

# Depth Recovery of Complex Surfaces from Texture-less Pairs of Stereo Images

S. Kumar\*, M. Kumar<sup>†</sup>, N. Sukavanam<sup>†</sup>, R. Balasubramanian<sup>†</sup> and R. Bhargava<sup>†</sup>

\* *Department of Mathematics and Computer Science, University of Udine, Via Della Scienze 206, Udine-33100, Italy*

<sup>†</sup> *Department of Mathematics, Indian Institute of Technology Roorkee, Roorkee-247 667, India*

Received 29 December 2008; accepted 29 April 2009

---

## Abstract

In this paper, a novel framework is presented to recover the 3D shape information of a complex surface using its texture-less stereo images. First a linear and generalized Lambertian model is proposed to obtain the depth information by shape from shading (SfS) using an image from stereo pair. Then this depth data is corrected by integrating scale invariant features (SIFT) indexes. These SIFT indexes are defined by means of disparity between the matching invariant features in rectified stereo images. The integration process is based on correcting the 3D visible surfaces obtained from SfS using these SIFT indexes. The SIFT indexes based improvement of depth values which are obtained from generalized Lambertian reflectance model is performed by a feed-forward neural network. The experiments are performed to demonstrate the usability and accuracy of the proposed framework.

*Key Words:* Function Approximation, Neural Network, Reflectance Model, Scale Invariant Features, Shape from Shading.

---

## 1 Introduction

Shape recovery of object surfaces is a special discipline in computer vision. This aims the recovery of object shapes or calculation of depth, i.e., the distance between the camera sensor and objects in the scene. It has a wide domain of applications like 3D reconstruction (surgery, architecture), distance measurement of obstacles (robotics, vehicle control), reconstructing surfaces of planets from photographs acquired by aircrafts and satellites etc. Shading is a unique cue to reconstruct the 3D surfaces because of its omnipresence property under all illumination conditions. However, most of the SfS algorithms are not able to provide accurate depth-map on the boundary of the surfaces and some of them have problem with variable albedo and spherical surfaces. Therefore, these drawbacks of SfS algorithms can be improved by combining it with other sources of depth information like stereo, shading, features and contour.

There are mainly four approaches used in SfS viz. minimization, local, propagation and linear. The minimization approach [8] has been used by enforcing the integrability constraint. An efficient propagation approach [3] has been developed for recovering the depth information directly for continuous surfaces. The local

---

Correspondence to: <malikdma@gmail.com>

Recommended for acceptance by <Joao Manuel Tavares>

ELCVIA ISSN:1577-5097

Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

approach [15] has been proposed by approximating the required surface with spherical patches. The linear SfS approach [19] have been proposed by linearizing Lambertian reflectance map in terms of depth. In the present work, we are using linear approach as [19], together with generalized Lambertian reflectance map [18] instead of Lambertian reflectance map.

From the last two decades several algorithms have been developed to integrate the various vision modules together. In [8], it has been pointed out that correspondence between stereo image pairs provides low frequency information which is not available in shading alone, and shading provides the high frequency information which is not available from sparse or low resolution stereo correspondences. In this context, a game-theoretic approach [4] has been given to integrate stereo and SfS together. An edge based stereo method for integrating stereo and SfS has been proposed in [5]. In [6], a method for recovering the 3D surfaces has been presented by integrating SfS and stereo data. The stereo data is obtained on few sparse points and used for correcting the SfS data. A method for combining stereo and shading modules by amplifying the low frequency information from the stereo, and add it with the amplified high frequency information from the SfS results has been presented in [7]. An algorithm for determining a depth-map from a pair of surface-orientation maps obtained by a dual photometric stereo has been given in [13]. The surface orientations have been determined using photometric stereo system with three images captured from the same position under different lighting conditions. A neural network based method [17] for integrating stereo and SfS together has been presented in order to improve the 3D reconstruction of visible surfaces of objects from intensity images. A feedforward neural network has been used to fit a surface to the error difference while extended Kalman filter algorithm has been used for the network learning.

