Stopping Region-based Image Segmentation at Meaningful Partitions

Tomasz Adamek and Noel E. O'Connor

Centre for Digital Video Processing, Dublin City University, Dublin 9, IRELAND
{adamekt,oconnorn}@eeng.dcu.ie

Abstract. This paper proposes a new stopping criterion for automatic image segmentation based on region merging. The criterion is dependent on image content itself and when combined with the recently proposed approaches to *syntactic* segmentation can produce results aligned with the most salient semantic regions/objects present in the scene across heterogeneous image collections. The method identifies a single iteration from the merging process as the stopping point, based on the evolution of an accumulated merging cost during the complete merging process. The approach is compared to three commonly used stopping criteria: (i) required number of regions, (ii) value of the least link cost, and (iii) Peak Signal to Noise Ratio (PSNR). For comparison, the stopping criterion is also evaluated for a segmentation approach that does not use syntactic extensions. All experiments use a manually generated segmentation ground truth and spatial accuracy measures. Results show that the proposed stopping criterion improves segmentation performance towards reflecting real-world scene content when integrated into a syntactic segmentation framework.

1 Introduction

The problem of partitioning an image into a set of homogenous regions or semantic entities is a fundamental enabling technology for understanding scene structure and identifying relevant objects. A large number of approaches to image segmentation have been proposed in the past [1]. This paper focuses on automatic *visual feature-based* segmentation of images that does not require developing models of individual objects, thereby making the approach generic and broadly applicable.

A popular class of automatic segmentation techniques are the *Graph Theoretic* approaches in which images are represented as weighted graphs, where nodes correspond to pixels or regions and the edges' weights encode the information about the segmentation, such as pairwise homogeneity or edge strength [2,3,4,5,6,7]. The segmentation is obtained by partitioning (cutting) the graph so that an appropriate criterion is optimized. Several approaches from this category utilize *Region Adjacency Graphs* (RAGs) where nodes represent regions while edges contain region pairwise dissimilarities [2,8,9]. The pairwise dissimilarities are typically computed using a distance between regions' colour features. Such dissimilarities are often referred to as *region homogeneity criteria*. RAGs can be simplified by successive mergings of neighboring regions. A fast and yet effective (and therefore popular) approach from this category is the *Recursive Shortest Spanning Tree* (RSST) algorithm [2,3,9]. Another example of an approach from this category is the *Normalized Cut* [5] approach that has attracted considerable attention within the content-based information retrieval (CBIR) community in recent years due to its state of the art performance [10].

Unfortunately however, since such approaches as based purely on low-level image features, the resulting segmentation results do not necessarily reflect the real-world content of the scene, thereby limiting their usefulness both for feature extraction for CBIR or as a pre-processing step prior to automatic or semi-automatic semantic object segmentation. This is manifested by either under-segmentation, where semantic objects or regions[1] are merged in the final segmentation mask, or over-segmentation, where single semantic objects or regions are composed of many small irrelevant regions. Thus, in an effort to make region-based segmentation more useful in applications that require a semantic interpretation of the scene, recently several researchers have proposed to improve the nature and quality of the segmentation produced by region merging by utilizing additional cues. These additional cues, whilst not semantic in themselves, reflect the semantic nature of the scene by encapsulating its geometrical and spatial configuration properties. In this way, the resulting approach is still generic but potentially provides a much more stable basis upon which to perform semantic knowledge extraction. *Ferran and Cassas* used the term *syntactic features* to refer to such quasi-semantic features [11]. In practice, integrating such features is achieved by using the geometrical properties and the spatial configuration of regions as merging criteria in one of the Graph Theoretic approaches referred to above [12,13,11,14,7]. In fact, using such a merging framework, some approaches have attempted to divide the merging process into stages and use different homogeneity criterion in each stage [11,14]. Whilst these approaches are promising, they are still hampered by the key difficulty of knowing when to stop the merging process in order to obtain the best possible segmentation result, particularly in the case of hierarchical segmentation.

The region merging (or splitting) approaches referred to above are particularly attractive in constructing hierarchical representations of images, e.g. all merges (or cuts) performed during the region merging (splitting) process can be stored in a *Binary Partition Tree* (BPT). The usefulness of such hierarchical representations, and BPT in particular, has been advocated by many researchers as an important pre-processing step in applications such as region-based compression, region-based feature extraction in the context of MPEG-7 description of content, and semi-automatic segmentation [4,15,7]. However, at the moment

[1] Note that we use the term 'semantic regions' to refer to image regions that would be defined by a human annotator of the scene. Whilst they may not constitute full semantic objects as required by any given application/user they are parts thereof that reflect real physical structure of an object.

many applications, including CBIR systems, can utilize only a single partition of the scene. Moreover, even in scenarios where the segmentation is used to produce a hierarchical representation of the image it is often necessary, e.g. due to reasons of efficiency, to identify a single partition within such a representation most likely to contain a meaningful segmentation (or the most representative impression of the scene) – see for example [15].

