Image and Vision Computing 46 (2016) 47–63

Contents lists available at ScienceDirect

Image and Vision Computing

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

Piecewise-planar reconstruction using two views ☆

Michel Antunes [a], João P. Barreto [b], Urbano Nunes [b]

[a] Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg
[b] Institute for Systems and Robotics (ISR), Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal

ARTICLE INFO

Article history:
Received 21 November 2013
Received in revised form 4 November 2015
Accepted 16 November 2015
Available online 20 January 2016

Keywords:
Piecewise-planar reconstruction
SymStereo
Stereo-rangefinding
Semi-dense reconstruction

ABSTRACT

The article describes a reconstruction pipeline that generates piecewise-planar models of man-made environments using two calibrated views. The 3D space is sampled by a set of virtual cut planes that intersect the baseline of the stereo rig and implicitly define possible pixel correspondences across views. The likelihood of these correspondences being true matches is measured using signal symmetry analysis [1], which enables to obtain profile contours of the 3D scene that become lines whenever the virtual cut planes intersect planar surfaces. The detection and estimation of these *lines cuts* is formulated as a global optimization problem over the symmetry matching cost, and pairs of reconstructed lines are used to generate plane hypotheses that serve as input to PEARL clustering [2]. The PEARL algorithm alternates between a discrete optimization step, which merges planar surface hypotheses and discards detections with poor support, and a continuous optimization step, which refines the plane poses taking into account surface slant. The pipeline outputs an accurate semi-dense Piecewise-Planar Reconstruction of the 3D scene. In addition, the input images can be segmented into piecewise-planar regions using a standard labeling formulation for assigning pixels to plane detections. Extensive experiments with both indoor and outdoor stereo pairs show significant improvements over state-of-the-art methods with respect to accuracy and robustness.

1. Introduction

Stereo cameras are becoming increasingly popular because of the recent advent of 3D visualization and display. A few years ago they were considered special purpose devices that could only be found in research laboratories and high-end equipments, but nowadays they are a consumer electronics product being available either as standalone hand-held cameras (e.g. Fujifilm Finepix 3D and Sony Bloggie 3D), or integrated into smart-phones (e.g. HTC Evo 3D). Our work is motivated by this proliferation of stereo cameras that is expected to create an urge for robust algorithms able to render complete, photo-realistic 3D models in an automatic manner.

Stereo reconstruction is a classical problem in computer and robot vision that deserved the attention of thousands of authors [3,4]. Despite the many advances in the field, situations of poor texture, variable illumination, severe surface slant or occlusion are still challenging for most stereo matching methods, making it difficult to find a tuning that provides good results under a broad variety of acquisition circumstances [5]. Since man-made environments are dominated by planar surfaces, several authors suggested to overcome the above mentioned difficulties by using the planarity assumption as a prior for stereo reconstruction [6-10]. These approaches have the advantage of providing piecewise-planar 3D models of the scene that are perceptually pleasing and geometrically simple, and, thus, their rendering, storage and transmission is computationally less complex. This article proposes a pipeline for two-view *Piecewise-Planar Reconstruction* (PPR) understood as the detection and reconstruction of dominant planar surfaces in the scene.[1]

PPR is in a large extent a *chicken-and-egg* problem. If there is accurate 3D evidence about the scene, such as points, lines, vanishing directions, etc, then the problem of detecting, segmenting, and estimating the pose of dominant planes can be potentially solved using standard model fitting techniques [2,13]. On the other hand, if there is a prior knowledge about dominant planar surfaces in the scene, then the matching process can be constrained to improve the accuracy of the final 3D reconstruction, e.g. the known plane orientations

☆ This paper has been recommended for acceptance by Thomas Pock.

E-mail addresses: michel.antunes@uni.lu (M. Antunes), jpbar@isr.uc.pt (J. Barreto), urbano@isr.uc.pt (U. Nunes).

[1] We mean by PPR something that is different from approximating surfaces by small planes, as typically done in several dense stereo methods (e.g. [11,12])

http://dx.doi.org/10.1016/j.imavis.2015.11.008
0262-8856/

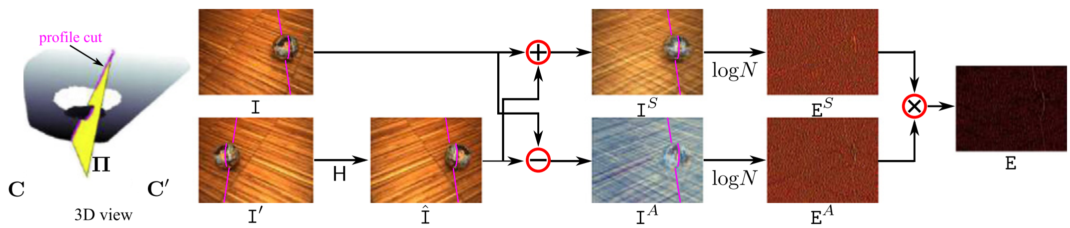

Fig. 1. The virtual cut plane $\mathbf{\Pi}$ (yellow) passes between the cameras and intersects the 3D scene in a non-continuous 3D curve (magenta). Let $\hat{\mathtt{I}}$ be the result of warping $\mathtt{I}'$ by the homography induced by $\mathbf{\Pi}$. The images $\mathtt{I}^S$ and $\mathtt{I}^A$ are, respectively, symmetric and anti-symmetric around the image of the profile cut (magenta). The output of the *logN* joint symmetry and anti-symmetry quantification method is the energy map $\mathtt{E}$ that highlights the image of the profile cut.

