Rule-based Reasoning for Semantic Image Segmentation and Interpretation

Petr Berka, Thanos Athanasiadis and Yannis Avrithis

Abstract— **In this paper, we propose the application of rule-based reasoning for knowledge assisted image segmentation and object detection. A region merging approach is proposed based on fuzzy labeling and not on visual descriptors, while reasoning is used in evaluation of dissimilarity between adjacent regions according to rules applied on local information.**

Index Terms— **Knowledge-assisted analysis, rule-based reasoning, semantic region merging.**

I. Introduction

Automatic segmentation of images is a very challenging task in computer vision and one of the most crucial steps toward image understanding. Although a great effort has been consumed in designing generic, robust and efficient segmentation algorithms [1], still human vision perception outperforms any state-of-the-art computer algorithm. One 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.

Knowledge assisted analysis can be defined as a tightly coupled and constant interaction between low level image analysis and higher level knowledge. In this paper we propose an algorithm that involves simultaneously both segmentation and detection of simple objects. Starting from traditional graph-based segmentation, the proposed technique continues with fuzzy region labeling and semantic region merging. More specifically the latter is based on rule-based reasoning, using knowledge about the possible labels of the candidate for merging regions and of their direct neighbors, to improve the initial segmentation and labeling.

To perform rule-based reasoning, we use the expert system NEST [2], developed at the University of Economics, Prague. This system follows the idea of compositional approach to inference introduced in mid 70s by the early expert systems MYCIN and PROSPECTOR. The rules used by NEST are in the form:

IF condition THEN conclusion (w)

where *condition* is a conjunction of propositions, *conclusion* is a single proposition and weight *w* expresses the uncertainty of the conclusion if the condition is true. During consultation, all rules for which the condition is true with some positive degree (weight) are activated and their contributions are used to compute the weights of goals.

P. Berka is with the Dept. of Information and Knowledge Engineering, University of Economics, Prague.

Th. Athanasiadis and Y. Avrithis are with Image, Video and Multimedia Systems Laboratory, National Technical University of Athens.

II. Initial Segmentation and Region Labeling

Our intention is to operate on a higher level of information where image regions are linked to possible labels rather than only to their visual features. For the representation of an image we adopt the Attributed Relational Graph (ARG) [3]. An ARG is a graph structure that holds a region-based representation of the image, where the set of vertices V corresponds to image regions and the set of edges E to links between adjacent regions. The ARG is constructed based on an initial color RSST segmentation that produces a few tens of regions. Each vertex of the graph holds the Dominant Color, Region Shape and Homogeneous Texture MPEG-7 visual descriptors extracted for this specific region.

Based on these features and the adjacency information provided by the ARG, the target is to assign to each region a a fuzzy set of labels L_a. In order to achieve this (for more details please refer to previous work in [4]), we compute a matching distance value between each one of these regions and each one of the prototype instances of all concepts in the domain ontology. This process results to an initial fuzzy labeling of the regions with concepts from the knowledge base, i.e. for region a we have the fuzzy set $L_a = \sum_k c_k/w_k$, where k is the cardinality of the (crisp) set of all concepts $C = \{c_k\}$ in the knowledge base and $w_k = \mu_a(c_k)$ is the degree of membership of element c_k in the fuzzy set L_a.

III. Rule-Based Image Analysis

This section presents the integration of a rule based reasoning system into an image segmentation algorithm. Based on the foundations described in the previous section, we introduce a novel segmentation algorithm that relies on fuzzy region labeling and rules to solve the problem of image oversegmentation.

A. Semantic region merging

Recursive Shortest Spanning Tree, or simply RSST, is a bottom-up segmentation algorithm that begins from the pixel level and iteratively merges neighbor regions according to a distance value until certain termination criteria are satisfied. This distance is calculated based on color and texture characteristics, which are independent of the area's size. In every step the two regions with the least distance are merged; visual characteristics of the new region are extracted and all distances are updated accordingly.

We introduce here a modified version of RSST, called Semantic RSST (S-RSST) that aims to improve the usual oversegmentation results by incorporating region labeling in the segmentation process [5]. In this approach the distance between two adjacent regions a and b (vertices v_a and v_b in the graph) is calculated using NEST, in a fashion described later on, and this dissimilarity value is assigned as the weight of the respective graph's edge e_{ab}.

Let us now examine in detail one iteration of the S-RSST algorithm. Firstly, the edge e_{ab} with the least weight is selected, then regions a and b are merged. Vertex v_b is removed completely from the ARG, whereas v_a is updated appropriately. This update procedure consists of the following two actions:

1) Re-evaluation of the degrees of membership of the labels fuzzy set in a weighted average (w.r.t. the regions' size) fashion.
2) Re-adjustment of the ARG edges by removing edge e_{ab} and re-evaluating the weight of the affected edges invoking NEST.

