A Learning Approach for Adaptive Image Segmentation

Vincent Martin, Nicolas Maillot, Monique Thonnat

A Learning Approach for Adaptive Image Segmentation

Vincent Martin, Nicolas Maillot and Monique Thonnat
INRIA Sophia Antipolis - Orion Team
2004 Route des lucioles - B.P. 93
06902 Sophia Antipolis, France
{*vincent.r.martin,nicolas.maillot,monique.thonnat*}*@sophia.inria.fr*

Abstract

As mentioned in many papers, a lot of key parameters of image segmentation algorithms are manually tuned by designers. This induces a lack of flexibility of the segmentation step in many vision systems. By a dynamic control of these parameters, results of this crucial step could be drastically improved.

We propose a scheme to automatically select segmentation algorithm and tune theirs key parameters thanks to a preliminary supervised learning stage. This paper details this learning approach which is composed by three steps: (1) optimal parameters extraction, (2) algorithm selection learning, and (3) generalization of parametrization learning.

The major contribution is twofold: segmentation is adapted to the image to segment, and in the same time, this scheme can be used as a generic framework, independant of any application domain.

keywords: *design methods for vision systems, image segmentation, learning techniques.*

1 Introduction

Image segmentation is a low-level task that consists on partitionning the image into homogeneous regions distincts from each other, according to some criteria. It is a crucial step in computer vision systems involving image processing (e.g. object recognition, content-based image retrieval) where the challenge is to perform an image segmentation with some semantic meaning.

Although promising results are presented in many papers, genericity is still not proven. In fact, many of these approaches suffer from a subjective tuning of key parameters. This problem also occurs in many vision systems where segmentation stage is narrowly tuned according to the application domain specificities by a human expert in image processing.

In order to cope with this lack of flexibilty, we propose an approach to automatic and adaptive segmentation based on learning optimal algorithm selection and key parameters tuning. We do not aim at building a new algorithm but rather add a control scheme to existing ones. The underlying idea is that we think that a segmentation process must be more directed by its goal than by the data. What are we expecting from a segmentation algorithm ? (1) It must be flexible enough to be ported from a domain to another one and (2) it must be adapted and well-tuned to the segmentation task. As Draper said in [8], we need to avoid relying on heuristically selected, domain specific features and methods, like *ad-hoc* algorithms and decision rules.

Program supervision techniques proved to be good candidates to control image processing programs [25, 5]. Such systems propose general architectures for planning, executing, evaluating and repairing image processing programs. But, as explained in [7], one negative point of these frameworks is that a lot of knowledge has to be provided in order to perform good parametrization. We aim at extending this approach by integrating learning at each step of the framework so as to have more dynamic and generic systems.

This paper is organized as follows. Section 2 gives a quick overview of key issues of existing image segmentation methods. Section 3 first presents an overview of the proposed approach then explains how knowledge on algorithm selection and parameters tuning is learned. Section 4 presents experimental results based on the proposed methodology applied to outdoor scenes. Finally, a conclusion and a discussion of future work are given in section 5.

2 Related Work

Over the last four decades, an increasing number of segmentation algorithms have been developed. In the first times, most of the efforts were devoted to build algorithms

based on low-level pixel cues as color, edges and texture. That makes them universally applicable but often leads to poor meaningful segmentation. Few of them were combined in a cooperative framework [24, 16] in order to avoid the weakness of each. However, the inability to specify how homogeneous a region should be causes the algorithm to fail. Thus, the challenge of achieving more perceptual oriented segmentation has motivated researchers to develop models for extracting, grouping and classifying more perceptual cues [18, 19, 10, 4]. Recent works [21, 2, 23, 12, 11, 3, 14] addressing these purposes apply learning techniques to capture models characteristics.

In this section, we describe three approaches devoted to produce perceptual segmentation by using various learning techniques: (1) algorithm parameters learning by synthetic object model matching, (2) object-class model learning by example, and (3) supervised parameters learning for perceptual segmentation of complex scenes.