can be used to guide the stereo aggregation [11]. Existing methods for PPR typically comprise three steps that are executed sequentially:

1. *3D reconstruction:* The objective is to collect 3D evidence about the scene from multiple views. This evidence can either be obtained from *sparse stereo* that matches a sparse set of features across views (e.g. [8,9]), or from *dense stereo* that performs dense data association between frames by assigning to each pixel a disparity value (e.g. [10]).
2. *Plane hypotheses generation:* Given the 3D data, the objective is to detect and estimate the pose of planar surfaces using some sort of multi-model fitting approach.
3. *Plane labeling:* The goal is to assign to each pixel one of the plane hypotheses generated in the previous step. This is usually done using a *Markov Random Field* (MRF) formulation with photo-consistency being used as data term.

While most methods were originally designed to receive multiple views [6,7,8,9,10,14], we propose a pipeline that uses only two views and makes no assumptions about the scene other than the fact of being dominated by planar surfaces. The novelty is mainly in the steps of *3D Reconstruction* and *Plane Hypothesis Generation*, and the contributions can be summarized as follows:

- *Reconstruction of line cuts using Stereo from Induced Symmetry (SymStereo):* Establishing dense stereo correspondence is computationally expensive specially when dealing with high-resolution images. On the other hand, two-view sparse stereo tends to provide insufficient 3D data for establishing accurate plane hypotheses. Thus, we propose to carry a semi-dense reconstruction of the scene by independently recovering depth along a set of pre-defined virtual planes using SymStereo [1]. Since the intersections of the virtual scan planes with the planar surfaces in the scene are lines, we extract line segments from the profile cuts and use these *line cuts* to generate plane hypotheses.
- *Improving SymStereo accuracy in the case of surface slant:* In a similar manner to what happens in conventional stereo, surface slant affects the depth estimation obtained from SymStereo. In this case, the line cuts are poorly reconstructed and the plane surface estimation is inaccurate. We study the problem of surface slant in the context of the SymStereo framework and devise a simple solution that enables the reconstruction of highly slanted planes.
- *Global plane fitting:* Most methods for PPR treat stereo matching and plane detection in a sequential manner [6,14,7,8,9,10]. This is problematic because the accuracy of the plane hypotheses is inevitably limited by the accuracy of the initial 3D reconstruction that does not take into account the fact of the scene being dominated by planar surfaces. We carry the 3D reconstruction and the plane fitting in a simultaneous and integrated manner using the recent PEARL framework proposed in [2]. The algorithm alternates between a global discrete optimization step, which merges plane hypotheses and discards spurious detections, and a continuous optimization step over the symmetry energy, which refines the plane pose estimation taking into account surface slant. The output is a set of plane hypotheses and a semi-dense PPR of the 3D scene, where the reconstructed line cuts are labeled according to the plane detections.

1.1. Related work

Several works in PPR start by obtaining a sparse 3D reconstruction of the scene (e.g. point clouds and lines), then establish plane hypotheses by applying multi-model fitting to the reconstructed data, and finally use these hypotheses to guide the dense stereo process and/or perform a piecewise-planar segmentation of the input images. In Ref. [6], Werner and Zisserman rely in several cues and assumptions to find the dominant surface orientations and use plane-sweeping along the detected normal directions to reconstruct the scene. Bartoli [14] obtains an initial sparse point reconstruction from multiple views and applies a RANSAC-like algorithm for generating and scoring the plane hypotheses. In a similar manner, Pollefeys et al. [7] propose to detect planar surfaces in urban environments from sparse 3D point features and use the estimated normals for guiding plane-sweep stereo. Furukawa et al. [8] reconstruct 3D patches in textured image regions from multiple views using Ref. [15], and use the normals of these patches to establish plane hypotheses assuming a Manhattan-world model. These hypotheses are then used in a MRF formulation for pixel-wise plane labeling. In Ref. [9], Sinha et al. introduce a probabilistic framework for assigning plane hypotheses to pixels with the evidences of planar surfaces being provided by point cloud reconstruction, matching of line segments, and estimation of vanishing points. Gallup et al. [10] propose a stereo method capable of handling both planar and non-planar objects contained in the scene. A robust procedure based on RANSAC is used for fitting plane hypotheses to dense depth maps, followed by a MRF formulation for plane labeling of the input images.

These pipelines were originally designed to work with multiple images. Moreover, depth estimation and plane fitting are carried in a sequential and decoupled manner that, as discussed previously, has the drawback that errors in 3D evidence affect the accuracy of plane pose estimation, and the inferred planar surfaces are not used for refining the initial depth estimates.

An alternative strategy is to over-segment the stereo images based on color information and fit a 3D plane to each non-overlapping region. The number of planes to be considered is defined by the segmentation result, which acts as a smoothness prior during the global optimization. This segmentation information is either used as a hard minimization constraint [16,17,18] or as a soft constraint [19]. The main weakness of this type of strategy is the assumption that planar surfaces in the scene have different colors, which is often