# A Semantic Region Growing Approach in Image Segmentation and Annotation

Thanos Athanasiadis, Yannis Avrithis and Stefanos Kollias

Image, Video and Multimedia Systems Laboratory
School of Electrical and Computer Engineering
National Technical University of Athens
9, Iroon Polytechniou St., 157 73 Zographou, Greece
{thanos,iavr}@image.ntua.gr, stefanos@cs.ntua.gr

**Abstract.** In this position paper we examine the limitation of region growing segmentation techniques to extract semantically meaningful objects from an image. We propose a region growing algorithm that performs on a semantic level, driven by the knowledge of what each region represents at every iteration step of the merging process. This approach utilizes simultaneous segmentation and labeling of regions leading to automatic image annotation.

## 1. Introduction

Automatic segmentation of images is a very challenging task in computer vision and one of the most crucial steps toward image understanding. A variety of applications such as object recognition, image annotation, image coding and image indexing, utilize at some point a segmentation algorithm and their performance depends highly on the quality of the latter. It is acknowledged that ages-long research has produced algorithms for automatic image [1] and video [2] segmentation, structuring of multimedia content [3] and recognition of low-level features within such content [4]. Comparatively to this effort, little progress has been made on machine-generated semantic descriptions of audiovisual information in a way familiar to humans. Still, human vision perception outperforms state-of-the-art computer's segmentation algorithms. The main reason for this is that human vision is based also in high level prior knowledge about the semantic meaning of the objects that compose the image.

We propose a segmentation technique that belongs to the general framework of region growing segmentation algorithms [5],[2]. Region growing algorithms start from an initial partition of the image and then an iteration of region merging begins, based on certain similarity criteria until the predefined termination criteria are met. Our contribution is an additional merging process that in comparison to previous merging, its criteria are not based on syntactic features like color or texture similarity, but on matching of concepts associated to each region. In other words, after a certain point where syntactic region merging stops, an initial region labeling is carried out using low-level features and detectors [6] and then segmentation continues based this time on fuzzy criteria that apply on a semantic level, i.e. the assigned concepts to each region along with a corresponding confidence value.

## 2. Semantic Region Growing Algorithm

The target of this novel algorithm is to improve both segmentation and recognition of objects at the same time, with obvious benefits for semantic annotation of images. In the following two subsections we describe the foundations of the Semantic Region Growing (SRG) algorithm, which are the graph representation of the images and the initial selection of the seeds. Finally the proposed algorithm is examined in subsection 2.3.

### 2.1 Graph Representation of an Image

An image can be described as a structured set of individual objects, allowing thus a straightforward mapping to a graph structure. In this fashion, many image analysis problems can be considered as graph theory problems, inheriting the solid theoretical grounds of the latter. Attributed Relation Graphs (ARGs) are a type of graph often used in computer vision and image analysis for the representation of structured objects. In this work we adopt the formal representation of an ARG given by Berreti et al in [7], where an ARG is defined precisely by spatial entities represented as a set of vertices $E$, each labeled with an attribute $a$ and, binary spatial relationships represented as pairs of vertices $E \times E$ each labeled with a spatial descriptor $w$.

In particular, the vertex's attribute $a$ is a complex structure that contains the following two (also complex) entities:

1. Three MPEG-7 Visual Descriptors that describe the low-level features of the corresponding region, namely *Dominant Color*, *Region Shape* and *Homogeneous Texture*.
2. A list of candidate labels, along with a degree of confidence for each one. This is the result of the initial region labeling, discussed briefly in the following section.

The spatial descriptor $w$ contains information regarding the spatial relation of the regions, which are actually extracted but not utilized so far from the algorithm under discussion, remaining hence an open issue for future research.

### 2.2 Initialization of Region Labeling

Our intention is to work on a higher level of information where regions are linked to possible labels rather than only to their visual features. The above described ARG contains low-level information extracted directly by the image itself, but it also has labels and confidence values assigned by a knowledge-assisted analysis (KAA) algorithm, discussed in depth in a previous work [6]. For each vertex (i.e. a region of the image) of the ARG a matching process is performed between the visual descriptors stored in the vertex and the corresponding visual descriptors of concepts, stored in the form of prototype instances in an ontological knowledge base. This process results to an initial fuzzy labeling of the regions with concepts from the knowledge base. This is of course not a simple task and results depend highly on the domain where it is applied, as well as on the quality of the knowledge base.

### 2.3 SRG Algorithm Description

Conducting thorough experiments trying to improve the results of the KAA algorithm, we came up with the idea presented in this paper: To adapt a well known segmentation technique, like region growing, to the problem of semantic annotation. More specifically, we adopt a watershed-like region merging [8] technique, starting from regions-seeds that are automatically selected.

