Image and Vision Computing 70 (2018) 21–31

Contents lists available at ScienceDirect

Image and Vision Computing

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

Joint segmentation of color and depth data based on splitting and merging driven by surface fitting☆

Giampaolo Pagnutti, Pietro Zanuttigh*

Department of Information Engineering, University of Padova, Via Gradenigo 6B, Padova 35131, Italy

ARTICLE INFO

Article history:
Received 10 June 2016
Received in revised form 19 June 2017
Accepted 8 December 2017
Available online 14 December 2017

Keywords:
Segmentation
Depth
Spectral clustering
Kinect
NURBS

ABSTRACT

This paper proposes a segmentation scheme based on the joint usage of color and depth data together with a 3D surface estimation scheme. Firstly a set of multi-dimensional vectors is built from color, geometry and surface orientation information. Normalized cuts spectral clustering is then applied in order to recursively segment the scene in two parts thus obtaining an over-segmentation. This procedure is followed by a recursive merging stage where close segments belonging to the same object are joined together. At each step of both procedures a NURBS model is fitted on the computed segments and the accuracy of the fitting is used as a measure of the plausibility that a segment represents a single surface or object. By comparing the accuracy to the one at the previous step, it is possible to determine if each splitting or merging operation leads to a better scene representation and consequently whether to perform it or not. Experimental results show how the proposed method provides an accurate and reliable segmentation.

1. Introduction

Scene segmentation by way of images is a long-term research topic that, despite a huge amount of research, remains very challenging. There is a large number of works addressing image segmentation [1], but even the best performing ones are not able to provide a reliable solution in all conditions since it is very difficult to properly understand the underlying scene structure from a single image. The recent introduction of matricial Time-of-Flight range cameras and of structured-light consumer depth cameras (e.g., the two versions of Microsoft Kinect) has made geometry acquisition available to the mass market. Depth data is a very valuable aid for segmentation since it conveys information about the 3D structure of the scene, making the recognition of the various structures in it much easier. This allows to re-formulate the segmentation problem as the search for effective ways of partitioning a set of samples featuring color and geometry information. It resembles what happens inside the human brain where the disparity between the images seen by the two eyes is one of the clues used to separate the different objects together with prior knowledge and other features extracted from the color data acquired by the human visual system.

Following this rationale, this paper introduces a novel segmentation scheme capable to jointly exploit color and depth information. The segmentation is performed within a region splitting and merging framework. In the first stage, following the idea we introduced in Ref. [2], the segments are recursively split in two parts using a joint color and depth segmentation scheme based on spectral clustering [3]. In particular for the subdivision step, we employ a modified version of the approach proposed in Ref. [4], exploiting orientation information in addition to geometry and color data. After the subdivision a Non-Uniform Rational B-Spline (NURBS) surface is fitted on each of the two resulting segments and the fitting accuracies before and after the splitting are compared. The motivation behind this approach is that segments representing single objects or surfaces are accurately fitted, while segments encompassing multiple surfaces will lead to a poor fitting accuracy. This procedure is followed by a bottom up recursive merging strategy that combines segments belonging to the same object [5]. The algorithm detects the neighboring segments with consistent depth, orientation and color values along the common edges, and attempts to merge them. Again, the surface fitting accuracies before and after the merging are used to evaluate which merge operations are properly joining two parts of the same object. Summarizing, this work extends our two recent conference works, i.e., the region splitting scheme of Ref. [2] and the region merging scheme of Ref. [5], and combines them into a general segmentation framework that gives better results than each of the two separate methods. In particular, by combining the splitting and merging stages it overcomes some limitations of the previous works, e.g., the merging scheme was not able to recover from errors in the initial over-segmentation.

☆ This paper has been recommended for acceptance by Y. Aloimonos.
* Corresponding author.

https://doi.org/10.1016/j.imavis.2017.12.004
0262-8856/

The paper is organized as follows. After presenting the related works in Section 2, Section 3 describes the general workflow of the segmentation algorithm. The joint color and depth segmentation scheme used for the various splitting operations is recalled in Section 4, while Section 5 presents the employed surface fitting algorithm. Then the two steps of the method, i.e, the recursive splitting and merging algorithms, are described in Section 6 and in Section 7 respectively. The results are presented in Section 8 while Section 9 draws the conclusions.

2. Related works

As already pointed out, image segmentation is a long-term research field and a huge number of techniques based on different insights have been proposed. Approaches based on graph theory and on clustering algorithms have been particularly successful [3,6,7]. Anyway, despite a huge amount of research none of the existing methods is able to obtain completely satisfactory results, since segmentation is an ill-posed problem and it is very difficult to properly estimate the scene structure from color data alone.

Even if the usage of depth data for segmentation purposes is a recent research field, several approaches addressing scene segmentation by way of color and geometry information have been proposed in the last few years. A simple solution is to perform two independent segmentations from the color and the depth data, and then join the two results as in Ref. [8].