In [21], Peng proposes a model-based multi-stage recognition system using reinforcement learning. In this paper, segmentation algorithm parameters and feature extraction algorithm parameters are trained to obtain the maximum model-matching confidence. However, the system is fully dependent on the object model (here a polygonal approximation of a sideway car) and cannot be considered in situations where objects are harder to model like natural objects from different points of view, scales, and so on.

In [3], a figure-ground learning scheme for class-based segmentation is described. It combines top-down and bottom-up segmentation processes to, respectively extract image class-relevant fragments, and thereafter to obtain more accurate object boundaries. Good results are presented for simple object-class like sideways horses or cars. The main drawback of the system is its sensitivity to regions variability. As mentioned by the authors, it relies on two main criteria. We observe that each of these criteria hides a key parameter, which is manually tuned from experience.

In [12], the authors present a method for figure-ground segmentation of objects in difficult real-world scenes (cars and cows) using a probalistic formulation to integrate learned knowledge about the recognized category with the supporting information on image. The main advantage of this work is that neither manual segmented image nor class-object models are needed during the learning process, excepted a codebook of local appearance of object category. However, the codebook grows proportionnaly to the complexity of object to extract. Even if this knowledge is easily available for cars or cows, this task is more difficult for natural objects.

In [4], Chen combines spatially adaptive texture features and local color composition features to perform robust and precise perceptual segmentation of complex scenes. As explained by the authors, several key parameters are determined by subjective preliminar tests, like threshold for smooth/nonsmooth texture classification and threshold for color composition feature similarity. Choice of these parameters could be assimilated to a manual learning stage.

Finally, we denote two main drawbacks in existing proposed methods for image segmentation learning. First, object-class model learning by examples approaches are limited in its applications: complex objects need too much knowledge to be easily modelizable (specially for the end-user). Secondly, perceptual segmentation approaches are still not able to dynamically adapt its parameters to all situations. An intermediate solution, which doesn't ask too much knowledge to the end-user (i.e. choice of segmentation algorithms and of theirs parameters) has to be found.

3 Proposed Approach

In the approach proposed in this paper, we avoid giving explicit models of the object to extract and hand-choosen parameters, because it implies too much knowledge and restricts the application domain. Because segmentation is an ill-defined problem, we argue that no generic segmentation algorithm can be found. A way to perform automatic meaningful segmentations is to be able to select best adapted and well-tuned algorithms according to a set of manually segmented examples. This scheme can be easily applied by end-users, non experts in image processing.

3.1 Overview

This approach has two main phases: a segmentation learning phase and an automatic segmentation phase.

The learning phase is subdivided into three stages (see figure 1): (1) optimal algorithm parameter extraction, (2) construction of a case base which contains processed cases. Each entry of this base is related to features describing an image with the corresponding optimal algorithm parameters and (3) algorithm selection learning.

The automatic phase uses this knowledge for automatic and adaptive segmentation (see figure 2). Features are given in input of the algorithm selection predictor trained in previous stage (1). Then, similarity is determined by looking up the case base for similar cases (2). When the closest one is found, image is segmented with corresponding optimal parameters.

3.2 Learning Phase

The goal of this stage is to extract optimal algorithm parameters (see figure 3), to build the case base and to train a predictor for selection of algorithm (see figure 5).
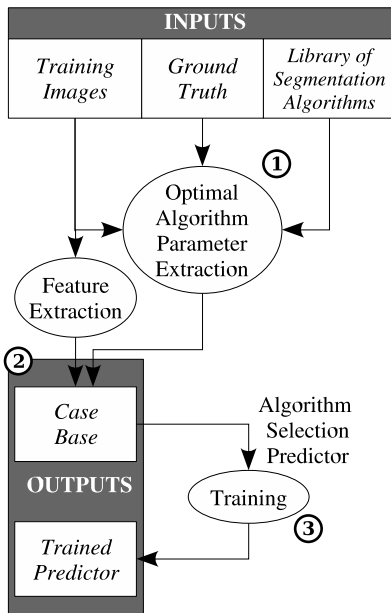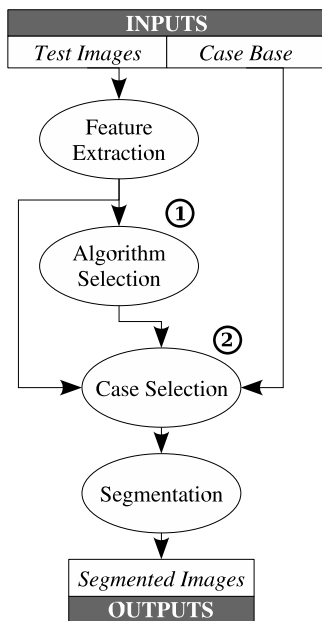
Figure 1. Overview of the learning phase