Recently, A method for integrating the orientation and depth information have been proposed by using a Gaussian-Markov random field model [10]. Again, a framework for combining stereo and shape-from-shading information by taking account of the local reliability of each shape estimate has been developed [11]. Local estimations of disparity and orientation have been modeled using Gaussian distributions. A Gaussian-Markov random field has been used to represent the disparity-map, taking into account interactions between disparity measurements and surface orientation, and the MAP estimate found using belief propagation. Local estimates of the precision of disparities and surface normal have been found and used to control the process so that the most accurate data source is used in each region. Motion, stereo and photometric stereo have been integrated into a single framework for providing accurate object models from a sequence of frames [20]. A method for viewing invariant gesture recognition by integrating range and intensity information has been given [12]. The intensity image has been used to define the region of interest for the relevant 3D data. This data fusion improves the quality of the range data and hence results in better recognition.

The main problem in depth-map obtained from stereo method is that it is not accurate for the image-pairs having poor texture. The integration of stereo and SfS for these types of image-pairs resulting more error in final depth-map. In this paper, a set of sparse values namely SIFT indexes are used instead of depth-map provide by stereo to correct the SfS data. These SIFT indexes are obtained from the matching of invariant features between stereo images, hence they are more accurate when compare to stereo data which contains calibration error. Some of the existing algorithms which integrate SfS and stereo into one system use stereo vision for the initialization and the boundary conditions are obtained from the shading [1, 14]. However, these types of algorithms may allow to propagate the error from stereo vision to the solution of shape from shading [2]. Here we are using SIFT indexes and SfS depth values are as constraints on the depth-map information simultaneously. A multilayer feed-forward neural network is used to correct SfS depth with the help of SIFT indexes. Neural network is trained based on a modified approach which is inspired by [9]. In this approach, the weight matrix of network is initialized in such a way that the integration process is able to meet the required gradient tolerance.

This paper is structured as follows. In section 2, a linear and generalized SfS algorithm is presented. Section 3 describes the process of generating SIFT indexes. The integration of SfS and SIFT indexes is given in section 4. A brief overview on overall framework is given in section 5. Experimental results are presented in section 6. Finally, the concluding remarks are given in section 7.

## 2 Shape from Shading (SfS)

SfS algorithms deal with the recovery of 3D shape information of an object/scene from its single shaded image by exploiting the shading information contained in the image. To recover a 3D 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  $(p, q) = (\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y})$  is the rate of change of depth in  $x$  and  $y$  directions. The surface slant  $\phi$  and tilt  $\theta$ , 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:

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

If we assume that the viewer and the light sources are far from the object, then we can introduce the reflectance map, 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:

$$E(x, y) = R(\hat{n}(x, y)) \quad \text{or} \quad E(x, y) = R(p, q) \quad (2)$$

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  $(p, q)$ .

### 2.1 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 illuminate vectors respectively and given as follows

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

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

$$R(p, q) = \rho \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2} \sqrt{1 + p_s^2 + q_s^2}} \quad (4)$$

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

### 2.2 General Lambertian Reflectance map

There are several real world objects, for which the Lambertian model can prove to be 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 to the source direction. To deal

with this problem, Oren *et al.* [18] have developed a comprehensive model which is the generalization of Lambertian reflectance map. The scene radiance in general Lambertian reflectance is given as follows:

$$L_r(\theta_r, \theta_i, \phi_r - \phi_i; \sigma) = \frac{\rho}{\pi} \cos \theta_i \{A + B \max[0, \cos(\phi_r - \phi_i)] \sin \alpha \sin \beta\} \quad (5)$$

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  $\theta_i$  and  $\theta_r$  are the tilt angles of incidence and reflection, while  $\phi_i$  and  $\phi_r$  are the slant angles for illumination and viewer.  $\sigma$  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  $\sigma$  is small for less rough surface and more for more rough surfaces. In experiments, we have taken the value of  $\sigma$  as 0.03 radian.

