ISPRS Journal of Photogrammetry and Remote Sensing 125 (2017) 33–49

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

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

Line segment matching and reconstruction via exploiting coplanar cues

Kai Li, Jian Yao *

Computer Vision and Remote Sensing (CVRS) Lab, School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, PR China

ARTICLE INFO

Article history:
Received 13 July 2016
Received in revised form 26 December 2016
Accepted 9 January 2017

Keywords:
Line segment matching
3D line segment reconstruction
Local region detector
Homography estimation
Markov random field

ABSTRACT

This paper introduces a new system for reconstructing 3D scenes from Line Segments (LS) on images. A new LS matching algorithm and a novel 3D LS reconstruction algorithm are incorporated into the system. Two coplanar cues that indicates image LSs are coplanar in physical (3D) space are extensively exploited in both algorithms: (1) adjacent image LSs are coplanar in space in a high possibility; (2) the projections of coplanar 3D LSs in two images are related by the same planar homography. Based on these two cues, we efficiently match LSs from two images firstly in pairs through matching the V-junctions formed by adjacent LSs, and secondly in individuals by exploiting local homographies. We extract for each V-junction a scale and affine invariant local region to match V-junctions from two images. The local homographies estimated from V-junction matches are used to match LSs in individuals. To get 3D LSs from the obtained LS matches, we propose to first estimate space planes from clustered LS matches and then back-project image LSs onto the space planes. Markov Random Field (MRF) is introduced to help more reliable LS match clustering. Experiments shows our LS matching algorithm significantly improves the efficiency of state-of-the-art methods while achieves comparable matching performance, and our 3D LS reconstruction algorithm generates more complete and detailed 3D scene models using much fewer images.

1. Introduction

Image-based 3D reconstruction is a widely studied research field and researchers have developed some remarkable works through exploiting feature points extracted from images (Agarwal et al., 2011; Furukawa and Ponce, 2010; Snavely et al., 2006, 2008; Wu, 2013). However, objects in man-made scenes are often structured and can be outlined by a bunch of Line Segments (LS). It is therefore advantageous to get the 3D wire-frame model of a scene by exploiting LSs on images. For example, for the house shown in Fig. 1(a), the 3D LS reconstruction method to be introduced in this paper generates the 3D model shown in Fig. 1(b) using only two images. It is easy to recognize the house from this 3D model, but hardly possible to do so from the extremely sparse point clouds obtained by some point based 3D reconstruction methods. Some works (Hofer et al., 2014; Sinha et al., 2009) also proved that 3D modeling by exploiting both feature points and LSs on images resulted in more accurate and complete results.

Despite of the above benefits of exploiting LSs for 3D scene reconstruction, it is yet hard to reliably reconstruct 3D LSs. The foremost reason is that LSs are difficult to be matched, such that even several 3D LS reconstruction algorithms (Hofer et al., 2013; Jain et al., 2010; Ramalingam and Brand, 2013) skipped LS matching and directly reconstructed extracted image LSs. The main cause for the difficulties of matching LSs is the absence of point-to-point correspondence between corresponding LSs. The endpoints of corresponding LSs do not reliably correspond with each other, and a short LS from one image is allowed to correspond to a long one from another image. This fact makes it unreliable to exploit some local region description based methods, which have been proved to be very effective in feature point matching, for LS matching because it is hard to extract invariant local regions around LSs.

Another factor complicating 3D LS reconstruction is the unstableness and low location accuracy of extracted LSs. LSs are the straight fittings of curve edges detected on images so that sometimes a 3D edge would result in two straight fittings that are not precisely corresponding on two images. The imprecise correspondence of corresponding LSs makes it difficult to reliably reconstruct their 3D correspondences. For example, to reconstruct 3D LSs in the scene shown in Fig. 1(c), all of which can roughly be regarded to lie on one space plane, when using traditional way to triangulate (forward intersect) LS correspondences identified from two

* Corresponding author.
E-mail address: jian.yao@whu.edu.cn (J. Yao).
URL: http://cvrs.whu.edu.cn/ (J. Yao).

