3D Reconstruction from Single Texture Image Based on Patch Matching and Optimization

Yujuan Sun*, Muwei Jian†, Shengke Wang† and Xin Sun†
* Ludong University, Yantai, China
E-mail: syj_anne@163.com Tel/Fax: +86-15853578596
† Ocean University of China, Qingdao, China
E-mail: jianmuwei@ouc.edu.cn Tel/Fax: +86-13326392372

***Abstract*—Most current three-dimensional reconstruction methods often require at least two or more input images to reconstruct the shape of the object. Traditionally, it is hard to reconstruct the 3D shape only from one single input image. In this paper, we propose an effective method to reconstruct the 3D texture surface from a single input image for the texture with similar appearances and reflectance properties. Exactly, the patch matching and optimization methods have been combined to resolve the under-constrained problem of 3D reconstruction from one single image. Experimental results have verified the effectiveness of the proposed method in 3D reconstruction according to realistic perception.**

I. Introduction

Based on the sample database, this paper proposes an effective algorithm to reconstruct the 3D texture surface from a single input image for the texture with similar appearances and reflectance properties. For the reconstruction process from a single texture image, too many unknown parameters cause that the 3D reconstruction is a serious ill-posed problem. In order to soften this ill-posed problem, the patch matching method has been used to estimate the initial surface normal of the texture by using the sample database to build the most similar image to the input texture image. At last, the optimization method has been used to refine the initial surface normal. Ten different kinds of rock texture images (160 sample image, where 20 samples as a test set, and 100 samples as sample database randomly selected from the rest samples) in Photex database have been used to verify our algorithm. Experimental results show that the proposed approach is effective and robust for the texture images with similar appearance and reflectance properties.

II. Related works

Recovering 3D shape from a single image is an open and challenging issue. But through an interactive operation with the users or adding constraints, a rough three-dimensional model of the target object can be obtained. Photogrammetry has been used in [?] for acquisition the three-dimensional models for virtual reality. Some literatures have been published aiming to obtain a reasonable solution by adding constraints. For example, a quick method of obtaining 3D models from a single image was presented in [?] based on three types of constraints: coplanarity of points, perpendicularity of directions or planes and parallelism of directions or planes. Delage et. al assumed that the indoor scenes consisted mainly of orthogonal planes, Markov random field model was used to identify the different planes and edges in the scene, and thereby a 3D reconstruction can be obtained in [?]. However, these methods are only applicable to the target object with a distinct geometry, such as buildings, or boxes of different shapes. Ability to build objects of height random variation is relatively weak.

The training method is the most widely used to reconstruct the 3D shape from a single image. Andrew Y. Ng published several related papers [?], [?], [?] about three-dimensional depth information reconstruction from a single image; these methods required a lot of images and three-dimensional depths were used as training information. The Markov Random Field was also used in these literatures to infer a set of plane parameters that captured both 3D location and 3D orientation. The goal of these methods is to create 3D models, which are quantitatively accurate as well as visual pleasing, but still far from detailed reconstruction.

For accurate reconstructing the 3D shape of the object, some researchers focused on reconstructing the object with similar characteristics. Through evolution or interaction with environment, human brain classifies the object into object-classes as to their shape. In [?], a 3D reconstruction method was presented by estimating the parameters of the principal components for human head, which can maintain facial detail and identity of the person. Then using the representation of PCA to solve and optimize shape from shading problem for any human head. Hence an accurate 3D face surface of any person from a single 2D image can be recovered. In [?], an efficient two-dimensional to three-dimensional integrated face reconstruction approach is introduced to reconstruct a personalized 3D face model from a single frontal face image. Then realistic virtual faces are synthesized based on the 3D face and provide the capability to conduct recognition under difficult conditions. Later, a novel method for 3D face shape recovery was proposed by exploiting the similarity of faces [?]. In this method, a single image and a reference face shape are required, and reconstruction results are accurate and robust, but this method is only limited to the human face reconstruction.