This procedure continues until the edge e^* with the least weight in the ARG is bigger than a threshold: $w(e^*) > T_w$. This threshold is calculated in the beginning of the algorithm, based on the histogram of all weights in E.

B. Estimation of regions dissimilarity

Expert system NEST serves in knowledge-assisted analysis as an estimator for the dissimilarity of two adjacent regions a and b in the image. The input consists of the fuzzy sets L_a and L_b of the two regions, with membership functions μ_A and μ_B respectively as well as the fuzzy set of their direct neighbors. More specifically, we consider as direct neighbors only the four dominant directional neighbors of regions a and b (north, south, west, east). With this input, NEST is able to reason about the (dis)similarity between two adjacent regions, according to rules applied on local information of what each region might represent given its neighborhood. The rule base is structured into two layers. The first layer determines the dominant label(s) of the regions according to the initial labeling of regions and their neighbors; the confidence of the initial labeling (expressed in these rules as "high", "medium", "low") is derived from the initial fuzzy labels. This layer contains 60 rules; a rule for each combination of label, of its confidence, of the region and its neighbor. The next layer increases/decreases the similarity of regions a and b if they share/don't share the same dominant label. This layer contains 48 rules; a rule for each combination of possible dominant labels of the two regions. All rules have been created empirically, by a knowledge engineer. Example rules for the first and second layer look like follows:

```
IF A_conf_sky(medium) AND A_conf_sea(medium) AND
A_conf_sand(low) AND leftofA_conf_sky(high) AND
rightofA_conf_sky(high) THEN A_dominant(sky) (0.8)

IF A_dominant(sky) AND B_dominant(sky,sea)
THEN A_B_similar (0.7)
```

Inferred weight of the (dis)similarity proposition is used to drive segmentation based on semantic information and to resolve oversegmentation problems.

IV. First Experiments

We will illustrate the rule-based reasoning step on a following simple example. Let 0, 2, 3, 7, 8 be five regions of an exemplar image of a beach scene, where 0, 2, 3, 8, are adjacent to 7. Table I illustrates the associated fuzzy set of each region, i.e. the degrees of membership of each concept. Correct concepts for each region, according to the ground truth, are highlighted with bold letters.

TABLE I
INITIAL FUZZY LABELING

Regions	Concepts					
	Sky	Sea	Cliff	Plant	Sand	Person
0	**0.66**	0.82	0.47	0.25	0.58	0.75
2	**0.74**	0.77	0.44	0.34	0.69	0.62
3	0.70	0.52	0.34	0.40	0.61	**0.73**
7	**0.78**	0.79	0.38	0.35	0.55	0.39
8	**0.66**	0.79	0.64	0.27	0.59	0.37

From this input, the rule base infers as dominant labels *Sea* and *Person* for 0, *Sea* and *Sky* for 2, *Person* and *Sky* for 3, *Sea* and *Sky* for 7, and *Sea* for 8. This will result in the following ordering of region pairs according to their similarity: 7-2 (most similar), 7-8, 7-0 (less similar), 7-3 (most dissimilar); 47 rules have been activated during this inference. Thus the regions 7 and 2 will be merged in this iteration of the S-RSST algorithm.

V. Conclusions

The methodology presented in this paper aimed in improving image segmentation based on initial labeling of regions and not only on visual features. This hybrid algorithm involves ruled-based reasoning in a region merging process, by means of calculating the semantic distance between two adjacent regions based on local knowledge. First experiments have shown the feasibility of our approach, nevertheless the current rule base must be thoroughly tested. In our future work, we will incorporate also some domain knowledge (e.g. sea cannot be above sky) to improve the whole process.

Acknowledgment

This research was supported by the European Commission under contract FP6-027026 K-SPACE. Thanos Athanasiadis is funded by the Greek Secretariat of Research and Technology (PENED Ontomedia 03 ED 475).

References

[1] P. Salembier and 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.
[2] P. Berka, V. Las, and V. Svatek, "Nest: re-engineering the compositional approach to rule-based inference," *Neural Network World*, pp. 367–379, May 2004.
[3] S. Berretti, A. D. Bimbo, and 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, Dec. 2001.
[4] Th. Athanasiadis, V. Tzouvaras, K. Petridis, F. Precioso, Y. Avrithis, and Y. Kompatsiaris, "Using a multimedia ontology infrastructure for semantic annotation of multimedia content," in *Proceedgins of 5th International Workshop on Knowledge Markup and Semantic Annotation (SemAnnot '05), Galway, Ireland*, 2005.
[5] Th. Athanasiadis, Y. Avrithis, and S. Kollias, "A semantic region growing approach in image segmentation and annotation," in *Proceedgins of 1st International Workshop on Semantic Web Annotations for Multimedia (SWAMM '06), Edinburgh, Scotland*, 2006.