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  $\alpha$  and  $\beta$  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 lie along the  $z$ -axis in the direction toward to the viewer. Therefore

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

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:

$$E(x, y) = L_r(p, q) \quad (6)$$

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 Lambertian model in case of  $\sigma = 0$  (zero). 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 loss of generality, we are assuming  $\rho = 1$  in further derivation. As given in the survey [21], the linear approach [19] 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

$$p = \frac{\partial Z}{\partial x} = Z(x, y) - Z(x - 1, y) \quad (7)$$

$$q = \frac{\partial Z}{\partial y} = Z(x, y) - Z(x, y - 1) \quad (8)$$

Now (6) can be written as:

$$E(x, y) - L_r(Z(x, y) - Z(x - 1, y), Z(x, y) - Z(x, y - 1)) = 0 \quad (9)$$

which is equivalent to

$$f(E(x, y), Z(x, y), Z(x - 1, y), Z(x, y - 1)) = 0 \quad (10)$$

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.

$$Z^{(n)}(x, y) = Z^{(n-1)}(x, y) + \frac{-f(Z^{(n-1)}(x, y))}{\frac{\partial}{\partial Z(x, y)} f(Z^{(n-1)}(x, y))} \quad (11)$$

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

$$\frac{df(Z(x, y))}{dZ(x, y)} = -\frac{1}{\pi} [AU' + BU'V + BUV'] \quad (12)$$

Let us assume

$$\begin{aligned} a &= \frac{1 + pp_s + qq_s}{(1 + p^2 + q^2)^{\frac{3}{2}}} \\ b &= \sqrt{(p^2 + q^2)(1 + p^2 + q^2)(1 + p_s^2 + q_s^2) - (1 + pp_s + qq_s)^2} \\ a' &= \frac{(1 + p^2 + q^2)^{\frac{3}{2}}(p_s + q_s) - 3(p + q)(1 + p^2 + q^2)^{\frac{1}{2}}(1 + pp_s + qq_s)}{(1 + p^2 + q^2)^3} \\ b' &= \frac{(p + q)(1 + p_s^2 + q_s^2)(1 + 2p^2 + 2q^2) - (p_s + q_s)(1 + pp_s + qq_s)}{\sqrt{(1 + p_s^2 + q_s^2)(p^2 + q^2)(1 + p^2 + q^2) - (1 + pp_s + qq_s)^2}} \\ c &= p^2 + q^2 - pp_s - qq_s \quad \text{and} \quad c' = 2p + 2q - p_s - q_s \end{aligned}$$

then the terms  $U, U', V$  and  $V'$  involved in (12) are given as

$$U = \frac{1}{(1 + p_s^2 + q_s^2)} [ab], \quad U' = \frac{1}{(1 + p_s^2 + q_s^2)} [ab' + ba'], \quad V = \frac{c}{b} \quad \text{and} \quad V' = \frac{bc' - cb'}{b^2}$$

By assuming the initial value of  $Z^{(0)}(x, y) = 0$  for all pixels, the depth can be iteratively found using (11). We calculate  $f(Z^{(n-1)}(x, y))$  and  $f'(Z^{(n-1)}(x, y))$  at each iteration. The iterative equation (11) 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 equals to  $f'(Z^{(n-1)}(x, y))$  but not zero.

### 3 Generation of SIFT Indexes

The process to identify locations in image scale space that are invariant with respect to image translation, scaling and rotation is based on the localization of a key. This task can be performed in the following steps:

1. Perform the convolution operation on input image  $I$  with the Gaussian function with variance  $\sigma = \sqrt{2}$ . Let this operation gives an image  $I_1$ .
2. Repeat the step 1 on image  $I_1$  to get a new image  $I_2$ .
3. Subtract image  $I_2$  from image  $I_1$  to obtain the difference of Gaussian function as  $\sqrt{2}$ .
4. Re-sample the image  $I_2$  using bilinear interpolation with a pixel spacing of 1.5 in each direction. The 1.5 spacing means that each new sample will be a constant linear combination of 4-adjacent pixels. From it, we generate a new pyramid level.
5. Determine the maxima and minima of this scale-space function by comparing each pixel in the pyramid to its neighbors.

6. Select key locations at maxima and minima of a difference of Gaussian function applied in scale space.

The scale invariant features can be detected from the locations of these keys. These features are detected on the exact key locations as well as their small neighborhood so that these can be used to perform reliable matching between different views of an object or scene. These features are invariant not only to image orientation but also to image scale, and provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination.

### 3.1 SIFT Matching

First, invariant features are extracted from the left image and stored in a database. Then the features extracted from right image are matched by individually comparing each feature to this database and finding candidate matching features based on Euclidean distance of their feature vectors. We have performed features matching between stereo pair using the process given in [16].

