ARTICLE IN PRESS

J. Vis. Commun. Image R. xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

journal homepage: www.elsevier.com/locate/jvci

Improved marching tetrahedra algorithm based on hierarchical signed distance field and multi-scale depth map fusion for 3D reconstruction

Dan Guo [a,*], Chuanqing Li [a], Lu Wu [b], Jianzhong Yang [c]

[a] School of Computer and Information, Hefei University of Technology, Hefei, China
[b] School of Information Engineering, Wuhan University of Technology, Wuhan, China
[c] Huazhong University of Science and Technology, Wuhan, China

A R T I C L E I N F O

Article history:
Received 18 September 2016
Revised 20 November 2016
Accepted 29 December 2016
Available online xxxx
This paper has been recommended for acceptance by Anan Liu

Keywords:
3D reconstruction
Surface reconstruction
Depth map fusion
Hierarchical signed distance field
Feature matching

A B S T R A C T

3D reconstruction systems are promoted by developments of both computer hardware and computing technologies. They still remain problems like high expense, low efficiency and inaccuracy. Especially for large-scale scenes, lack of full use of multi-scale depth information will cause blurring and irreal reconstruction results. To solve this problem, we construct the structure of hierarchical signed distance field (H-SDF) and design an improved marching tetrahedra algorithm for multi-scale depth map fusion. In addition, to improve efficiency, we also propose a two-phase search strategy in image feature matching: the bag-of-features model (BOF) is adopted in a coarse search to narrow search scope and then the SIFT descriptor is used in exact matching to pick reconstruction image points. Experiment results indicate that coarse search makes matching time shorter; using the H-SDF to fuse multi-scale depth maps, and isosurface extraction with improved marching tetrahedra algorithm can improve visual effect.

1. Introduction

3D reconstruction has widely applications in the fields of computer vision and graphics. For example, in film and animation area, it is used to build the 3D models of objects and scenery [1–3]. In cultural heritage protection researches [4], 3D reconstruction can be utilized for virtual reproduction of heritages and monuments. Moreover, in fields like reverse engineering [5], location-based service [6–9] and action recognition [10,11], 3D reconstruction also plays an important role and bring significant economic and cultural benefits.

In this paper, we focus on 3D reconstruction of large-scale scenes, such as geography and terrain reconstruction system [12]. The problem is somewhat different. 3D reconstruction aims at modeling scenes exactly, rapidly and less costly. However, current sensor-based methods, such as structured light sensor [13,14], can reconstruct object surface well, but their expensive costs and limited modeling scopes make them impractical for large-scale data collection. Image-based methods can achieve great progress [15]. But they always sacrifice computational efficiency for high accuracy. While constructing several objects in a large scene simultaneously, traditional image-based methods may become time-consuming. What is important that large-scale scenes usually have multi-scale analysis from different views and scales. Without considering this analysis may result in blurry and irreal reconstruction effect. Therefore, to be efficient and accurate for large-scale scenes, we build two scene datasets by collecting 2D image samples with Unmanned Aerial Vehicle (UAV), and take multi-scale information in account to research 3D reconstruction.

Major steps of the image-based 3D reconstruction are shown in Fig. 1: feature matching, camera calibration, depth map extraction, depth map fusion and surface reconstruction. The first three steps can be handled with varieties of mature technologies. But depth map fusion is still difficult to be solved. Usually, a 2D image merely covers a partial surface of the object or scene due to the limited camera view and the object occlusion problem. How to get complete surface information, multi-scale depth map fusion is considered.

Early fusion approaches transform different depth maps into a unified object centered coordinate system [16–18]. These depth maps must be simple and compact. Subsequently, surface-based fusion approaches are proposed. For example, Pito picked appropriate depth map for each sub-surface region and then sutured remaining surfaces [19]. In theory, it is able to fuse maps of different scales. However it is sensitive to exterior points and not fit for large curvature surfaces. In addition, to simulate sensor noises,

* Corresponding author.
E-mail addresses: guodan@hfut.edu.cn (D. Guo), lichuanqing@mail.hfut.edu.cn (C. Li), wulv@whut.edu.cn (L. Wu), yangjzh11@mail.hust.edu.cn (J. Yang).

http://dx.doi.org/10.1016/j.jvcir.2016.12.016
1047-3203/© 2017 Elsevier Inc. All rights reserved.

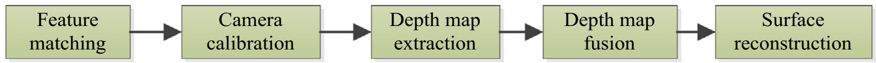

Fig. 1. The framework of image-based 3D reconstruction.

Curless and Levoy proposed an uncertainty method based on the signed distance field [20]. With the moving least square method, Zach et al. [21] transformed the fusion problem into a global optimality problem of energy function of the variation model. However, without considering different depth scales in the fusion process, the signed distance field will cause 3D structure errors. In this paper, we construct hierarchical signed distance field (H-SDF) for fusion of different scale depth maps.

