J. Vis. Commun. Image R. 40 (2016) 831–837

Contents lists available at ScienceDirect

J. Vis. Commun. Image R.

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

Plane surface detection and reconstruction using segment-based tensor voting ☆

Liangbing Feng [a,b], Zhihan Lv [a,*]

[a] Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China
[b] Guiling University of Electronic Technology, GuiLing, Guangxi, China

ARTICLE INFO

Article history:
Received 4 February 2016
Revised 10 July 2016
Accepted 20 August 2016
Available online 22 August 2016

Keywords:
Tensor voting
Multi-view stereo
3D reconstruction
Image segmentation

ABSTRACT

A Segment-based Tensor Voting (SBTV) algorithm is presented for planar surface detection and reconstruction of man-made objects. Our work is inspired by piecewise planar stereo reconstruction. During the vital procedure to detect and label the planar surface, the two main contributions are: first, tensor voting is used for obtaining the geometry attribute of the 3D points cloud. The candidate planar patches are generated through scene image segment of low variation of color and intensity. Second, we over-segment the scene image into the segment and the candidate 3D planar patch is generated. The SBTV algorithm is used on 3D points cloud sets to identify the co-plane on the candidate patch. After detecting every planar patch, the geometry architecture of object is obtained. The experiments demonstrate the effectiveness of our proposed approach on either outdoor or indoor datasets.

1. Introduction

Piecewise planar models for stereo reconstruction have recently become popular for modeling man-made indoor and urban outdoor scenes [1], since man-made structures mainly consist of planes [12]. The approaches recovering 3D structure from stereo can be coarsely divided in two groups: sparse stereo; that inputs a sparse set of matching features across views; and dense stereo, that preforms dense data association between images by assigning each pixel a disparity value [2]. Our proposed method is somewhere in between the sparse stereo and dense stereo. With a small set of images the reconstruction result is incomplete, since leaving small holes on the texture-less surfaces and specular reflections like uniformly painted walls. Planar approximations to the scene can obtain more visually pleasing 3D reconstructions [3,4]. However, these algorithms use image features such as key points and strong lines, which may be absent in texture-less surfaces. Furukawa et al. [3] use a very specific Manhattan-world model, where all planes must be orthogonal to each other. Similarly, Sinha et al. [4] use a general piecewise planar model for 3D reconstruction. The 3D lines are extracted, and then the candidate planar surfaces are detected. Kowdle et al. [2] attempt to obtain a piecewise planar reconstruction of the scene automatically through an energy minimization framework. However, the users need provide some interactions to indicate the coplanar regions, non-coplanar regions and so on. Gallup et al. [1] proposed a technique to segment an image into piecewise planar regions as well as regions labeled as non-planar. The reconstruction work of Zebedin et al. [5], focuses on aerial imagery, in addition to planar rooftops. This work is suitable for a surface of revolution representation to handle domes and spires. Hoiem et al. [6] and Saxena et al. [7] use color, texture, and other image features to infer geometric context.

These papers use strong planarity assumption as a prior for stereo construction to tackle problems caused by poorly textured surfaces and non-Lambertian reflections (e.g. windows). The achieved 3D models are perceptually pleasing and geometrically simple, thus, their rendering, storage and transmission process have lower computational complexity. However, locating planar surfaces in the scene for establishing plane hypotheses can be a very challenging task. In these papers, some need user interactions; some need plane orientation constraints, and so on. In this paper, a Segment-based Tensor Voting algorithm is proposed for man-made objects' planar surface detection and reconstruction. The proposed method need not compute 3D lines. The candidate planes are generated according to 3D point's geometry feature which directly extracted from 3D point cloud. In addition, a robust multiple planes detection approach is proposed from noisy and corrupted sparse 3D points cloud data sets. In [13–15,17–22], some learning algorithms have proposed for feature extracting in plane detection.

☆ This paper has been recommended for acceptance by Zicheng Liu.

* Corresponding author.
E-mail address: lvzhihan@gmail.com (Z. Lv).

http://dx.doi.org/10.1016/j.jvcir.2016.08.012
1047-3203/

In this paper, a novel plane surface detection method is presented. The mainly contributes of this paper are below. 1. Tensor voting is used for obtaining the geometry properties of the 3D points cloud. The candidate planar patches are generated through scene image segment of low variation of color and intensity. 2. The scene image is over-segmented into the candidate 3D planar patch is generated. The SBTV algorithm is used on 3D points cloud sets to identify the co-plane on the candidate patch. 3. The 3D model is reconstructed based on plane detection.

2. Overview of the proposed approach

Fig. 1 depicts the proposed reconstruction approach. The main steps of the proposed approach are: sparse points obtaining, segmenting image and clustering 3D point into different patch, computing planar salience (planar normal) using tensor voting, merging 3D points patch and generating reconstruction result.

First, the sparse point cloud of scene is recovered by structure-from-motion method as shown in Fig. 1a. In this paper, we employed the state-of-the-art tool so-called Bundler [8] to achieve this step. And then, the correspondence between the generated 3D points and the key points of the 2D image can be recorded as shown in Fig. 1b. When a 2D image is segmented, the 3D points can be clustered into different patches according to the spatial relationship of the 3D points corresponding to the segmentation of the 2D image. We named this kind of 3D points as the segmented 3D points. Here, our segmenting approach is built upon the assumption that large disparity discontinuities only occur on the boundaries of homogeneous color segments. Therefore any color segmentation algorithm that decomposes an image into several homogeneous color regions can work for us.