The random sample consensus (RANSAC) is used to remove the outliers from the pairs of matching points obtained from SIFT matching. More generally speaking, the basic assumption is that the data consist of inliers, i.e., data points which can be explained by some set of model parameters, and outliers which are data points that do not fit the model. In addition, the data points are subject to noise. An advantage of RANSAC is its ability to robustly estimate the model parameters. It finds reasonable estimates of the parameters even if a high percentage of outliers are present in the data set.

### 3.2 SIFT Indexes

The index for a matching pair of features is computed based on the absolute differences of their horizontal location in left and right images. Suppose that a point  $(P_l, P_r)$  is a pair of matching features, where  $P_l(x_l, y_l)$  is a point in left image while  $P_r(x_r, y_r)$  is the corresponding point of  $P_l$  in right image. For a rectified stereo pair, the values  $y_l$  and  $y_r$  will be same. The SIFT index for this matching pair is defined as

$$S_P = |x_l - x_r| \quad (13)$$

From the above equation a set of sparse SIFT indexes  $S_i : i = 1, 2, \dots, n$  is computed for the SIFT matching pairs  $\Pi_i : i = 1, 2, \dots, n$ . The image pair must be rectified before obtaining the SIFT indexes.

## 4 Integration of SIFT Indexes and SfS

Integration of SIFT indexes and SfS data is performed by using a feed-forward neural network. The main aim of integration process is the correction of the depth-map obtained from SfS and the resolution of ambiguity of 3D visible surface upto 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  $W = f(u)$ , where the vector  $u$  is the input and the vector  $W$  is the output. The mapping function  $f(\cdot)$  is unknown and highly nonlinear. Now for a known set of input-output values  $(u_i, W_i); i = 1, 2, \dots, n$ , the problem is to find the function  $F(\cdot)$  that approximates  $f(\cdot)$  over all inputs. That is,

$$\| F(u) - f(u) \| < \epsilon \quad \text{for all } u, \quad (14)$$

where  $\epsilon$  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 in following steps as given in [9]:

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,  $\gamma$  is a user define scalar that can be adjusted to obtain input-to-node that do 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

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

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.

For the integration process, the network is trained using the SfS 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 neural network 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 obtain the final depth-map  $Z_f$  as network output.

A multilayer feed-forward neural network is used because this function approximation problem is a nonlinear problem and very difficult to solve using single layer network. Although one can not fix the architecture of an ideal network for the purpose of solving problems and it can be evaluated by experiments only. However, variation in architecture and algorithm effects only the convergence time of the solution. In our network, each output in a layer is connected to each input in the next layer. In this case, the output layer has simple linear neurons, while all the neurons in the hidden layer have the same transfer function, with a sigmoidal nonlinearity. Also, there is no feedback between layers, 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), three input neurons in the hidden layer and one output neuron corresponding to the SIFT indexes. We train the network on a range of inputs and outputs, such that the network could train and give the more accurate depth-map for any depth-map obtained using shape from shading.

## 5 A Brief Overview on Overall Process

In this section, a brief overview of the whole process is presented. The complete process can be divided into three parts as shown in figure 1.

1. Computation of SfS depth values and SIFT indexes from the right image and the stereo pair, respectively.
  - (a) Let  $I_l$  and  $I_r$  be the left and right stereo images respectively.
  - (b) Using linear generalized Lambertian model, obtain  $Z_{SFS}$  from right stereo image.
  - (c) Compute the SIFT features in  $I_l$  and  $I_r$  separately.