The focused reconstruction object in this paper is the texture image, which has the overall similarities for the same type texture, but has the local random height map. Hence it is

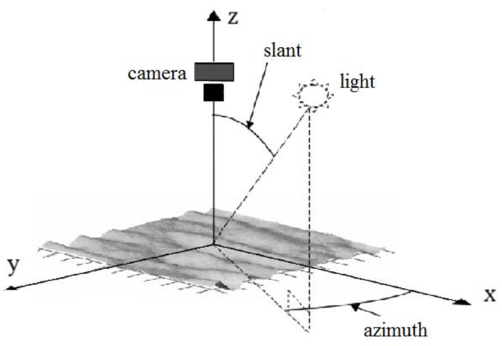

Fig. 1. The light environment model for imaging.

difficult to establish a uniform 3D model. By combining the patch matching and optimization method, an effective method has been proposed in this paper to reconstruct the 3D shape of the texture image with similar characteristics.

III. Our Method

To simplify the discussion, we assume reflection characteristics satisfies Lambertian model, which is shown in Fig.1. The light source is uniform and parallel, and the reflectance of the object can be defined using equation (1).

$$I = \rho(n \cdot l^T) \quad (1)$$

where ρ, n and l represent the albedo, surface normal and light direction. $[n_x, n_y, n_z]$ and $[l_x, l_y, l_z]$ represents the components of n and l respectively in x, y, z direction of Cartesian coordination. $l = [l_x, l_y, l_z]$ can also be expressed as equation (2) using azimuth angle and slant angle in Fig.1.

$$l = [cos(\phi)sin(\theta), sin(\phi)sin(\theta), cos(\theta)] \quad (2)$$

where ϕ and θ represents the azimuth angle and slant angle of the light source, respectively.

A. Build the Training Database

Our algorithm is based on the training method, and the training database needs to build in advance. In order to simplify the complexity of building training database, the texture images in Photex databases are used to indirectly reconstruct the database and verify the proposed algorithm. A lot of different texture images in different lighting conditions (mainly changes the light directions in the azimuth and slant angle) are provided in Photex database, which also includes photometric images under different perspectives. But only the textures of orthogonal projection are in our scope.

Each sample in this database should have the similar characteristics or belong to the same type. The albedo and the corresponding height map constitute the feathers of sample in the training database. The albedo can be represented using the image with 0 degree azimuth and 0 degree slant in Photex database; the surface normal can be estimated from images of the sample in different lighting conditions using photometric stereo method (PMS) [?]. Then the sample can be written as follow:

$$sample_i = \{\rho_i, n_i\}, i = 1, 2, \ldots N \quad (3)$$

where i represents the ith sample in training database, which includes N samples. ρ_i and n_i represent the albedo and surface normal of the ith sample respectively.

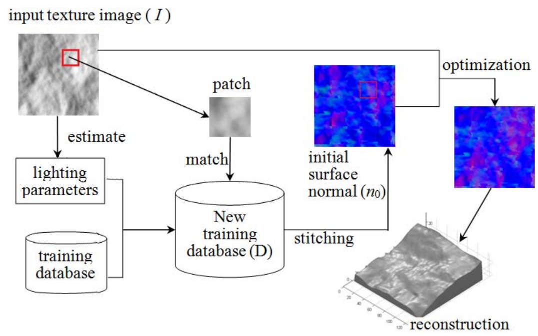

Fig. 2. The framework of the proposed method.

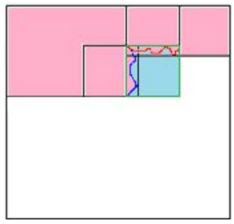

Fig. 3. The stitching diagram of quilting method.

B. 3D Surface Reconstruction Based on Patch Matching and Optimization

Our method need just one input texture image, moreover the lighting parameters of this image need not to know.

The framework of the proposed method is shown in Fig.2, and the details are listed as follows:

(1) The input data is a texture image I with unknown lighting parameters.

(2) The method in [?] is used to estimate the lighting parameters (l, which can be represented by ϕ and θ) of the input texture image I.

(3) By using the training database and the estimated lighting parameters (ϕ, θ), the new training database (D) can be built. D includes the sample surface normal (n_i) and the sample image (S_i). S_i has the same lighting conditions with the input texture image. Equation (4) shows the building of the new training database:

$$D = \begin{cases} n_i & i = 1, 2, \ldots, N \\ S_i & S_i = \rho_i(n_i \cdot l), i = 1, 2, \ldots, N \end{cases} \quad (4)$$

(4) By using the patch matching (discussed detailed in section $III.C$) between the input texture image and the sample image in D, the initial surface normal (n_0) of the input image can be obtained.

