Pattern Recognition 43 (2010) 494–505

Contents lists available at ScienceDirect

Pattern Recognition

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

Registration and interactive planar segmentation for stereo images of polyhedral scenes

Javier Flavio Vigueras*, Mariano Rivera

Center for Mathematical Research, Centro de Investigacion en Matematicas (CIMAT), Jalisco S/N, Colonia Valenciana, 36240 Guanajuato, Gto, Mexico

ARTICLE INFO

Article history:
Received 31 May 2008
Received in revised form 28 December 2008
Accepted 2 March 2009

Keywords:
Interactive computer vision
Stereo
Registration
Segmentation
Coplanarity
Color

ABSTRACT

We introduce a two-step iterative segmentation and registration method to find coplanar surfaces among stereo images of a polyhedral environment. The novelties of this paper are: (i) to propose a user-defined initialization easing the image matching and segmentation, (ii) to incorporate color appearance and planar projection information into a Bayesian segmentation scheme, and (iii) to add consistency to the projective transformations related to the polyhedral structure of the scenes. The method utilizes an assisted Bayesian color segmentation scheme. The initial user-assisted segmentation is used to define search regions for planar homography image registration. The two reliable methods cooperate to obtain probabilities for coplanar regions with similar color information that are used to get a new segmentation by means of quadratic Markov measure fields (QMMF). We search for the best regions by iterating both steps: registration and segmentation.

1. Introduction

Planar surfaces are often found in artificial manmade environments: outdoor scenes are commonly formed by polyhedral buildings (some examples are shown in Fig. 1); indoor scenes contain floors, walls, desks, etc. Planes have a constrained representation and ease various computer vision tasks such as camera calibration [10,34], camera localization [13,31], robot navigation [24], and 3D reconstruction [7,37]. Plane-based algorithms are commonly stable but they may become ill-conditioned when they are applied to wrong coplanar features, and therefore it is very important to know which image regions correspond to individual planes. By these reason, several works have been conducted on plane detection and segmentation.

1.1. Review of state-of-the-art segmentation approaches

A more general problem than the one issued in this paper is to estimate simultaneously the regions in the image corresponding to a given model (segmentation) and the set of parameters for each region model. In our specific case, each model corresponds to a planar surface.

If the model parameters are known, generic clustering algorithms as K-Means or isodata have been used with relative success [12]. Other algorithms consider spatial interactions among pixel labels providing useful constraints on the problem: region merging [8], active contour [5] approaches, eigendecomposition [39] and variational methods [29]. Among these, Bayesian formulations [16,18] have been successfully used for finding the solution to the segmentation problem. In this framework, the solution is computed by maximizing the a posterior probability distribution (MAP estimator).

In the general case, the model parameters are not known and some of these methods are extended using two-step procedures, following that an initial estimate is given and then the method: (1) estimates the model parameters given the segmentation, (2) estimates the label map (segmentation) given the model parameters, iterating these two steps until convergence [6,20,36].

However, the MAP estimator for the label field requires the solution of a combinatorial optimization problem. Graph-cuts-based algorithms [6] can be used for computing the exact MAP estimator in the case of binary segmentation or an approximation for problems with more than two classes, but make the two-step algorithm prone to be trapped in local minima.

A better strategy is to compute, instead of binary label variables, the probability that the observed data at each pixel is generated by a particular model (i.e. the posterior marginal distributions). Posterior marginal probabilities can be estimated with Markov chain Monte Carlo (MCMC) methods [16], by mean field (MF) approximations [36], or using a Gauss–Markov measure field (GMMF) model [19]. Nevertheless, MCMC approaches may be computationally

* Corresponding author.
E-mail address: flavio@cimat.mx (J.F. Vigueras).

0031-3203/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2009.03.012

expensive and the other two methods only guarantee convergence to a local maximum of the posterior distribution: if the procedure starts from a *wrong* initial set of parameters, the method may not converge to the global maximum [20]. On the other hand GMMF methods are not adequate for simultaneous estimation of the segmentation and the model parameters [20]. Recently, the quadratic Markov measure field (QMMF) approach was proposed [27], a computationally efficient method based on a Bayesian framework, that overcomes the GMMF model limitations, since the posterior marginal probabilities are efficiently computed by minimizing a quadratic, linearly constrained energy function; hence, any standard linear algorithm will converge to the global minimum.

1.2. Previous works on plane segmentation

Unsupervised plane detection and segmentation are commonly solved using sparse image key-points (using structure from motion techniques [2,31,32]), disparity maps [14,35], optical flow approaches [42], triangular surfaces [22] or range images [38]. Nevertheless, many of these approaches require to perform a 3D reconstruction; other methods assume that the plane is mostly textured [2], that a single plane is dominant in the image, or the camera require a rough calibration [1,35], constraining the range of application of those methods.

