Neurocomputing 161 (2015) 81–88

Contents lists available at ScienceDirect

Neurocomputing

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

Hierarchical segmentation of range images inside the combinatorial pyramid ☆

R. Marfil, E. Antúnez, A. Bandera *

Grupo ISIS, Dpto. Tecnología Electrónica, E.T.S.I. Telecomunicación, Universidad de Málaga, Campus de Teatinos s/n 29071-Málaga, Spain

ARTICLE INFO

Article history:
Received 8 June 2014
Received in revised form
11 December 2014
Accepted 18 January 2015
Available online 24 February 2015

Keywords:
Irregular pyramids
RGB-D cameras
Image segmentation

ABSTRACT

RGB-D cameras are not only able to provide color (Red-Green-Blue -RGB-) information from the scene but also a relatively accurate cloud of 3D points. Using information coming from this organized cloud, it is possible to define around each image pixel a small planar patch and obtain its normal vector. Within the framework of the combinatorial pyramid, this paper describes a method to abstract from these normals to parametric surface models. The method works at two consecutive stages. Firstly, normals are hierarchically grouped to divide up the image into superpixels. These superpixels capture small patches on the scene that belong to the same surface. Then, they are merged to segment the scene into simple geometric models. Curvature information and model information are used to divide up the image into planes, cylinders and/or spheres. This paper shows how, in the higher levels of abstraction of the combinatorial pyramid, scenes can be described using these geometric items and their topological relationships.

1. Introduction

Scene Understanding has become a popular topic in Computer Science which combines abilities such as perception, signal analysis and interpretation. The abilities related to understanding what we see in a visual scene are also called *visual recognition*. Visual recognition from 2D images has been widely explored [8,7]. However, due to the availability of affordable RGB-D sensors, a significant effort has been recently put into RGB-D scene understanding. As a significant advantage over their 2D counterparts, such 3D representations are able to incorporate physical information, such as the 3D volume of the objects, supporting relations and stability [20]. In this paper, we present an approach for segmenting and fitting depth maps or point clouds (range images) to shape primitives. Thus, this approach covers the first tasks of the visual recognition process. But the proposed framework could integrate in the future, within the same representation structure, the recognition task, as we recently demonstrated for planar images [2].

Segmenting a range image consists of dividing the input image into regions so that all the points of the same surface in the scene are grouped into the same region. The union of all these regions will generate the whole image, and there will not exist overlap between them. Given this definition, this segmentation process is not an ill-defined problem, as there is a 'good' result for each acquired image. Range segmentation has been widely employed not only for tridimensional (3D) object recognition [14], but also for 3D scene reconstruction or visualization.

Assuming the existence of a spatial coherence on each image surface, region-based approaches for range segmentation classify into the same segment those pixels that present similar properties. Typically, they have been grouped into region-growing and model-based approaches. Region-growing approaches to range segmentation segment the image into an initial set of regions, which are then merged or extended using a specific criterion. On the other hand, model-based approaches assume that surfaces in the scene are described by parametric models. The image is encoded as a cloud of 3D points and surface models are fitted to these points. Every data point which presents a small distance to the resulting fit is assumed to belong to this surface. Both approaches have advantages and disadvantages. Region-growing approaches work at region level, having problems to correctly define the boundaries on the image [3]. In addition, they require a large seed for good performance, which is usually difficult to locate, especially for complex scenes. Model-based approaches work at pixel level, but they need to choose the right underlying surface models. If these models are not correctly chosen, it might be an over- or an under-estimation of the true models, provoking a subsequent problem on the segmentation. Furthermore, model selection criteria are critical [15].

A major difference on the behavior of both kinds of approaches can be then observed. Using a large seed, region-growing approaches allow us to divide up the image into blobs belonging to different

☆This work has been partially granted by the Spanish Ministerio de Economía y Competitividad (MINECO) and FEDER funds under project TIN2012-38079-C03-03.

* Corresponding author.
E-mail address: rebeca@uma.es (R. Marfil).

http://dx.doi.org/10.1016/j.neucom.2015.01.075
0925-2312/

surfaces. The main problem arises when they try to merge larger regions, as they are not usually able to correctly deal with the boundaries on the image. On the contrary, model-based approaches have problems to define the models at low level, as noise usually makes difficult to choose the correct model. But, when the underlying model for each surface in the scene is known, they are able to provide a robust segmentation [3]. The solution can then be to work at two consecutive stages: data is first pre-segmented into patches and then model-based surfaces are fitted [15]. This paper proposes to apply both stages within the same hierarchical representation of the input image. For this end, the range image is encoded using an irregular pyramid. Irregular pyramids represent the image as a stack of graphs with a decreasing number of nodes. Each layer of the pyramid is built from the layer below by applying a specific decimation process. This process allows us to adapt the structure of the pyramid to the image content. Given the behavior of the region-based approaches for range segmentation, we propose to use a region-based criterion for building the lower layers of the pyramid and a model-based one for building the higher ones. Thus, the first approach provides a decomposition of the range image into quasi-planar patches, which are associated to a given surface of the scene. The resulting decomposition reduces image complexity while avoiding under-segmentation. These patches are then fitted to parametric surface models (planes, cylinders and spheres) using RANSAC [18,10]. The process is summarized in Fig. 1.