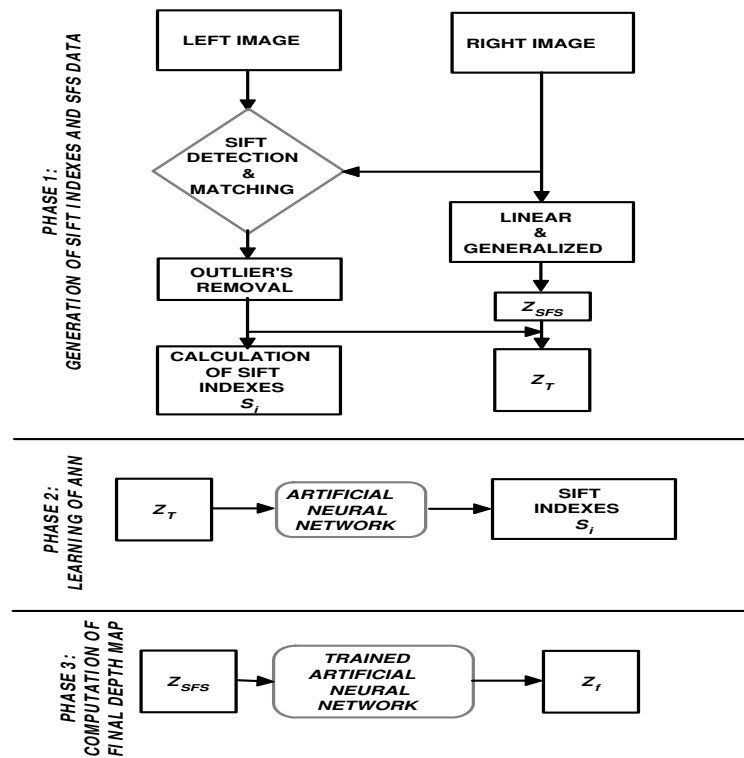


Figure 1: Block diagram for overall process.

- (d) Match the SIFT features to construct a sparse set  $\Pi_i$  of stereo matching pairs.
  - (e) Calculate SIFT indexes  $S_i$  for each pair  $\Pi_i$ .
  - (f) Select the 3D depth data  $Z_T$  from  $Z_{SFS}$  based on the locations of  $\Pi_i$ .
2. Train a feed-forward neural network using error-backpropagation (BP) algorithm by considering  $Z_T$  and  $S_i$  as input-output network pair.
  3. Simulate  $Z_f$  from the trained neural network using  $Z_{SFS}$  as network input.

## 6 Results and Discussions

Several experiments have been performed to recover the depth information using real as well as synthetic images. The integration process has been performed using neural network based function approximation (sigmoidal interpolation). The gradient tolerance has been fixed as  $10^{-6}$  in network training process for all surfaces. The algorithm for neural network training process has been implemented in C++ on a Core2Duo (processor speed 3.0 GHZ) machine with 2GB RAM. MATLAB has been used as the visualization tool by importing the data from the C++ program.

Figure 2 represents the pairs of stereo images for two synthetic surfaces, i.e., Mozart and Vase. Figure 3 shows the 3D shape recovery of Mozart and Vase surfaces using the proposed framework. It can be seen from these results that the surfaces obtained using integration framework are better and more similar to ground truths when compared to surfaces obtained from Sfs alone. Errors in recover-depths have been estimated and reported in table 1 for the proposed framework. There are several ways to report the error behavior. We have reported the error in the following ways:



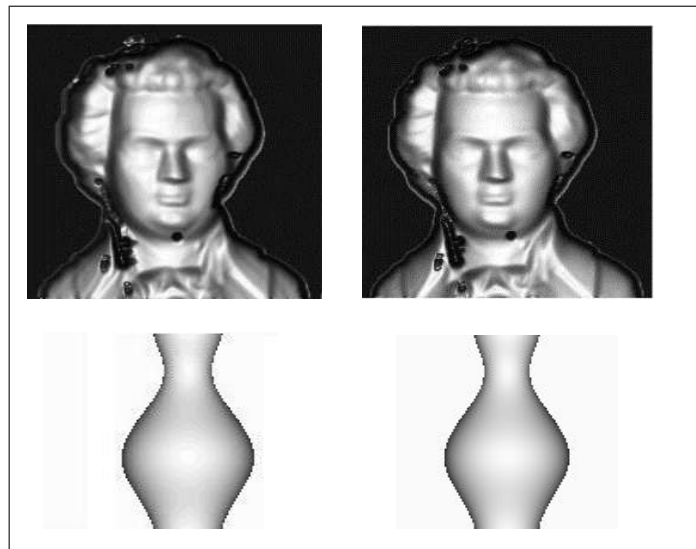


Figure 2: Synthetic pairs of stereo images: Mozart (top row) and Vase(bottom row).

