one hidden layer and the number of nodes in this layer can be fixed as per the dimension of the input data.

7. Conclusion

The objective of this paper is to present a framework for the integration of SfS and stereo depth data. The SfS data have been obtained using a modified linear generalized Lambertian reflectance model. The integration process is performed using a feed-forward neural network. It is observed that the integration of stereo and SfS greatly improves the 3-D visible surface reconstruction obtained from SfS only and also produces 3-D visible surfaces representation with nearly accurate metric representation. Furthermore, the approximation based on neural network is more accurate when compared to the linear and nonlinear interpolations based approximation in this framework. Moreover, the results are slightly better when compared to the extended Kalman filter based neural network approach. The effect of the network parameters on the final results have been analyzed and best tuned parameters have been used for obtaining the results for the integration framework.

References

Bae, K.Y., Behabib, B., 2003. A hybrid scheme incorporating stereo-matching and shape from shading for spatial object recognition. Journal of Engineering Manufacture (IMechE) 217, 1533–1542.

Banerjee, S., Sastry, P.S., Venkatesh, Y.V., 1992. Surface reconstruction from disparate shading: an integration of shape from shading and stereopsis. International Conference on Pattern Recogination, 141–144.

Bichsel, M., Pentland, A.P., 1992. A simple algorithm for shape from shading. In: IEEE proceedings of CVPR, pp. 459–465.

Bozma, H.I., Duncan, J.S., 1994. A game-theoritic approach to integration of modules. IEEE Transactions on Pattern Analysis and Machine Applications 16 (11), 1074–1086.

Bulthoff, H.H., Mallot, H.A., 1988. Integration of depth modules: stereo and shading. Journal of Optical Society of America A 5 (10), 1749–1758.

Chiradia, M.T., Distante, A., Stella, E., 1989. Three dimensional surface reconstruction integrating shading and sparse stereo data. Optical Engineering 28 (9), 935–942.

Cryer, J., Tsai, P., Shah, M., 1995. Integration shape from shading and stereo. Pattern Recognition 28 (7), 1033–1043.

Ferrari, S., Stengel, R.F., 2005. Smooth function approximation using neural networks. IEEE Transaction of Neural Networks 16 (1), 24–38.

Frankot, R.T., Chellappa, R., 1988. A method for enforcing integrability in shape from shading algorithms. IEEE Transactions of pattern Analysis and Machine Intelligence 10, 439–451.

Haines, T.S., Wilson, R.C., 2007. Integrating stereo with shape-from-shading derived orientation information, British Machine Vision Conference, vol. 2, pp. 910–919.

Ikeuchi, K., 1987. Determining a depth map using a dual photometric stereo. International Journal of Robotic Research 6 (1), 15–31.

Jin, H., Yezzi, A., Soatto, S., 2000. Stereoscopic shading: integrating multi-frame shape cues in a variational framework. In: IEEE International Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 169–176.

Lee, C.H., Roseneld, A., 1985. Improved methods of estimating shape from shading using the light source coordinate system. Artificial Intelligence 26, 125–143.

Mostafa, M., Yamani, S.M., Farag, A.A., 1999. Integrating stereo and shape from shading. In: IEEE International Conference on Image Processing, vol. 3, pp. 130–134.

Mostafa, M., Yamani, S.M., Farag, A.A., 1999. Integrating shape from shading and range data using neural networks. In: IEEE International Conference on Computer Vision and Pattern Recognition, 1063-6919/99.

Oren, M., Nayar, S.K., 1993. Diffuse reflectance from rough surfaces. In: Proceedings of ICCV, pp. 763–764.

Tsai, P.S., Shah, M., 1994. Shape from shading using linear approximation. Journal of Image and Vision Computing 12 (8), 487–498.

Zhang, R., Tsai, P.S., Cryer, J.E., Shah, M., 2000. Shape from shading: a survey. IEEE Transactions of Pattern Analysis and Machine Intelligence 21, 690–706.

Zheng, J.Y., Kishino, F., 1992. Verifying and combining different visual cues into a 3D model. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 777–780.