(5)The initial surface normal can be refined by using optimization method (discussed detailed in section $III.C$) and then optimized surface normal can be obtained.

(6)The height map of the input texture image can be integrated from the optimized surface normal by using the method in [?].

C. Patch Matching and Optimization

1) Patch Matching: 3D reconstruction from a single input image is the under-constrained problem. In order to get a reasonable solution, the initial surface normal estimation for the input texture image is essential. Based on the similarity of the same type texture, the initial surface normal can be obtained by using the patch matching between the input texture image and the sample image in the new training database.

The input texture I is the gray scale image and its size is 128×128. I will be divided into patches with size of 28×28. The most similar patch (called similar patch) with each patch (called input image patch) in I can be found if the minimum of equation (5) is solved.

$$fun_i(k,l) = min\{Sel_{k=1}^{N}[(1 - Cor_{l=1}^{M}(I_i, S_{kl})]\} \quad (5)$$

where I_i represents the ith input image patch; S_{kl} (the lth patch of kth sample image in the new database) represents the most similar patch with I_i. N represents the number of the samples in new training database. M represents the number of the patches in one input image. Cor represents the similar function, which will find the most similar patch at one sample image, and the similar function in this paper is computed by comparing the gray value similarity between input image patch and similar patch. Sel represents the selection function, which can find the most similar sample image in all samples of the new training database. Equation (5) is valid on the condition that the input texture image and the sample image of the new training database are in the same scale (if not in the same scale, muti-scale method can be used, and such a situation is not in the scope of this paper). If the similar patch S_{kl} is found, the corresponding surface normal (called similar normal patch) will be used as the initial surface normal of input image patch. The quilting method in [?] is used to stitch all the similar normal patches (called initial surface normal). Fig.3 shows the stitched result of the initial surface normal for the input texture image. The pink area represents the stitched similar normal patch, and there will be overlap area when the new patch (the green box) is to be stitched. The red and blue lines in overlap represent the minimum cutting path when stitching, which is detailed defined in [?].

2) Optimization: The initial surface normal provides the basic height information, but the details still have stitching errors and artifacts, which can be seen from Fig.2. We build the objective function of equation (6) to further refine the initial surface normal.

$$F(n) = min\{(I - \rho n \cdot l)^2 + \lambda(\sum_{j \in \Omega} angular\_err(n, n_j))\} \quad (6)$$

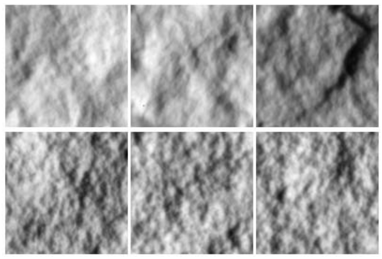

Fig. 4. The input rock images with unknown lighting parameters (from left to right, from top to bottom, the number is 1, 2, 3, 4, 5 and 6 respectively).

where Ω represents the neighborhood of surface normal n; $angular\_err$ is the angle error function, which represents that the estimated surface normal n should have the minimum angle error with the surface normal of its neighborhood.

IV. The Experiments and the Analyses

The images in the Photex texture database is monochrome, and the rock images with resolution of 512×512 have been used in our experiments. There are ten kinds of rock images, which include aab, aaf, aai, aaj, aam, $aanaaoaap$, aar and aas. We divid each rock image into 16 pieces (each piece with a resolution of 128×128), hence a total of 160 samples have been produced. 20 samples are used as the test set, and random 100 samples in the rest are regarded as the training set. Many intensity images in different kinds of lighting directions for each kind of rock image have been captured in this database. Then our new training database with the random 100 samples can be built according to section $III.A$. The input texture image can be selected from the test set of 20 samples, and the lighting direction is not known in advance. Fig.4 shows several input rock images with unknown lighting parameters.