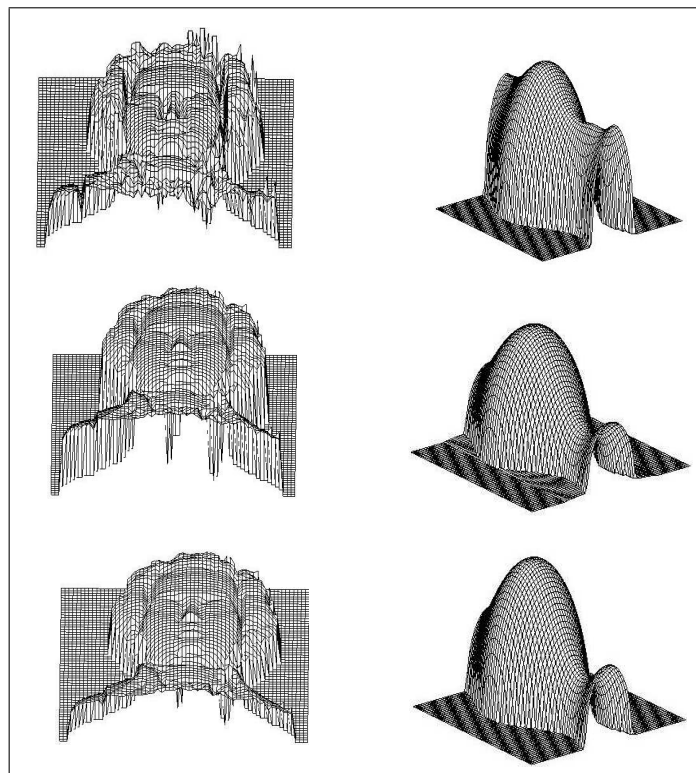


Figure 3: 3D plot of depth values for synthetic surfaces of Mozart (left) and Vase(right) obtained using SfS alone(top row), proposed integration framework (middle row) and ground truth (bottom row).

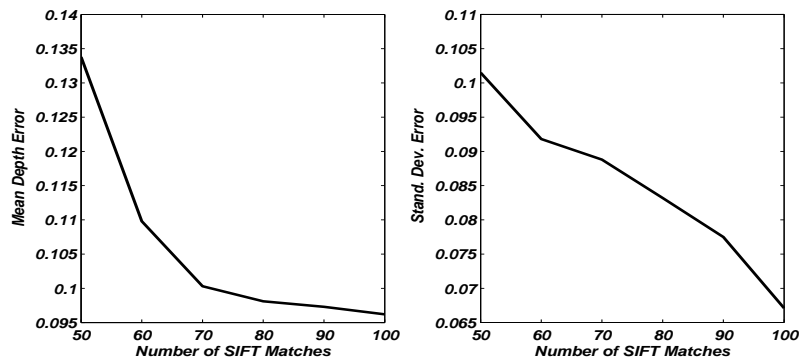


Figure 4: Mean and standard-deviation errors for reconstruction of Mozart surface corresponding to different number of SIFT matches.

- **Mean of depth error:** The obtained depth and ground truth values have been normalized between 0 (zero) to 1 (one). The mean of absolute differences between the ground truth and obtained depth values has been calculated.
- **Standard deviation of depth error:** The obtained depth and ground truth values have been normalized between 0 (zero) to 1 (one). The standard deviation from the absolute differences between the ground truth and obtained depth values has been calculated.

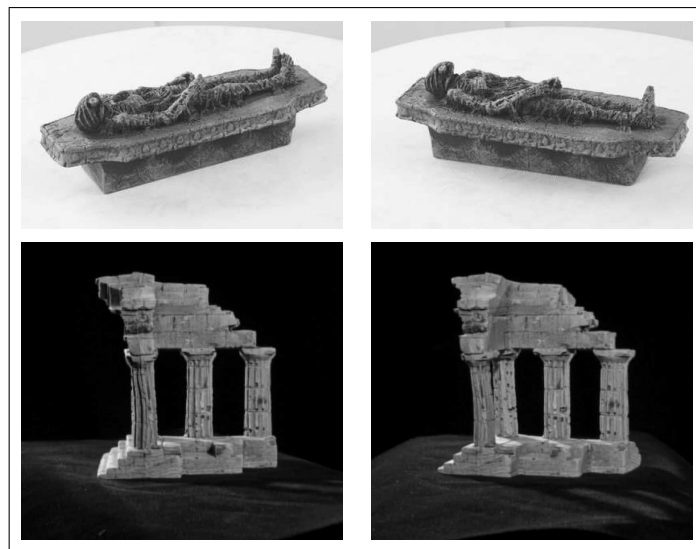


Figure 5: Pairs of stereo images (real): Mummy (top row) and Temple (Bottom row).