Irregular pyramids can use a simple graph (i.e. a region adjacency graph (RAG)) to encode each layer of the hierarchy. However, objects and scenes are not only characterized by features or parts, but also by the spatial relationships among these parts. RAGs do not permit us to know if two adjacent regions have one or more common boundaries, and they do not allow us to differentiate an adjacency relationship between two regions from an inclusion relationship. Instead of simple graphs, each level of the hierarchy can be represented using a pair of dual graphs or a combinatorial map. In this paper, we employ the second type of encoding, being the input scene represented by a stack of combinatorial maps of reduced resolution. This hierarchy is called a combinatorial pyramid [4,1]. Specifically, the combinatorial pyramid [4] is defined by an initial combinatorial map that can be successively reduced using the general scheme proposed by Kropatsch [13].

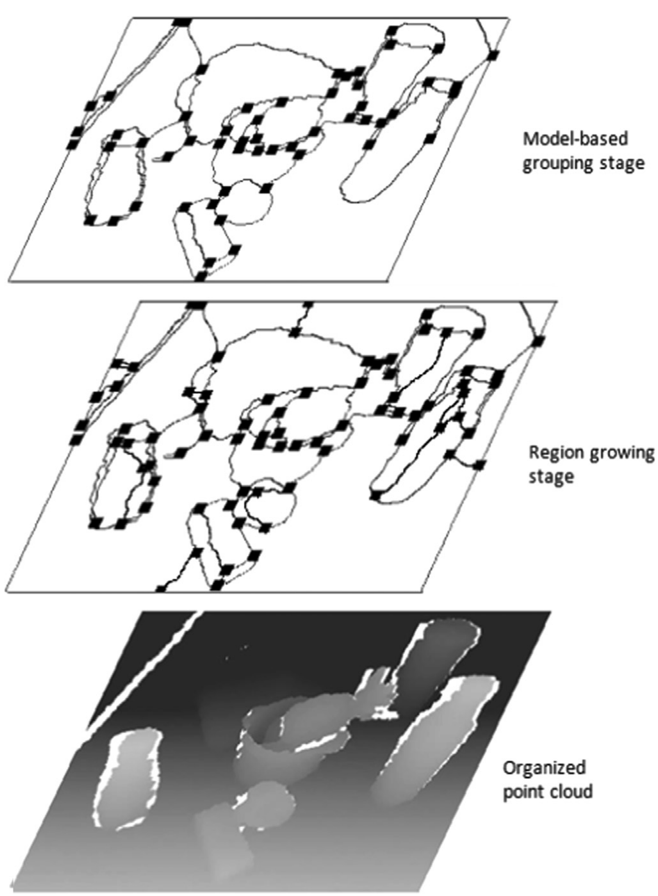

Fig. 1. Schematic overview of the proposed approach: normals obtained from the organized point cloud are firstly pre-segmented into superpixels and then, these superpixels are grouped into surfaces.

1.1. Contribution and outline of the paper

In the multiscale framework provided by the combinatorial pyramid, this paper presents an approach to the segmentation of range images. Contributions include

- A novel, multi-stage algorithm to combine data- and model-driven processes for surface description inside the hierarchy of the combinatorial pyramid.
- Region merging is conducted using different metrics inside the same hierarchy, generating a representation of the image at different levels of abstraction or scales. At the base level, depth information is considered to estimate the normals for every image pixel. Then, at low scales, image pixels are grouped into planar patches taking into account proximity and continuity criteria. These patches or superpixels [16] reduce image complexity while avoiding under-segmentation, and they are described by their centroids and normals. At high levels, superpixels are grouped into larger structures (planes) using the same scheme. Curvature information is used as a previous estimator for testing if two superpixels should be merged. However, RANSAC is employed to evaluate if those segments which are not fully explained using a planar model fit a sphere or a cylinder model.

The proposed approach has been evaluated using real images acquired with the Kinect from Microsoft and annotated images from the Object Segmentation Database (OSD) [17]. Results show that it can be compared with other leading approaches. Furthermore, with respect to these other methods, the proposed framework presents the main advantage that the combinatorial pyramid preserves at all levels of the hierarchy the topological relationships of the original image. Thus, the decomposition of the image into regions at each level is represented by a combinatorial map that encodes correctly these relationships [4,6].

The rest of the paper is organized as follows: the approach is presented in Section 2. Experimental results revealing the efficiency of the proposed method are presented in Section 3. Finally, the paper concludes along with discussions and future work in Section 4. A brief summary about the main terms related to combinatorial maps and pyramids is provided at Appendix A.

2. The segmentation approach

2.1. Overview of the proposed approach

The key idea in the proposed approach is to reduce the perceptual grouping computation to an efficiently solvable clustering problem. Fig. 1 should be correctly understood: the layers of the hierarchy are not images but combinatorial maps whose size is reduced when we go up in the hierarchy. Furthermore, there are arcs that link the data at each layer with the pixels of the original range image. Thus, it is possible to obtain a segmentation result from the map encoding each layer of the hierarchy. Fig. 1 shows two of these maps but the number