Figure 2. Overview of the automatic phase

3.2.1 Optimal Algorithm Parameter Extraction

From experience, in many segmentation algorithms, we have been able to come up with key parameters that reduce the complexity of the search space for the user and make it simple to achieve a reasonable segmentation while only modifying one or two parameters. The goal of this step is to automatically tune such key parameters for the considered images to segment. The only provided knowledge on algorithms is the key parameters and some constraints on theirs scales of values (e.g. minimum and maximum). Others paremeters are set by default.

We pose the optimal algorithm parameter extraction as an optimization procedure. The purpose of an optimization procedure is to find a set of parameter values for which an objective function gives the *best* maximum/minimum measure values. This objective function is based on a measure of goodness/discrepancy, called performance metric. A large variety of performance metrics have been proposed for evaluating segmentation results [26]. In this paper, we use a supervised evaluation method (also known as empirical discrepancy method) which requires beforehand to generate manually reference segmented images. In that way, we can directly evaluate segmentation within a perceptual ground truth and thus, optimize algorithm parameters for perceptual segmentation, as far as possible. But this job is also subjective and time-consumming, especially for complex natural images.
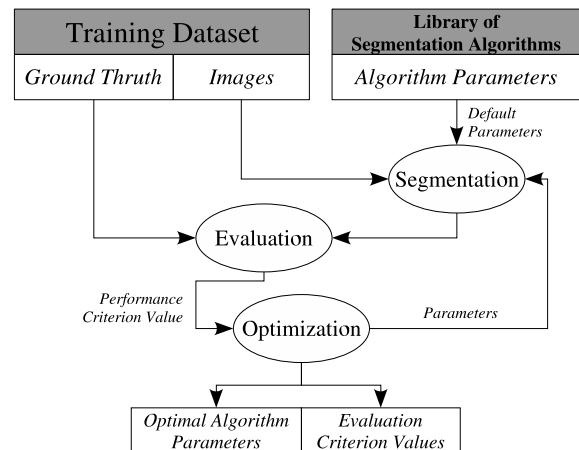

Figure 3. Optimal algorithm parameters extraction scheme. Default parameters are set according to algorithm designers recommendations.

Our performance metric is area-based. It captures deficiencies such as inaccurate boundary localization, oversegmentation, and under-segmentation. First, each region

of the segmented image is associated with a region of the reference segmented image on the basis of region overlapping. By this way, we obtain three sets of region pairs: a set of identified region pairs, a set of non-associated regions of the segmented image with a region of the reference segmented image (over-segmented regions) and a set of non-associated regions of the reference segmented image with a region of the segmented image (under-segmented regions). For the inaccurate boundary localization error measure, a weighted sum of misclassified pixels for identified region pairs is computed. Similar calculation is applied to each region pair of the two others sets. So, the final output is a weighted sum of misclassified pixels, indicating how well the segmentation masks correspond to the reference ones. The smaller output value is, the better is the segmentation quality. Note that value zero is achieved when segmentation result and reference fit exactly. More details on this evaluation metric can be found in [15].