http://dx.doi.org/10.1016/j.isprsjprs.2017.01.006
0924-2716/

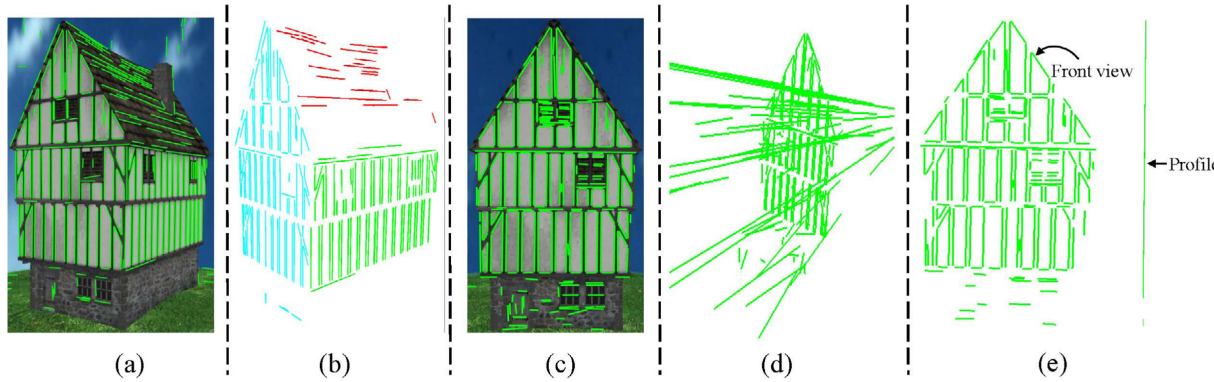

Fig. 1. Examples showing the benefits and difficulties of exploiting LSs on images for 3D reconstruction. (a) A multi-planar scene and the extracted LSs. (b) The reconstructed 3D LSs for the scene (a) obtained by the proposed 3D LS reconstruction method using two images. Different colors are used to differentiate 3D LSs lying on different space planes. (c) An image of a roughly planar scene and the extracted LSs. (d) The 3D LS reconstruction result for the scene shown in (c) by triangulating LS correspondences identified from two images. (e) The 3D LS reconstruction result obtained by our proposed algorithm after solving the problem existing in (d). The front view and profile of the obtained 3D model are shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

images, the 3D LSs shown in Fig. 1(d) are obtained. As we can see, a big fraction of the 3D LSs are incorrectly reconstructed.

We solve the aforementioned problems by exploiting the following two coplanar cues that indicate the coplanarity image LSs in space.

- $\mathcal{C}1$: Adjacent image LSs are coplanar in space in a high possibility.
- $\mathcal{C}2$: The projections of coplanar space LSs in two images shall be related by the same planar homography.

As for the no point-to-point correspondence problem in LS matching, based on $\mathcal{C}1$, we intersect adjacent LSs to form V-junctions in both images to be matched. Since adjacent image LSs are very likely to be coplanar in space and the intersecting junctions of coplanar space LSs are projectively invariant with camera motions, a portion of V-junctions constructed in one image would reappear in the other image. Through matching V-junctions from the two images, the LSs forming the obtained V-junction matches are matched accordingly. While matching V-junctions, we propose to extract for each V-junction a scale and affine invariant local region and describe it with SIFT. For LSs unable to be matched along with V-junctions (due to that they are not adjacent enough to others as to be used to form V-junctions) based on $\mathcal{C}2$, we use local homographies estimated from their adjacent V-junction matches to evaluate their correspondence.

When reconstructing 3D LSs, based on $\mathcal{C}2$, we group LS matches obtained from two images according to a set of homographies, such that LS matches in each group are related by the same homography, which is induced by the space plane where the 3D LSs corresponding to the LS matches in the group lie. The space plane for each LS matches group can then be recovered from the 3D LSs obtained by triangulating all the pairs of LS correspondences. As the space plane being recovered, the 3D LSs corresponding to LS matches in the group can be obtained easily by back-projecting LSs from one image onto the space plane. To reduce the incidence of incorrect LS match grouping, we introduce coplanar cue $\mathcal{C}1$ into LS match grouping, frame it to Markov Random Field (MRF) and solve it as a multi-label optimization problem. Fig. 1(e) shows our 3D LS reconstruction result for the scene shown in Fig. 1(d). It is easy to observe that our algorithm has successfully remedied the problem existing in Fig. 1(d).

In summary, the novelties of this paper are threefold: First, we propose to match V-junction from two images by extracting for each of them a scale and affine invariant local region. Second, we propose a new solution for solving the ambiguities in 3D LS reconstruction through LS match grouping, space plane estimation and image LS back-projection. Third, we formulate to solve the LS match grouping problem by solving a multi-label optimization problem.

The rest of this paper is organized as this: Section 2 presents relevant works to ours. The proposed LS matching algorithm and 3D LS reconstruction algorithm are introduced in Sections 3 and 4, respectively. Experimental results are reported in Section 5, and conclusions are drawn in Section 6.

2. Related works

We give in this section a brief introduction of existing LS matching and 3D LS reconstruction methods.

2.1. Line segment matching

LS matching methods in existing literatures can generally be classified into two groups: methods that match LSs in individuals and those in groups. Some methods matching LSs in individuals exploit the photometric information in the local regions around LSs, like intensity (Baillard et al., 1999; Schmid and Zisserman, 1997), gradient (Verhagen et al., 2014; Wang et al., 2009b; Zhang and Koch, 2013), and color (Bay et al., 2005). All these methods underlie the assumption that there are considerable overlaps between corresponding LSs, which might lead to the failure of these methods when corresponding LSs share insufficient overlapping parts.

Other methods matching LSs in individuals leverage point matches for LS matching (Chen and Shao, 2013; Fan et al., 2010, 2012; Lourakis et al., 2000). These methods first find point matches using the existing point matching methods, and then exploit geometric invariants between coplanar points and line(s) under certain image transformations to evaluate LSs from two images. The LSs which meet the invariants are regarded to be in correspondence. A common disadvantage of these methods is that they depend heavily on point matching results so that once insufficient point matches were found before, these methods would generate inferior results.

Methods matching LSs in groups are more complex, but more constraints are available for disambiguation. Wang et al. (2009a) exploited the stability of the relative positions of the endpoints of a group of LSs in a local region under various image transformations to describe and match LS groups. This method is robust in