Second, tensor voting algorithm is used in all 3D points, thus the geometry feature (normal of plane which the points on it) of every point is obtained as shown in Fig. 1c top. The planar salience and normal of every segmented 3D points (candidate plane) are generated. The normal of neighbor segmented 3D points are compared, and then 3D point patches which have near normal are merged. This process is iterative until the plane is extracted completely. Fig. 1c below is the point patches after 9 times neighbor segments mergence. Fig. 1d is the point patches after 10 times neighbor segments mergence. Fig. 1e is the point patches after final mergence. Finally, the mesh of the reconstruction result is gotten as shown in Fig. 1f. The reconstruction result which pasted the textures is shown in Fig. 1g.

In the next section, the segment-based tensor voting method will be elaborated. The experiment and result will be descripted in Section 4. Section 5 summaries this paper.

3. Segment-based tensor voting algorithm

The purpose of tensor voting is to extract geometric feature such as regions, curves, surfaces, and the intersection between them. Here, tensor voting is used for finding plane saliency and plane normal from the sparse 3D point cloud.

3.1. Tensor voting from 3D points cloud

Tensor Voting is a robust computational approach for grouping and segmentation [9]. Tensors are simply mathematical objects that can be used to describe physical or geometrical properties, just like scalars, vectors and matrices. The methodology is grounded into two elements: tensor calculus for representation of geometrical properties, and voting for data communication. Every point embedded in R3 can be encoded as second order, symmetric and positive semi-definite 3×3 tensor and decomposed as:

$$\boldsymbol{T} = \lambda_1 \vec{e}_1 \vec{e}_1^T + \lambda_2 \vec{e}_2 \vec{e}_2^T + \lambda_3 \vec{e}_3 \vec{e}_3^T \tag{1}$$

Here, $\vec{e}_1$, $\vec{e}_2$, $\vec{e}_3$ are the eigenvectors associated with eigenvalues λ_1, λ_2 and λ_3 ($\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$), respectively. The above formula can be rewritten as

$$\boldsymbol{T} = (\lambda_1 - \lambda_2)\vec{e}_1 \vec{e}_1^T + (\lambda_2 - \lambda_3)(\vec{e}_1 \vec{e}_1^T + \vec{e}_2 \vec{e}_2^T) + \lambda_3(\vec{e}_1 \vec{e}_1^T + \vec{e}_2 \vec{e}_2^T + \vec{e}_3 \vec{e}_3^T) \tag{2}$$

Fig. 2 shows a generic tensor visualized as a 3D ellipsoid. It has three components: a stick, a plate and a ball, each of them is formulated as $\vec{e}_1 \vec{e}_1^T$, $\vec{e}_1 \vec{e}_1^T + \vec{e}_2 \vec{e}_2^T$ and $\vec{e}_1 \vec{e}_1^T + \vec{e}_2 \vec{e}_2^T + \vec{e}_3 \vec{e}_3^T$, respectively. The size of the each component, termed 'saliency', indicates how much the corresponding component is involved in the tensor $\boldsymbol{T}$. Through tensor voting between points, each point

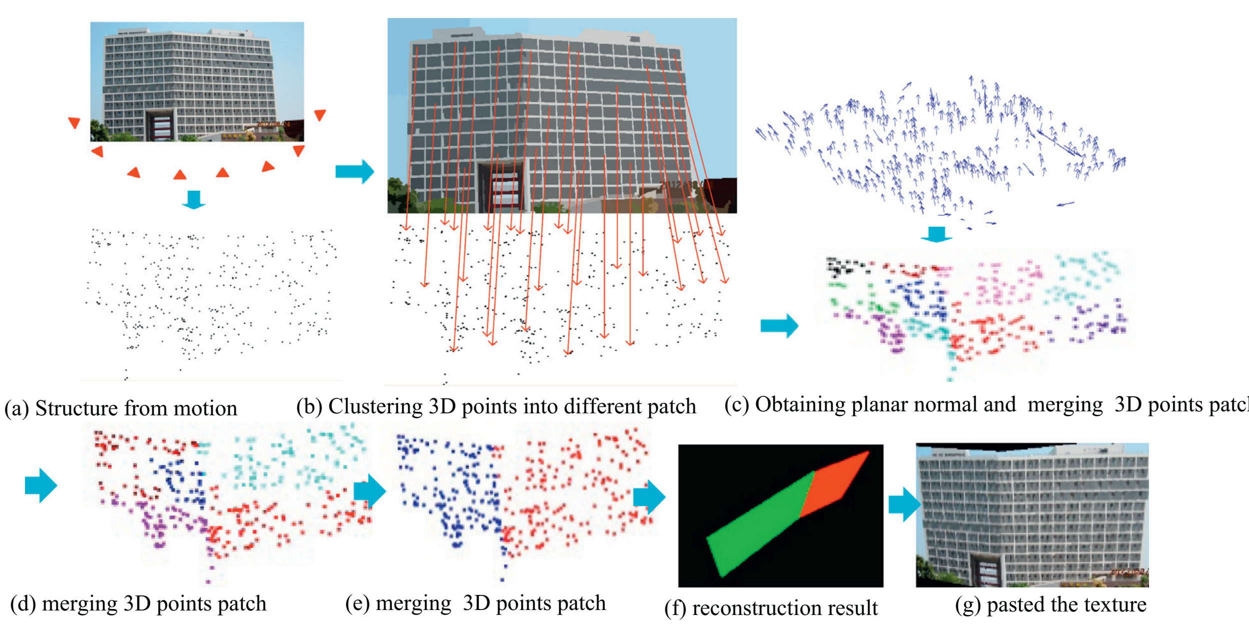

Fig. 1. Workflow of the proposed reconstruction approach.