Let i be an image of the training dataset $\mathcal{I}$, G_i be its ground truth (manual segmentation), A be a segmentation algorithm of the library of segmentation algorithms $\mathcal{A}$ and $\mathbf{p}^A$ a vector of parameters for the algorithm A. The result R_i^A of the segmentation of i with algorithm A is defined as $R_i^A = A(i, \mathbf{p}^A)$, where R is a set of regions. The goal is to obtain R_i^A as closed to G_i as possible. The performance evaluation of this result is noted $E_i^A = \rho(G_i, R_i^A)$ where ρ is the performance metric and E_i^A a scalar. The purpose of the optimization procedure is to find a set of parameter values $\mathbf{p}_i^A$ which minimizes E_i^A:

$$\mathbf{p}_i^A = \arg \min_{\mathbf{p}^A} \rho\big(G_i, A(i, \mathbf{p}^A)\big) \qquad (1)$$

Because ρ has no explicit mathematical form and is non-differentiable, standard powerful optimization techniques like Newton-based and quasi-Newton methods cannot be applied effectively. General methods suitable for such a problem are usually called *direct search* method [9]. Here, we use a modified simplex search technique[1] introduced by Nelder and Mead [17].

This optimization procedure has many advantages: first, simplex technique is appropriate for optimizing several algorithm parameters at the same time. Then, the used performance metric allows algorithms performance scores to be objectively ranked. A third aspect we have experimented is the possibility to constraint the criterion measures to be more sensitive to some regions of interest: error measures for boundary localization, under-segmentation and over-segmentation can be weighted differently to take more into account a highlighted region. By this way, parameters will be specifically optimized for a better segmentation of this region.

---

[1]P.L. Shepherd. http://www.cs.wm.edu/plshep

Some results from the optimization step are given in figure 4.
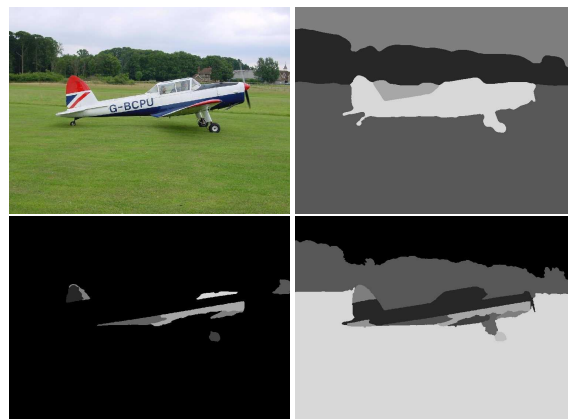

Figure 4. Example of optimal parameters extraction. From left to right and from top to bottom: input image, manual segmentation, segmentation with default parameters, segmentation after optimal parameters extraction.

This optimization is performed for each image $i \in \mathcal{I}$ and for each algorithm $A \in \mathcal{A}$. The output of this stage is a set of vectors $\mathbf{p}_i^A$ with associated E_i^A (one parameter vector per algorithm and per image).

3.2.2 Case Base Construction

The first step (1) consists on ranking optimization results from the previous stage. For each image i, according to the smallest value of E_i^A, the best algorithm is associated to the image and is denoted A_i.

In parallel (2), a vector of features $\mathbf{F}_i$ is extracted. We use color distribution descriptors (color coherence vectors [20]), texture descriptors (steerable oriented gaussian derivatives features [1]) and some global statistic descriptors (global entropy, energy and variance) to construct feature vectors.

Then, a new case $c_i^A = (\mathbf{F}_i, A_i, \mathbf{p}_i^A)$ is stored in the case base (3).

3.2.3 Algorithm Selection Learning

When all images of $\mathcal{I}$ are processed (i.e. when the case base is entirely constructed), the predictor is trained with a neural network (a multi-layer perceptron). This network takes as input a vector of features $\mathbf{F}_i$ and its output is the identifier of the associated algorithm A_i (4). The output of this stage
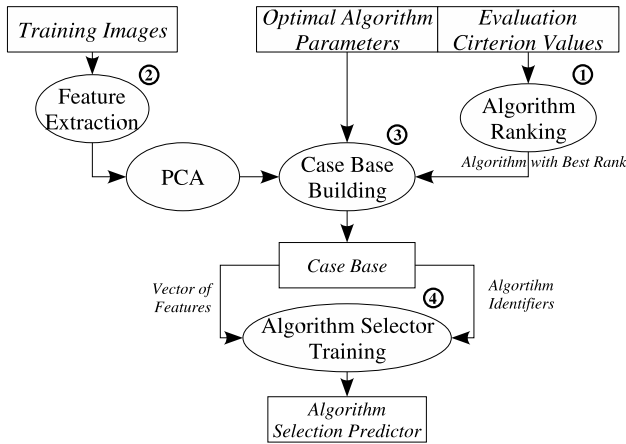
Figure 5. Learning phase scheme (case base construction and algorithm selection learning

is a predictor trained for algorithm selection and a case base for algorithm parametrization.