In the case of segmentation via region merging, a single partition is obtained simply by defining a criterion for stopping the merging process. In other words, the stopping criterion is needed to identify those elements within the hierarchical structure which are most likely to be relevant in a given application. To date, only very simple stopping criteria have been used, e.g. the required number of regions [3] or the minimum value of *Peak Signal to Noise Ratio* (PSNR) between the original and the segmented image reconstructed using mean region colour [15]. Although intuitive, in cases of heterogeneous image collections such criteria often fail to produce partitions containing the most salient objects present in the scene.

This paper proposes a new stopping criteria for a syntactic region-based segmentation approach that facilitates the generation of single partitions that contain the most salient objects present in the scene (or partitions corresponding to the most representative impression of the scene) and that works across heterogeneous image collections. The method identifies a single iteration from the merging process corresponding to the most salient partition based on the evolution of the accumulated merging cost during the overall merging process. The proposed approach is compared to three different commonly used stopping criteria: (i) required number of regions, (ii) value of the least link cost, and (iii) *Peak Signal to Noise Ratio* (PSRN) between the original and segmented image (reconstructed using mean region colour) [15]. The different stopping criteria are evaluated within a merging framework that includes two different region homogeneity criteria corresponding to the commonly used RSST [3] that uses only low-level features, and a recently proposed improvement of this [7] that uses syntactic features.

The remainder of this paper is organized as follows: the next section presents an overview of the region merging framework used. Then, section 3 discusses three different commonly used stopping criteria and section 4 introduces our new approach. The results of exhaustive evaluation with two different merging criteria and an image collection with ground-truth segmentation of semantic regions are presented and discussed in section 5. Section 6 concludes the paper.

2 Region Merging Framework

The proposed stopping criteria are evaluated using two region merging approaches. The first is the commonly used *Recursive Shortest Spanning Tree* (RSST) algorithm [2,3,9]. The second one is a new extension to the original RSST, proposed recently in [7], that uses so-called syntactic features. However, theoretically the criterion could be integrated into any approach based on region merging or splitting.

2.1 Original RSST

The original RSST algorithm starts by mapping the input image into a weighted graph [2], where the regions (initially pixels) form the nodes of the graph and the links between neighboring regions represent the merging cost, computed according a selected homogeneity criterion. Merging is performed iteratively. At each iteration two regions connected by the least cost link are merged. Merging two regions involves creating a joint representation for the new region (typically its colour is represented by average colour of all its pixels [3]) and updating its links with its neighbors. The process continues until a certain stopping condition is fulfilled, e.g. the desired number of regions is reached or the value of the least link cost exceeds a predefined threshold.

The merging order is exclusively controlled by the function used to compute the merging cost. Let r_i and r_j be two neighboring regions. The merging cost is based solely on a simple colour homogeneity criterion defined as [16,17]:

$$C_{orig}(i,j) = \|\mathbf{c}_i - \mathbf{c}_j\|_2^2 \cdot \frac{1}{a_{img}} \cdot \frac{a_i a_j}{a_i + a_j} \tag{1}$$

where $\mathbf{c}_i$ and $\mathbf{c}_j$ are the mean colours of r_i and r_j respectively, a_i and a_j denote region sizes, and $\|\cdot\|_2$ denotes the $\mathcal{L}_2$ norm. a_{img} is the size of the entire image. Such normalization by a_{img} does not affect the merging order allowing to use the value of C_{orig} as a stopping criterion with collections containing images of different sizes. Alternative merging criteria based on colour can be found in [2,15,11].

2.2 Enhanced RSST

In the region merging approach proposed in [7] additional evidence for merging is provided by the syntactic visual features advocated by Ferran and Casas [11], representing geometric properties of regions and their spatial configuration, e.g. homogeneity, compactness, regularity, inclusion or symmetry. It has been shown that these features can be used in bottom-up segmentation approaches as a way of partitioning images into more meaningful entities without assuming any application dependent semantic models.

The segmentation process is divided into two stages. The initial partition is obtained by the RSST algorithm with the original region homogeneity criterion implemented as in [3] since it is capable of producing good results when regions are uniform and small. This stage ensures the low computational cost of the overall algorithm and also avoids analysis of geometrical properties of small regions with meaningless shape. This initial stage is forced to stop when a predefined number of regions is reached (100 for all experiments presented in this paper). Then, in the second stage, homogeneity criteria are re-defined based on colour and syntactic visual features and the merging process continues until a certain stopping criterion is fulfilled. Regions' colour is represented using a fine and compact representation motivated by the *Adaptive Distribution of Colour Shades* (ADCS) [18] whereby each region contains a list of pairs of colour/population (where population refers to the ratio between the number of pixels with this colour and the total size of the region) that represents its complex colour variations more precisely than the mean value. Two region geometric properties, adjacency and changes in global shape complexity, are included as syntactic features. The new merging order is based on evidence provided by different features (colour and geometric properties) fused using an integration framework based on *Dempster-Shafer* (DS) theory [19] which takes into account the reliability of different sources of information as well as the fact that certain measurements may not be precise (doubtful) or even "unknown" in some cases. Full details can be found in [7].

