



Learning of perceptual grouping for object segmentation on RGB-D data [☆]



Andreas Richtsfeld*, Thomas Mörwald, Johann Prankl, Michael Zillich, Markus Vincze

Vienna University of Technology, Automation and Control Institute (ACIN), Gusshausstraße 25-29, 1040 Vienna, Austria

ARTICLE INFO

Article history:

Available online 18 April 2013

Keywords:

Computer vision
Object segmentation
Perceptual organization
RGB-D images
B-spline fitting
Object reconstruction
SVM learning
Graph-based segmentation

ABSTRACT

Object segmentation of unknown objects with arbitrary shape in cluttered scenes is an ambitious goal in computer vision and became a great impulse with the introduction of cheap and powerful RGB-D sensors. We introduce a framework for segmenting RGB-D images where data is processed in a hierarchical fashion. After pre-clustering on pixel level parametric surface patches are estimated. Different relations between patch-pairs are calculated, which we derive from perceptual grouping principles, and support vector machine classification is employed to learn Perceptual Grouping. Finally, we show that object hypotheses generation with Graph-Cut finds a globally optimal solution and prevents wrong grouping. Our framework is able to segment objects, even if they are stacked or jumbled in cluttered scenes. We also tackle the problem of segmenting objects when they are partially occluded. The work is evaluated on publicly available object segmentation databases and also compared with state-of-the-art work of object segmentation.

© 2013 The Authors. Published by Elsevier Inc. All rights reserved.

1. Introduction

Wertheimer, Köhler, Koffka and Metzger were the pioneers of studying Gestalt psychology, when they started to investigate this theory about hundred years ago. Wertheimer [1,2] first introduced *Gestalt principles* and Köhler [3], Koffka [4] and Metzger [5] further developed his theories. A summary and more recent contributions can be found in the modern textbook presentation of Palmer [6]. *Gestalt principles* (also called *Gestalt laws*) aim to formulate the regularities according to which the perceptual input is organized into unitary forms, also referred to as wholes, groups, or Gestalten [7]. In visual perception, such forms are the regions of the visual field whose portions are perceived as grouped or joined together, and are thus segregated from the rest of the visual field. These phenomena are called *laws*, but a more accurate term is *principles of perceptual organization*. The principles are much like heuristics, which are mental short-cuts for solving problems. *Perceptual organization* can be defined as the ability to impose structural organization on sensory data, so as to group sensory primitives arising from a common underlying cause [8]. In computer vision this is more often called *perceptual grouping*, when Gestalt principles are used to group visual features together into meaningful parts, unitary forms or objects.

There is no definite list of *Gestalt principles* defined in literature. The first discussed and mainly used ones are *proximity*, *continuity*, *similarity*, *closure* and *symmetry*, defined by Wertheimer [1], Köhler [3], Koffka [4] and Metzger [5]. *Common region* and *element connectedness* were later introduced and discussed by Rock and Palmer [9–11]. Other principles are *common fate*, considering similar motion of elements, *past experience*, considering former experience and *good Gestalt (form)*, explaining that elements tend to be grouped together if they are part of a pattern, which describes the input as simple, orderly, balanced, unified, coherent and as regular as possible. For completeness we also have to mention the concept of *figure-ground articulation*, introduced by Rubin [12]. It describes a fundamental aspect of field organization but is usually not referred to as a *Gestalt principle*, because this term is mostly used for describing rules of the organization of somewhat more complex visual fields. Some of these rules are stronger than others and may be better described as tendencies, especially when principles compete with each other.

Perceptual grouping has a long tradition in computer vision. But many especially of the earlier approaches suffered from susceptibility to scene complexity. Accordingly scenes tended to be “clean” or the methods required an unwieldy number of tunable parameters and heuristics to tackle scene complexity. A classificatory structure for perceptual grouping methods in computer vision was introduced by Sarkar and Boyer [13] in their review of available systems. They listed representative work for each category at this time and updated it later in [8]. More than ten years after Sarkar and Boyer wrote their status on perceptual grouping, cheap and powerful 3D sensors, such as the Microsoft Kinect or Asus Xtion, became available and sparked a renewed interest in 3D

[☆] This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

* Corresponding author.

E-mail address: ari@acin.tuwien.ac.at (A. Richtsfeld).