The main difficulty of this stage is to train the neural network with only relevant features. For this challenge, two solutions are conceivable: first, intrinsic knowledge on the segmentation algorithm enables an heuristic selection of features. For example, a threshold-based algorithm is sensitive to the gray-level value of pixels. Hence, a relevant feature will be simply an histogram. But the relationship between algorithms and features cannot be always readily established, especially for complex algorithms with many parameters. Second, we can extract a broad set of general global features and then, reduce it to a more relevant subset with a PCA. This is the solution we have adopted.

For the presented results, the dimensionality of the computed vector is 209 features. PCA reduces it to 66 features.

3.3 Automatic Segmentation Phase

The automatic phase aims at using knowledge learned from learning phase for best adapted segmentation of new images (see figure 2).

The case base can be decomposed into subsets. Let consider that s^A is the subset of cases $\{c_i^A\}_{i \in \mathcal{I}}$ where algorithm A has performed the best segmentation. For each new test image j of the test dataset $\mathcal{J}$, a feature vector $\mathbf{F}_j$ is first computed then reduced by PCA. This vector is used as input to the algorithm selection predictor. An algorithm $A_j \in \mathcal{A}$ is selected. The optimal parametrization $\mathbf{p}_j^A$ for A_j is defined as:

$$\forall c_i^A \in s^A, \quad \mathbf{p}_j^A \,|\, j = \arg\min_i dist(\mathbf{F}_i, \mathbf{F}_j) \qquad (2)$$

where $dist(\mathbf{F}_i, \mathbf{F}_j)$ is the euclidean distance between $\mathbf{F}_i$ and $\mathbf{F}_j$.

4 Experimental Results

We have experimented our approach on an image database composed of 140 samples images of aircrafts in outdoor scenes. Images are very heterogeneous: some of them have homogeneous background, others are strongly contrasted or have complex object of interest and background structures. This dataset is randomly divided into 67 training images and 73 testing images.

Currently, three candidate image segmentation algorithms compose the library: a meanshift segmentation algorithm [6], a region growing algorithm, and an inherently parallel hierarchical color segmentation algorithm [22]. The meanshift algorithm has three key parameters to tune: the maximum neighbour color distance parameter which controls the region merging, the range radius of the mean shift sphere (relative to the first parameter) and the spatial radius of the mean shift sphere which controls the smoothing of the region boundaries. The region growing key parameter is a threshold relative to the gradient image. Four seeds at each corner are also defined for the starting points. The third algorithm has also one key parameter that defines the smallest allowed euclidian distance between two similar rgb color vectors.

Parameters of the neural network are: two layers with a sigmoid activation function, 66 hidden units (number of features), conjuguate gradients training method and maximal number of epochs of 800 (number of presentations of the entire training set).

Table 1 presents first results of the automatic phase. It can be seen, for the presented examples, that the system achieves good algorithm selection. For the first image, the system has selected the color segmentation algorithm. The segmentation is quite good since the different perceptual regions related to the sky, the plane and the ground are well separated. In contrary, region growing and meanshift algorithms have merged too many regions. For the second image, the homogeneous background has inducted the system to select the region growing algorithm which is visually the best choice. Meanshift and color segmentation have produced very close results for the third and the fourth image.