3 Existing Stopping Criteria

This section describes and briefly discusses three simple but commonly used stopping criteria that are used to evaluate our approach. They are: (i) required number of regions, (ii) maximum value of the least link cost, and (iii) minimum value of *Peak Signal to Noise Ratio* (PSRN) between the original and segmented image (reconstructed using mean region colour) [15].

3.1 Number Of Regions

In this very simple method, the merging process continues until a desired number of regions is reached indepenent of the image content. Clearly, application of this criterion is limited to cases where the number of required regions/object is known.

3.2 Merging Cost

In this case, the merging process is stopped when the merging cost exceeds a predefined threshold. The threshold is typically chosen in an ad-hoc manner to suit a particular application. Although the method is intuitive, one should note that for heterogeneous collections it is often impossible to chose a singe threshold resulting in segmentation corresponding to the most salient regions present in the scene. Looking at Figure 1, for example, a threshold high enough to segment the person's torso in Figure 1(c)) as one region would result in the "jet" being merged into the backgroundd in (Figure 1(a)).

3.3 PSNR Value

In this case, a single partitioning of the image is achieved based on value of PSNR between the original and segmented image (reconstructed using mean region colour) – see for example [15]. Typically, all merges performed during the region merging process are stored in a *Binary Partition Tree* (BPT) together with values of PSNR at each merging iteration. A single partition is then obtained from the BPT by deactivating nodes following the merging sequence until the value of PSNR falls below a pre-defined threshold. In our work, only the luminance information is used in order to limit the computational complexity. Formally PSNR for an $M \times N$ image is computed as:

$$PSNR = 20 \log_{10} \frac{MAX_L}{\sqrt{\frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} \Big(L(i,j) - K(i,j) \Big)^2}} \tag{2}$$

where L is the original image and K is the segmented image reconstructed using mean region intensities and MAX_L is the maximum luminance value. Again, it may be impossible to chose a singe threshold resulting in segmentation corresponding to the most salient regions present in the scene for heterogeneous collections of images.

4 Proposed Stopping Criterion

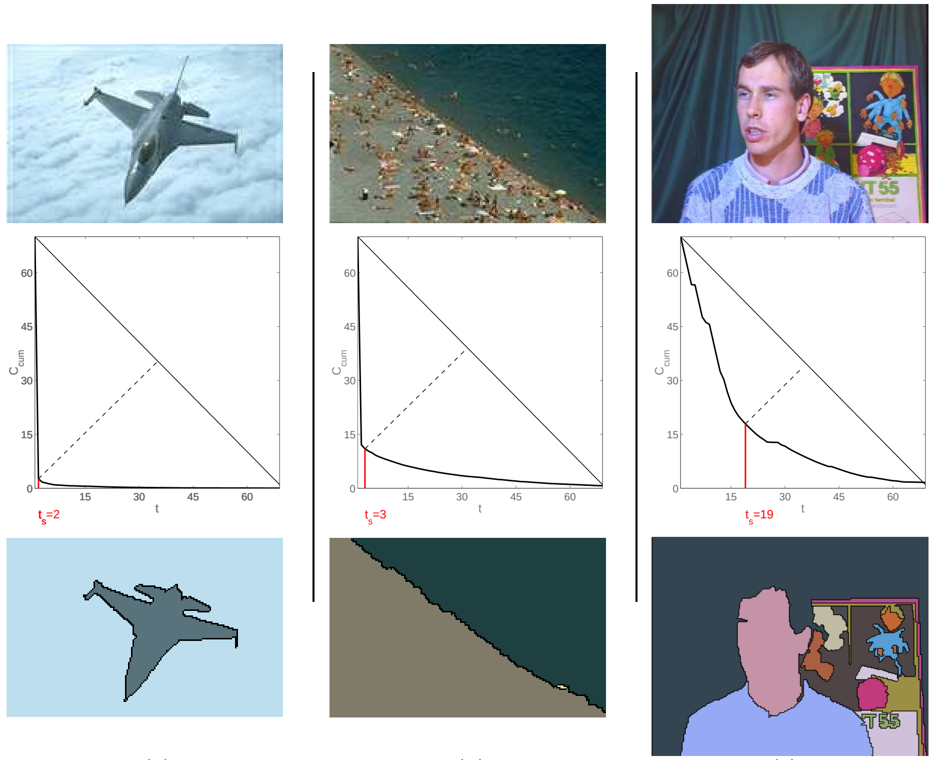
This section discusses a new stopping criteria to obtaining a single partition that reflects meaningful image content. The approach is not based on a pre-defined threshold, but rather evaluates the 'goodness' of segmentation at a single iteration with a final decision made based on the evolution of the merging cost accumulated during the overall merging process. As in the PSNR-based approach, the selected partition is obtained by deactivating nodes from the BPT built during the merging process.