In figure 4, mean and standard-deviation of depth errors are plotted for the Mozart surface corresponding to different number of SIFT matches used in integration. These errors are computed using minimum 50 matches and maximum 100 matches. From the figure, it is clear that the error is reduced gradually when the number of matches are increased. In other words, these both are inversely proportional and this is the main objective of proposed framework, i.e., integration of SIFT indexes to reduce the error in reconstruction.

The pairs of real images used to recover the Mummy and Temple surfaces are shown in figure 5. The obtained depth results for these surfaces are shown in figure 6. It can be seen from the presented results that

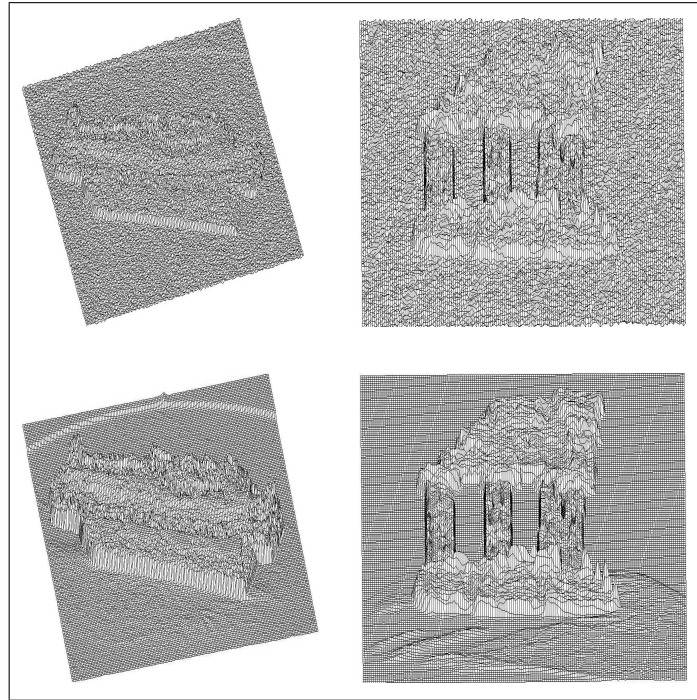


Figure 6: 3D plot of depth values for real surfaces of Mummy (left) and Temple (right) obtained using SfS alone(top row), proposed integration framework (bottom row).

Methods	Mean Error		Stand. Deviation Error	
	Mozart	Vase	Mozart	Vase
SfS alone	0.2567	0.1050	0.1500	.1407
Integration framework	0.0962	0.0354	0.0671	0.0448

Table 1: Reduction in Mean and Standards Deviation Errors due to the integration of SIFT indexes in SfS data.

the quality of 3D reconstruction of surfaces is improved in terms of visibility and accuracy by integrating SIFT indexes in the SfS depth data.

## 7 Conclusions

In this paper, a framework for depth-recovery of complex surfaces has been presented using their texture-less pair of stereo images. The SIFT indexes has been obtained using scale invariant feature matching in stereo images of the surface. The integration process has been performed using a feed-forward neural network. It has been observed that the proposed framework is able to recover more accurate depth-map when compared to SfS process alone in case of texture-less images. The final results for depth-map is more accurate when we have integrated more number of SIFT matches with SfS depth data. Moreover, the proposed framework is simple to apply and can be integrated with any SfS algorithm. The predefined gradient tolerance in the training of neural network has been achieved for all cases using the presented training algorithm.

## Acknowledgements

One of the authors, Manoj Kumar is thankful to Ministry of Human Resources and Development (MHRD), India for the financial support to carry out this research work.