5 Conclusion

This paper presents a method for learning how to perform adaptive image segmentation. This learning approach is structured in three main stages: a parameter optimization stage, a construction of a case base stage and an algorithm selection learning stage. The first stage consists of
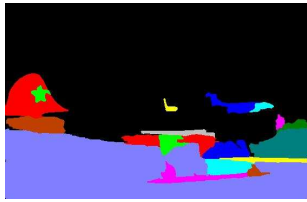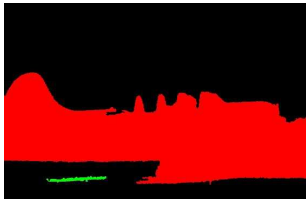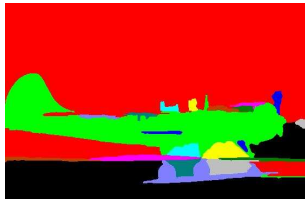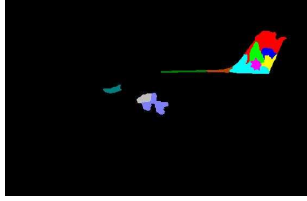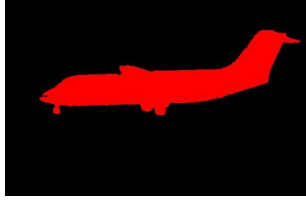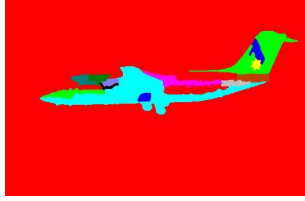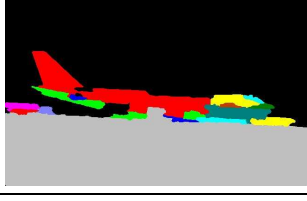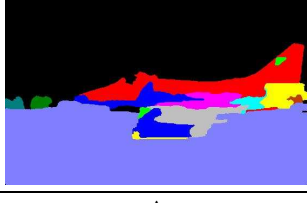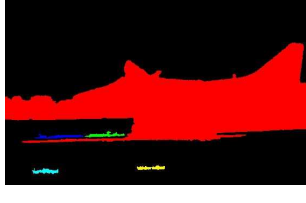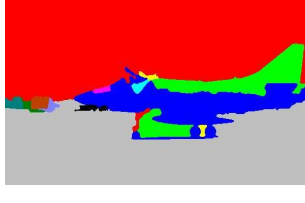
original images	meanshift [6]	region growing	color segmentation [22]

Table 1. Segmentation results in automatic phase. For each test image, results of the three algorithms are presented. Surrounded images correspond to the selected algorithm by the system.

extracting optimal algorithm parameters for a training image dataset. We pose this optimal parameters extraction as an optimization problem. A performance metric based on region segmentation accuracy criterions is used for evaluating segmentation result within ground truth (manual segmentation). This performance metric is considered as an objective function to be minimized. Nelder-Mead Simplex method is then used to solve the optimization problem. The result is a set of optimal parameters and an objective evaluation value for each algorithm and for each image of the training dataset. The second stage consists of learning the optimal algorithm selection and then of case base construction. From the results of the previous stage, algorithms are ranked and relevant features are extracted. For each training image, a new case, composed of a vector of image features, choosen algorithm and optimal parametrization of this algorithm is stored in the case base. The third stage make use of this stored knowledge to train a MLP neural network for algorithm selection.

Final segmentation performance is limited by algorithm individual performance and by the size of the learning dataset. In order to fully validate our approach, we have to test it and to evaluate it on large image databases from various domains and to expand the algorithm library. This is our current work.

The main drawback of our approach is the difficulty to draw manual segmentation of complex images during the learning phase. Human manual segmentation cannot be reproduced exactly by a segmentation algorithm. In order to fill this gap, we have to reduce the weight of manual segmentation in the learning process. More simplified ground truths (like figure ground segmentation) coupled with multiscales approach could be used. We also want to use more *a priori* knowledge on the objects to extract. It can be knowledge for optimal merging of region of interests. Another way is to guide the segmentation task according to a visual concept based description [13].

6 Acknowledgements

References

[1] P. Alvarado, P. Doerfler, and J. Wickel. Axon2 - a visual object recognition system for non-rigid objects. In *SPPRA*, July 2001.