Matching sparse features often fails in cases like the tower (Fig. 1a) and the roof (Fig. 1b) because both planes have mostly the same texture. Although some heuristics (like RANSAC [9] or *a contrario* statistics [21]) have demonstrated to be successful on improving matching for a single surface having repeated structures (if a parametric model relating two views is given), their success is widely related with the ratio of inliers/outlier correspondences. However, when this heuristics are applied to distinguish between two or more planes, this ratio has a tendency to be very small and, as a result, the heuristics performance is noticeably poor.

On the other hand, untextured surfaces (like the three walls in Fig. 1c) are also difficult to distinguish as different planes for optical flow techniques, disparity maps and feature-based methods, because image features are usually mismatched.

Image acquisition from a moving camera over a polyhedral scene imposes known constraints on matching information for every couple of views and it is possible to extract planar segmentation from them without explicitly performing a 3D reconstruction [41]. Few approaches have tried to conduct segmentation on dense disparity models [33] but they often fail because several ambiguities arise on considering disparity information alone.

1.3. Interactivity

Several automatic solutions [4,6,15,28,40] have been developed for the planar segmentation problem from stereo views. Nevertheless, correct results are not reachable under some circumstances (described below in this paper), and only few automatic approaches may attain satisfactory results but in an unmanageable amount of time for some applications [17,30]. On the other hand, humans are capable to distinguish correctly distinct planar surfaces from a single image in a very short time. However, doing this task as a manual procedure could be unpractical. A completely user-assisted segmentation may be tedious if the boundaries of planar surfaces are not clearly distinguished into the image, when these boundaries have complex shapes, or when the interfaces are not friendly enough.

Computational complexity is not the only problem to cope. Automatic segmentation algorithms often have problems to make a decision on the number of region models to employ, thus, the number of planar surfaces observed in the images. Some approaches tend to over-segment the image, while others tend to merge different planes

Fig. 1. Example of planar piecewise environments. Right images acquired by a stereo pair of (a) tower scene, (b) roof scene and (c) house scene.

[17]. Interactive approaches are also very useful in this sense because they allow the user to designate by himself/herself the number of planes in the images, or correct an over-segmented or under-segmented partition by assistance.

It often happens also in stereo that a plane is visible in one image but invisible in another view. Without an appropriate occlusion management, the matching process certainly may not succeed or may match to a wrong region in the second image; hence, the corresponding planar projective transformation (homography) may be wrong due to the occlusion. This is one of the major problems to cope with the automatic initialization processes. The interactive region selection has the advantage that the user may prevent such a problem with his/her previous knowledge of the scene.

1.4. Proposed approach

Several computer vision methods seem to be very sensitive to the initial estimations required as input and, for many applications, choosing the adequate initial values may become an important or a tedious problem by itself. Interactive approaches have allowed to separate clearly the initialization stage from the automatic image processing, and improve the development of efficient computer vision tools due to the inclusion of knowledge given by human experts. We have chosen an assisted strategy to initialize our algorithm, although automatic planar segmentation methods have been found in the literature: interactivity may help to reduce computational time and to correct wrong segmentations.

In order to cope with the problems stated in Section 1.2, we propose a user-assisted segmentation method combining color information and motion matching of observed coplanar features in a two-image set. The novelty of this paper is to directly compute both homography parameters and dense region segmentation by means of a brief interaction with the user, instead of using unsupervised state-of-the-art methods that generally consider either case: (i) finding sparse coplanar points and then fitting a planar surface (implying occlusion and convexity problems) or (ii) computing general optical flow or disparity maps and later trying to fit a planar surface.

Our approach is based on a rough solution given into an initial stage by the user and an iterative two-step algorithm: (1) registering two views of the scene in order to find the corresponding planar homographies for every plane and (2) segmenting the image using a Bayesian segmentation approach. We refine the homography models using the new marginal a posteriori probabilities obtained from segmentation as a registration mask, and repeat the two-step procedure until convergence. As input for the planar segmentation step, model likelihoods are estimated by combining planar homographies fitting between the two views and their corresponding color information.

For the algorithm proposed in this paper, we have selected the QMMF segmentation approach. QMMF performance was already demonstrated by numerical experiments that compare this approach [27] with other state-of-the-art algorithms, such as graph-cut [6], random walker, Gaussian Markov measure field [19] and hidden Markov measure field methods [20].