Fig.5 shows the reconstructed surface normal of the images in Fig.4. From left to right, the first column is the single input images; the second column is the initial surface normals; the third column is the optimized surface normals; the fourth column is the surface normals estimated using PMS [?]. Due to the calibrated multi-images are used in PMS method and the accuracy of reconstructed results is high. Many literatures [?], [?] used the results of PMS as the ground truth (G.T). Hence the fourth column in Fig.5 has been regarded as the G.T. We can see from Fig.5 that the initial surface normal is similar with the G.T on the whole. But there are obvious artifacts at the patch edge; moreover large errors exist in details. The artifacts of the optimized surface normals in third column have been reduced largely and are closer to G.T. Fig.6 shows the reconstructed height maps for the input images in Fig.4. From top to bottom, the number of the input image is 1, 2, 3, 4, 5 and 6 respectively; the first column is the optimized height maps using proposed method; the second column is the height maps using PMS; the third column is the height maps using the method of shape from shading (SFS) [?].

As can be seen from Fig.6, the reconstructed results of the proposed method are closer with G.T (height maps in second column) than those of SFS.

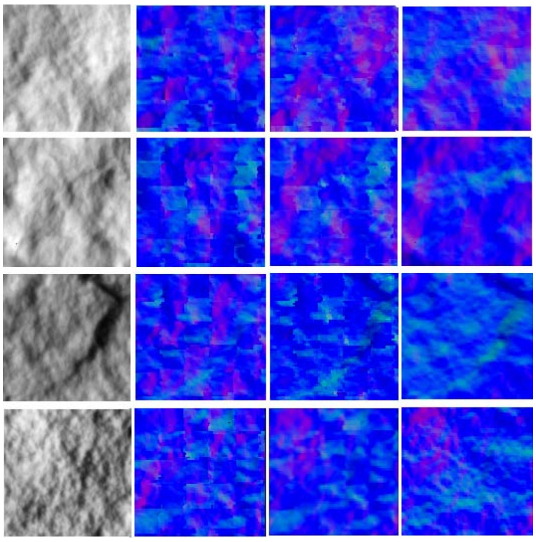

Fig. 5. The reconstructed surface normals of the images in Fig.4 (the first column is the input image, and its number is 1,2,3 and 4 respectively from top to bottom; the second column is the initial surface normals; the third column is the optimized surface normals; the fourth column is the surface normals estimated using PMS).

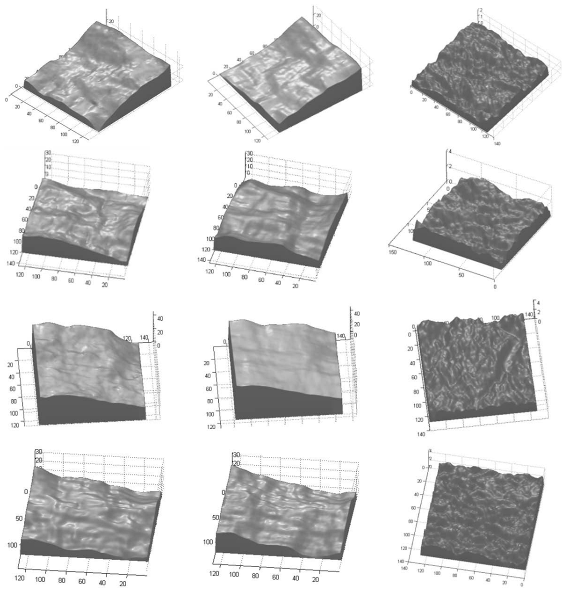

Fig. 6. The reconstructed height maps of different methods (From top to bottom is the height maps for the input image of 1, 2, 3 and 4 in Fig.4 respectively. The first column is the optimized height maps using proposed method; the second column is the height maps using PMS; the third column is the height maps using SFS.)

V. Conclusions

An effective method of reconstructing 3D shape from a single input texture image has been proposed in this paper. The patch matching and optimization are used for resolving the under-constrained problem of 3D reconstruction. The experiment results on rock images in Photex database verified the proposed method is effective according to human realistic perception.

Acknowledgment

This work was supported by Natural Science Foundation of Shandong (ZR2015FQ011; ZR2014FQ023); National Natural Science Foundation of China (NSFC) (61271405, 61401413); The Fundamental Research Funds for the Central Universities (201511008, 30020084851); Open Fund of Shangdong Coputer Federation and International Science & Technology Cooperation Program of China (ISTCP) (2014DFA10410).