Clustering techniques can easily be adapted to joint depth and color segmentation and different approaches based on this idea have been proposed. In Ref. [9], Mean Shift clustering [6] is applied to 6D vectors containing both the color and spatial component for each sample, while superparamagnetic clustering is used in Ref. [10]. A joint color, spatial and directional clustering method coupled with a planar region merging scheme is used in Ref. [11] and in the refined version of the same approach proposed in Ref. [12]. The method of Ref. [13] exploits instead a multi-layer clustering strategy.

A method based on graph cuts has been applied for joint color and depth segmentation in Ref. [14]. The approach of Ref. [15] exploits Normalized Cuts segmentation [3] together with saliency maps. In Ref. [4], we proposed a segmentation scheme based on normalized cuts that is capable to automatically balance the relevance of color and depth.

Region splitting and region growing methods have also been used. The work of Ref. [16] starts from an initial superpixel segmentation and joins the segments exploiting a saliency metric. Superpixels and region merging are used also in Ref. [17] that uses a graph-based approach for the merging stage. Superpixels are combined together also in Ref. [18] where regions corresponding to planar surfaces are computed using an approach based on Rao-Blackwellized Monte Carlo Markov Chain. The approach has been extended for the segmentation of multiple depth maps in Ref. [19]. A bottom-up approach is used also by Ref. [20], that builds an adjacency graph of surface patches. We exploited surface fitting in Ref. [2], where we fitted NURBS surfaces over the segments. The fitting accuracy is then used to evaluate the consistency of the segmented regions in order to further split segments not encompassing a single surface in an iterative approach. The work in Ref. [5] applies the NURBS fitting scheme within a region merging procedure, starting from an initial over-segmentation and joining adjacent segments on the basis of the fitting accuracy. These two works constitute the starting point for the approach of this paper.

Hierarchical segmentation based on the output of contour extraction has been used in Ref. [21], that also deals with object detection from the segmented data. Another combined approach for segmentation and object recognition has been presented in Ref. [22], where an initial over-segmentation based on the watershed algorithm is followed by a hierarchical scheme. Combined segmentation and labeling is also addressed in Ref. [23] that exploits a MRF superpixel segmentation associated with a tree-structured approach. Finally, dynamic programming has been used in Ref. [24] to extract the planar surfaces in indoor scenes.

Some approaches deal with the close but less general problem of separating the foreground from the background [25–29]. In Ref. [29], two likelihood functions, based on color and depth data, are combined together for this task. The work of Ref. [25] uses two distinct Gaussian Mixture Models for the foreground in the depth and color spaces and combine them in a Bayesian Framework. Mixture of Gaussians are used in Ref. [26] as well. Both this approach and Ref. [27] consider also temporal constraints in depth and color videos. Finally, some works try to jointly solve the segmentation and stereo disparity estimation problems, e.g., Refs. [30–32].

3. General overview

The proposed segmentation procedure is divided into 3 main steps as depicted in Fig. 1. The initial pre-processing step produces the input data for the clustering algorithm, then a recursive splitting scheme (*top-down phase*) divides the scene into smaller and smaller segments and finally a region merging algorithm (*bottom-up phase*) combines together close segments belonging to the same object.

In the first phase, the color image and the depth map are converted to a unified representation. Color data, depth data and surface orientation information are represented by a set of 9D vectors containing the three sources of information.

The data is then recursively split into two segments using both color and depth information. At each step, the segment being processed is divided into two sub-segments using the approach based on spectral clustering described in Section 4. Then, a NURBS surface is fitted over each of the two sub-segments (Section 5). The fitting accuracy is compared to the one obtained at the previous step for the same segment. If the split operation has provided a better fitting then the process is iterated by recursively dividing the two sub-segments, otherwise it is stopped. The process is iterated until it is not possible to obtain any improvement by further subdividing any of the produced segments, as described in detail in Section 6.

At this point, the algorithm starts to analyze adjacent clusters corresponding to surfaces with a common contour and similar orientation and color properties. For each couple of candidates, the fitting accuracy on the union of the two segments is calculated and compared against the fitting accuracy on the original segments. If it is improved, the two segments are joined and replaced by the new resulting segment. The procedure continues until there are no more possible merging operations. This part is detailed in Section 7.

4. Joint color and depth segmentation

Following an approach similar to Refs. [4,5,33], before entering the proposed iterative clustering algorithm, a nine-dimensional representation of the scene samples is built from the geometry and color data. Firstly, the depth and color cameras are jointly calibrated. Various methods are possible for this task [34], we employed the method of Ref. [35] for the Kinect sensor and the one of Ref. [36] for ToF sensors. After calibration, it is possible to compute the coordinates (x, y, z) of each depth sample in 3D space and to reproject the depth samples on color data, thus associating to each sample also a vector containing the (R, G, B) color components. The surface orientation information, i.e, the surface normals (n_x, n_y, n_z) at each location, is also computed from the geometric data using the approach of Ref. [37]. Notice that the normals information, not considered in Refs. [4] and [2], is very useful to separate surfaces with similar colors and close spatial positions but with different orientations (e.g., the walls of a room).