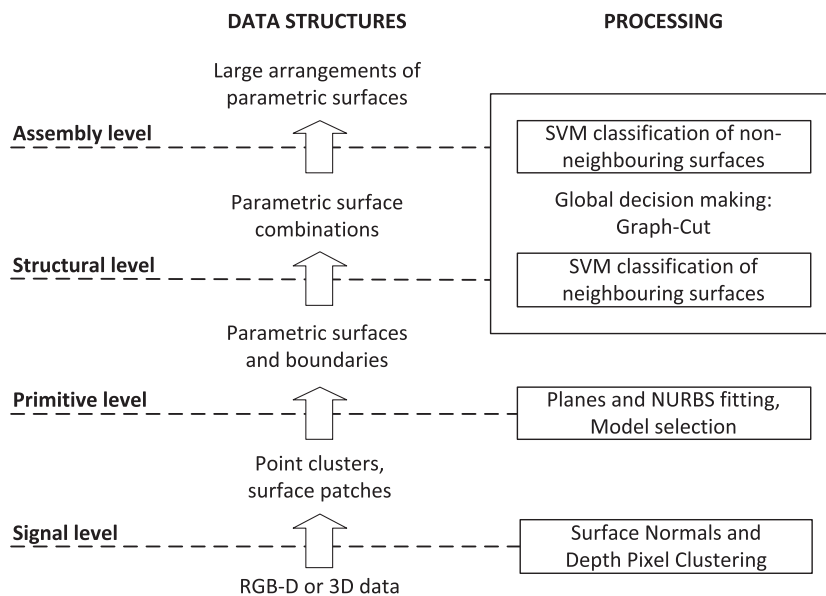


Fig. 1. System overview: From raw input data to object hypotheses.

methods throughout all areas of computer vision. Making use of 3D or RGB-D data can greatly simplify the grouping of scene elements, as structural relationships are more readily observable in the data rather than needing to be inferred from a 2D image.

In describing our system we follow the structure of Sarkar and Boyer [13,8], where input data is organized in bottom-up fashion, stratified by layers of abstraction: *signal*, *primitive*, *structural* and *assembly level*, see Fig. 1. Raw sensor data, occurring as RGB-D data, is grouped in the signal level to point clusters (surface patches), before the primitive level produces parametric surfaces and associated boundaries. Perceptual grouping principles are learned in the assembly and structural level to form groupings of parametric surface patches. Finally, a globally optimal segmentation is achieved using Graph-Cut on a graph consisting of surface patches and their learned relations.

Signal level – Raw RGB-D images are pre-clustered based on depth information. The relation between 2D image space and the associated depth information of RGB-D data is exploited to group neighboring pixels into patches.

Primitive level – The task on the primitive level is to create parametric surfaces and boundaries from the extracted pixel clusters of the signal level. Plane and B-spline fitting methods are used to estimate parametric surface representations. Model Selection finds the best representation and therefore the simplest set of parametric models for the given data.

Structural level – Features, derived from Gestalt principles, are calculated between neighboring surface patches (in the 3D euclidean space) and a feature vector is created. During a training period, feature vectors and ground truth data are used to train a support vector machine (SVM) classifier to distinguish between patches belonging to the same object and belonging to different objects. The SVM then provides a value for each feature vector from a neighboring patch pair which represents the probability that two neighboring patches belong together to the same object.

Assembly level – Groups of neighboring parametric surfaces are available for processing. Feature vectors are again constructed from relations derived from Gestalt principles, but now between non-neighboring surface patches of different parametric surface groups. A second SVM is trained to classify based on this type of feature vector. Creating object hypotheses directly from the assembly level is difficult, as the estimated probability values from the

SVM are only available between single surfaces, but not between whole groupings of surface patches from the structural level. Wrong classifications by the SVMs (which after all only perform a local decision) pose a further problem, possibly leading to high under-segmentation of the scene for only a few errors.

Global Decision Making – To overcome these problems, the decision about the optimal segmentation has to be made on a global level. To this end we build a graph where parametric surfaces from the primitive level represents nodes and the above relations implementing Gestalt principles represent edges. We then employ Graph-Cut using the probability values from the SVM of the assembly level as well as from the structural level as energy terms of the edges to finally segment the most likely connected parts, forming object hypotheses.

The main contribution of our work is the combination of perceptual grouping with SVM learning following a designated hierarchical structure. The learning approach of the framework enables segmentation of unknown objects of reasonably compact shape and allows segmentation for a wide variety of different objects in cluttered scenes, even if objects are partially occluded. Fig. 2 shows segmentation of a complex scene, processed with the proposed framework. Furthermore, the system provides beside image segmentation a parametric model for each object, enabling efficient storage for convenient further processing of the segmented structures.

The paper is structured as follows: The next section discusses representative related work and sets the work in context. Sections 3, 4, 5 and 6 explain the bottom-up processing for each data abstraction level before Section 7 shows global decision making. Experiments and evaluation results are presented in Section 8 and the work ends with a conclusion and a outlook in Section 9.

2. Related work

Many state-of-the-art approaches in literature formulate image segmentation as energy minimization with an MRF [14–17]. Reasoning on raw sensor data without usage of any constraints is a hard and ill-defined problem. So various approaches added constraints using a shape or a location prior, others exploited active segmentation strategies.

Download English Version:

<https://daneshyari.com/en/article/10360053>

Download Persian Version:

<https://daneshyari.com/article/10360053>

[Daneshyari.com](https://daneshyari.com)