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