Let us now introduce the necessary mathematical notation used in this paper. The Semantic Region Growing (SRG) algorithm acts on a higher level than other region growing algorithms; this higher level we call it *Semantic Level*:

$$SL = \left\{ \left\langle L^{g_q}, D^{g_q} \right\rangle \right\}, \forall g_q \in RG \quad \textbf{(1)}$$

The fuzzy set $SL$ that represents the *Semantic Level*, consists of pairs $\left\langle L^{g_q}, D^{g_q} \right\rangle$, for all regions of the image: $\forall g_q \in RG$, where $RG$ : the set of all regions in the image . The entities $L^{g_q}$ and $D^{g_q}$ are two sets containing for the specific region $g_q$ all candidate labels and confidence values respectively and, are defined as:

$$L^{g_q} = \{l_k\} \subseteq L, \text{ where: } g_q \in RG, l_k \in L, \quad \textbf{(2)}$$
$$\text{where } L \text{ : the set of all possible labels}$$

$$D^{g_q} = \left\{ d_{l_k}^{g_q} \right\}, \quad \textbf{(3)}$$
$$\text{where } d_{l_k}^{g_q} \in [0,1]$$

The aforementioned pair formulates that every single region $g_q$ has been assigned to a number of candidate labels (equation 2) accompanied by the respective confidence values (equation 3):

$$\forall g_q \in RG \text{ : has a set of labels } L^{g_q} = \{l_k\} \subseteq L$$
$$\forall\, l_k \text{: has a confidence value } d_{l_k}^{g_q} \quad \textbf{(4)}$$

A number of regions $g_q$ are selected to be used as seeds for the initialization of the SRG algorithm and form an initial set, let it be $S$ . The criteria for selecting a region to become a seed are two: i) The region's best confidence value should be above a threshold. ii) the rest concepts have low confidence values. These two constrains ensure that the specific region has been correctly selected as seed of the particular concept.

An iterative process begins that checks whether the direct neighbors (as defined in the ARG) of the initial regions-seeds have been assigned to the same concept its propagator region-seed has and, with what confidence value. Some of those regions, that satisfy two additional criteria, form a new set of regions $N^i$ ($i$ denotes the iteration step, with $N^0 @ S$), which will be the new seeds for the next iteration of the algorithm. These two criteria are:

1. Confidence value of the propagator region $g_p$ for the particular label $l_k$ should be above a threshold: $d_{l_k}^{g_p} > T_{prop}$
2. Confidence value of the region under examination $g_q$ for the same label $l_k$ should be above another threshold: $d_{l_k}^{g_q} > a^i \cdot T_{child}$, where $a$ is a constant slightly above one, that increases the threshold in every iteration $i$ of the algorithm in a non-proportional way to the distance from the initial regions-seeds.

When the above criteria are satisfied, region $g_q$ is merged with its propagator $g_p$ and its confidence value is re-evaluated as the minimum between their confidence values, thus: $d_{l_k}^{\hat{g}_q} = \min\left(d_{l_k}^{g_p}, d_{l_k}^{g_q}\right)$

The termination criteria of the algorithm are quite straightforward: Repeat while the set of regions-seeds in step $i$: $N^i \neq \varnothing$. In this point, we should underline that when neighbors of a region are examined, previous accessed regions are excluded, i.e. each region is reached only once and that is by closest region-seed, as defined in the ARG.

Schematically, this algorithm looks like clusters of regions (each cluster corresponding to a specific concept) expanding in every iteration, until either the coherency of the cluster is smaller than allowed to be, or the borders of two such clusters meet. We use the term watershed-like because the decision for which regions to be merged depends on both their confidence value and their distance from the seed (catchment basin, in watershed segmentation terminology) and the iteration keeps on until two expanded regions meet (basins are flooded till the watershed).

## 3. References


1. H. Gao, W.-C. Siu and C.-H. Hou, "Improved techniques for automatic image segmentation", IEEE Trans. on Circuits and Systems for Video Technology, vol.11, no. 12, pp. 1273-1280, December 2001
2. P. Salembier, F. Marques, "Region-Based Representations of Image and Video - Segmentation Tools for Multimedia Services", IEEE Trans. on Circuits and Systems for Video Technology, vol.9, no.8, December 1999
3. S. Chang and H. Sundaram, "Structural and Semantic Analysis of Video", IEEE International Conference on Multimedia and Expo (II), 2000
4. E.L.Andrade Neto, J.C.Woods, E.Khan, M.Ghanbari "Region Based Analysis and Retrieval for Tracking of Semantic Objects and Provision of Augmented Informa-

tion in Interactive Sport Scenes“, IEEE Trans. on Multimedia Vol. 7, Issue 6, Dec. 2005 Page(s):1084 – 1096

5. R. Adams and L. Bischof, “Seeded Region Growing”, IEEE Trans. on Pattern Analysis and Machine Intelligence, vol 16, no. 6, pp. 641-647, June 1994
6. T. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis and Y. Kompatsiaris, “Using a Multimedia Ontology Infrastructure for Semantic Annotation of Multimedia Content“, Proc. of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05)
7. S. Berretti, A. Del Bimbo, E. Vicario, “Efficient matching and indexing of graph models in content-based retrieval”, IEEE Trans. on Circuits and Systems for Video Technology, vol.11, no. 12, pp. 1089-1105, December 2001
8. S. Beucher and F. Meyer, “The Morphological Approach to Segmentation: The Watershed Transformation”, in: Mathematical Morphology in Image Processing, E.R.Doughertty (Ed.), Marcel Dekker, NY, 1993.