[2] B. Bhanu and S. Fonder. Learning based interactive image segmentation. In *International Conference on Pattern Recognition (ICPR'00)*, volume 1, pages 1299–1302, Barcelona, Spain, 2000. IEEE.

[3] E. Borenstein and S. Ullman. Learning to segment. In LNCS, editor, *ECCV'04*, volume 30, pages 315–328, 2004.

[4] J. Chen, T. Pappas, N. Mojsilivic, and B. E. Rogowitz. Adaptive perceptual color-texture image segmentation. *Transactions on Image Processing*, 2004.

[5] R. Clouard, A. Elmoataz, C. Porquet, and M. Revenu. Borg : A knowledge-based system for automatic generation of image processing programs. In *PAMI*, volume 21, pages 128–144, France, February 1999. IEEE.

[6] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *PAMI*, 24(5):603–619, 2002.

[7] B. Draper, A. Hanson, and E. Riseman. Knowledge-directed vision: Control, learning, and integration. In IEEE, editor, *Signals and Symbols*, volume 84, pages 1625–1637, 1996.

[8] B. A. Draper. From knowledge bases to markov models to pca. VSCA03, 2003.

[9] P. Gill, W. Murray, and M. Wright. *Practical Optimization*, chapter 4. Academic Press, London and New York, 1993.

[10] A. Hoogs, R. Collins, R. Kaucic, and J. Mundy. A common set of perceptual observables for grouping, figure-ground discrimination, and texture classification. *PAMI*, 25(4):458–474, 2003.

[11] J. Kaufhold and A. Hoogs. Learning to segment images using region-based perceptual features. In IEEE, editor, *CVPR'04*, 2004.

[12] B. Leibe and B. Schiele. Interleaved object categorization and segmentation. In *BMVC'03*, NORWICH, UK, Sept 2003.

[13] N. Maillot, M. Thonnat, and A. Boucher. Towards ontology-based cognitive vision. *MVA*, 16:33–40, 2004.

[14] D. R. Martin, C. C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 26(1), January 2004.

[15] V. Mezaris, I. Kompatsiaris, and M. Strintzis. Still image objective segmentation evaluation using ground truth. In *Fifth COST 276 Workshop on Information and Knowledge Management for Integrated Media Communication*, pages 9–14, Prague, Czech Republic, October 2003.

[16] X. Munoz, J. Freixenet, X. Cufi, and J. Marti. Strategies for image segmentation combining region and boundary information. *Pattern Recognition Letters*, 24:375–392, June 2003.

[17] J. Nelder and R. Mead. A simplex method for function minimization. *Computer Journal*, 15:1162–1173, 1965.

[18] W. Osberger and A. J. Maeder. Automatic identification of perceptually important regions in an image. In *ICPR'98*, volume 1, pages 701–104, 1998.

[19] A. Pardo. Extraction of semantic objects from still images. In *ICIP'02*, New York, USA, 2002.

[20] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *ACM Multimedia*, pages 65–73, 1996.

[21] J. Peng and B. Bahnu. Delayed reinforcement learning for adaptive image segmentation and feature extraction. *Systems, Man, and Cybernetics*, 28(3), 1998.

[22] V. Priese, Lutz Rehrmann and P. Sturm. Color structure code, 2002.

[23] X. Ren and J. Malik. Learning a classification model for segmentation. In IEEE, editor, *ICCV'03*, volume 1, pages 10–18. Nice, France, 2003.

[24] E. Saber and A. M. Tekalp. Integration of color, edge, shape, and texture features for automatic region-based image annotation and retrieval. *Journal of Electronic Imaging*, 7(3):684–700, 1998.

[25] M. Thonnat, S. Moisan, and M. Crubézy. Experience in integrating image processing programs. In H. Christensen, editor, *1st International Conference on Vision Systems. Lecture Notes in Computer Science*, Gran Canaria, Spain, 1998. Springer-Verlag.

[26] Y. Zhang. A survey on evaluation methods for image segmentation. *Pattern Recognition*, 29(8):1335–1346, 1996.