## References

- [1] K. Y. Bae, and B. Behabib, 2003, "A hybrid scheme incorporating stereo-matching and shape from shading for spatial object recognition", *Journal of Engineering Manufacture (IMEchE)*, 217:1533–1542, 2003.
- [2] S. Banerjee, P.S. Sastry, and Y.V. Venkatesh, "Surface reconstruction from disparate shading: an integration of shape from shading and stereopsis", *IEEE proc. of Int. Conf. on Pattern Recognition*, The Hague, The Netherlands, 141–144, 1992.
- [3] M. Bichsel and A.P. Pentland, "A simplr algorithm for shape from shading", *IEEE proc. of Int. Conf. on Computer Vision and Pattern Recogintion*, Urbana-Champaign, Illinois, USA, 459–465, 1992.
- [4] H.I.Bozma and J.S. Duncan, "A Game-Theoritic approach to integration of modules", *IEEE transactions on Pattern Analysis and machine Applications*, 16(11):1074–1086, 1994.
- [5] H.H. Bulthoff and H.A. Mallot, "Integration of depth modules: stereo and shading", *Journal of opt. Soc. of America A*, 5(10):1749–1758, 1988.
- [6] M.T. Chiradia, A. Distanto, E. Stella, "Three dimensional surface reconstruction integrating shading and sparse stereo data", *Journal of Optical Engineering*, 28(9):935–942, 1989.
- [7] J. Cryer, P. Tsai, and M. Shah, "Integration Shape from Shading and Stereo", *Pattern Recognition*, 28(7):1033–1043, 1995.
- [8] R.T. Frankot, and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms", *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 10:439–451, 1988.
- [9] S. Ferrari and R.F. Stengel, "Smooth Function Approximation Using Neural Networks", *IEEE Transaction of Neural Networks*, 16(1):24–38, 2005.
- [10] T.S. Haines and R.C. Wilson, "Integrating Stereo with Shape-from-Shading derived Orientation Information", in *proc. of British Machine Vision Conference*, Warwick, UK, 2007.
- [11] T.S. Haines and R.C. Wilson, "Combining Shape-from-Shading and Stereo using Gaussian-Markov Random Fields", *IEEE proc. of Int. Conf. on Pattern Recognition*, Tampa, Florida, USA, 1–4, 2008.
- [12] M.B. Holte, T.B. Moeslund and P. Fihl, "Fusion of range and intensity information for view invariant gesture recognition", *IEEE proc. of Int. Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW-08)*, Alaska, USA, 1–7, 2008.
- [13] K. Ikeuchi, "Determining a depth map using a dual photometric stereo", *Int. Journal of Robotic Research*, 6(1):15–31, 1987.
- [14] H. Jin, A. Yezzi and S. Soatto, "Stereoscopic Shading: Integrating Multi-frame Shape Cues in a Variational Framework", *IEEE proc. of Int. Conf. on Computer Vision and Pattern Recognition*, Hilton Head, SC, USA, 1:169–176, 2000.
- [15] C.H. Lee, and A. Roseneld, "Improved methods of estimating shape from shading using the light source coordinate system", *artificial Intelligence*, 26:125–143, 1995.

- [16] D. G. Lowe, "Distinctive image features from scale-invariant keypoints", *International Journal of Computer Vision*, 60(2):91–110, 2004.
- [17] M. Mostafa, S.M. Yamani and A.A. Farag, "Integrating Stereo and Shape from Shading", *IEEE proc. of Int. Conf. on Image Processing*, Kobe, Japan,3:130–134, 1999.
- [18] M. Oren, and S.K. Nayar, "Diffuse reflectance from rough surfaces", *IEEE Proc. of Int. Conf. on Computer Vision*, Berlin, Germany, 763–764, 1993.
- [19] P.S. Tsai and M. Shah, "Shape from shading using linear approximation", *Journal of Image and vision computing*, 12(8):487–498, 1994.
- [20] L. Zhang, B. Curless, A. Hertzmann and S. M. Seitz, "Shape and motion under varying illumination: Unifying structure from motion, photometric stereo, and multiview stereo", *IEEE proc. of Int. Conf. on Computer Vision*, Nice, France, 618–625, 2003.
- [21] R. Zhang, P.S. Tsai, J.E. Cryer and M. Shah, "Shape from Shading: A Survey", *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 21:690–706, 1999.