Then for the surface reconstruction step, there are also varieties of approaches including 2D contour line stitch method [22], marching cubes algorithm [23] and marching tetrahedra algorithm [24]. The marching cubes algorithm is the most popular, but it is only applicable for regular arrayed voxels. To break through this constraint, Kazhdan et al. [25] proposed an isosurface extraction method, which requires that each non-leaf node of the octree structure has eight child nodes. To solve connection ambiguity in the marching cubes [23] algorithm, Doi and Koide proposed the marching tetrahedra algorithm [24], which firstly partitioned each voxel into multiple tetrahedrons and then constructed isosurfaces of the extracted tetrahedrons. However, for multi-scale depth maps, approaches mentioned above cannot extract tetrahedrons based on hierarchical signed distance field (H-SDF) for different scale depth maps.

In this paper, we focus on depth map fusion and surface reconstruction steps in the 3D reconstruction procedure, and propose an improved marching tetrahedra algorithm based on hierarchical signed distance field. Our algorithm is fit for large-scale scenes. We verify it on two large scale scene datasets collected by UAV.

(1) For multi-scale fusion, the hierarchical signed distance field (H-SDF) is calculated based on triangulations with different depth levels. The confidence weight of each voxel should be updated and unreliable voxels in the field should be removed.
(2) With the H-SDF, classical marching cubes algorithm [23] and marching tetrahedra algorithm [24] are not suitable to solve the isosurface extraction problem. We introduce an improved marching tetrahedra method to filter unnecessary isosurfaces and useless generated triangles. This improvement makes reconstructed surfaces much smoother.
(3) In addition, these point-based approaches that calculate all points of depth maps usually result in high time and space complexity. Especially, when the image dataset is large, feature matching with local descriptors is time consuming. To improve efficiency, our paper uses two-phase search strategy in image matching: we firstly use the bag-of-features model (BOF) in a coarse search to narrow the search scope and then use the SIFT descriptor in exact matching to pick reconstruction image points.

The paper is organized as follows: Section 2 introduces details of H-SDF for multi-scale depth-map fusion. Then we propose an improved marching tetrahedra algorithm based on H-SDF for surface reconstruction in Section 3. Section 4 gives our whole algorithm and Section 5 compares our method with the VRIP method [20]. Finally, we conclude the paper in Section 6.

2. H-SDF for multi-scale depth map fusion

Motived by the signed distance field [26], we improve the triangulation construction process to adapt to multi-scale depth map fusion. As shown in Fig. 2, we construct the hierarchical signed distance field (H-SDF) based on the octree structure. Octree is a tree structure that describes the 3D construction space [25]. Its root node indicates the whole cubic space (one big voxel) and each sub-cube (eight small voxels) respectively indicates its eight vertices. Similarly, each small voxel can be further split into its own 8 smaller sub-voxels. We can use the duplicated information of those voxels' (vertices') to represent different layers. The octree structure of each point (node) in a depth map can be expressed as the index structure (l, I_l), where l denotes the located layer of the point and I_l indexes the corresponding layer of each generated voxel. Here is $I_l \in \{0,\ldots,2^{3l}-1\}$.

We initialize a to-be-fused map as null and generate its multiple triangulation-based depth maps by the octree structure. A node in the octree structure is created when a new triangle is interpolated into the fused map. The process is shown in Fig. 2. When a point (vertex) in a depth map is first visited, its corresponding node is created; otherwise its node weight value in the octree structure is updated.

2.1. H-SDF calculation

2.1.1. Extraction of to-be-interpolated triangles

We firstly need to ascertain to-be-interpolated triangles, and their placed layers and affected voxels. We use the definition of pixel trace F_O to determine the generated triangles whether are chosen to be interpolated. As shown in Fig. 3, given the optical center of the camera O, a point N in the 2D image plane and the 3D point P related to N, the trace of N is $F_O = \|MP\|$.

We use the extraction method proposed by Fuhrmann and Goesele [26] to triangulate the depth maps. We project depth maps into 3D space and connect any three adjacent vertices. Besides, we define the threshold $\tau = \rho \cdot F_o$. For each generated triangle, if the depth difference between any two vertices is bigger than τ, we consider the corresponding vertices of these two pixels are not adjacent, and the triangle will be removed. By the characteristic of our large-scale scene data collected by the UAV, the redundancy

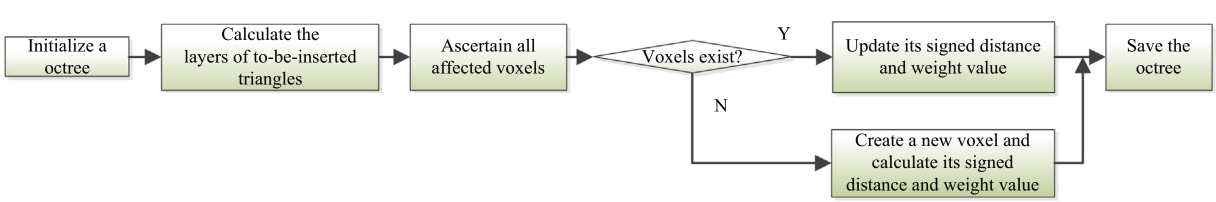

Fig. 2. Flowchart of the H-SDF construction.