First, let us define an accumulated merging cost measure C_{cum} which measures the total cost of all mergings performed to produce t regions as:

$$C_{cum}(t) = \begin{cases} \sum_{n=t}^{N_I - 1} C_{mrg}(n) & \text{if} \quad 1 \le t < N_I \\ 0.0 & \text{otherwise} \end{cases} \tag{3}$$

where $C_{mrg}(n)$ denotes cost of merging of a single pair of regions reducing the number of regions from $n+1$ to n computed using a given merging criterion. N_I denotes the number of regions in the initial partition, i.e. produced in the first merging stage. In the case of the enhanced RSST, only the costs of merging performed during the second stage contribute to the value of C_{cum} (i.e. $N_I = 100$). Also, N_I was set to 100 in the case of the original RSST. It should be stressed that the value of N_I has a minimal effect on the final result since C_{cum} changes very slowly for the initial merges.

Figure 1 shows measure C_{cum} plotted for each iteration t of the merging process for three images, each presenting a different segmentation challenge. The basic idea behind the stopping criterion is to find the number of regions t_s which partitions the curve $C_{cum}(t)$ into two segments in such a way so that with decreasing values of t, values of $C_{cum}(t)$ within segment $[1, t_s]$ increase significantly faster than for the segment $[t_s, N_I]$.



(a) (b) (c)

Fig. 1. Stopping criterion based on accumulated merging cost (C_{cum}) for $T_{cum} = 70$. The merging order is computed using the extended colour representation together with syntactic visual features. Values of C_{cum} are re-scaled to the range $[0, T_{cum}]$ for visualization purposes.

In the proposed approach t_s is found using the method proposed in [20] for bi-level thresholding designed to cope with uni-modal distributions (histogram based thresholding). The algorithm is based on the assumption that the main peak of the uni-modal distribution has a detectable corner at its base which corresponds to a suitable threshold point. The approach has been found suitable for various problems requiring thresholding such as edge detection, image difference, optic flow, texture difference images, polygonal approximations of curves and even parameter selection for the split and merge image segmentation algorithm itself [21].

Here, the approach is used to determine the stopping criterion t_s based on the accumulated merging cost measure C_{cum}. Let us assume a hypothetical reference accumulated merging cost measure C_{ref} having a form of a line passing through points $(1, C_{cum}(1))$ and $(T_{cum}, C_{cum}(T_{cum}))$, where T_{cum} is a parameter whose role is explained later. The threshold point t_s is selected to maximize the perpendicular distance between the reference line and the point $(t_s, C_{cum}(t_s))$. It can be shown that this step is equivalent to an application of a single step of the standard recursive subdivision method for determining the polygonal approximation of a curve [22].

Parameter T_{cum} can be used to bias the stopping criterion towards under- or over-segmentation, i.e. smaller values of T_{cum} bias the stopping criterion towards under-segmentation while larger values tend to lead to over-segmentation. However, it should be stressed that it is the shape of $C_{cum}(t)$ which plays the dominant role in selecting t_s.

5 Experiments

5.1 Image Collection

The collection used consists of 100 images from the Corel gallery and 20 images from various sources such as keyframes from well known MPEG-4 test sequences and a private collection of photos. Ground-truth segmentation masks of semantic regions in the scene were created manually using a tool developed specifically for this task [7]. Using this tool creation of a new reference region involves the annotator choosing a label for the new region, drawing its boundary on the original image and then filling its interior. To ensure high accuracy of the masks, the area around the cursor is automatically zoomed. Furthermore, the mask can be edited in a *Boundary Mode* in which only region boundaries superimposed on the original image are displayed allowing accurate localization of the borders on the original image and fine corrections. The typical time required to manually partition an image was between 5-10 minutes, depending on user drawing skills and complexity of the scene. Although this time seems acceptable, in fact manual segmentation of several images proved extremely tiresome.

Taking into account the relatively small number of images available the proposed criteria are evaluated by two-fold cross validation. The dataset is divided into two subsets, each containing 60 images, i.e. 50 images from the Corel dataset and 10 from the other sources. Each time one of them is used for parameter tuning, the other subset is used as the test set. The final result is computed as the average error from the two trials. All images showing selected segmentation results are generated during the cross-validation.

5.2 Evaluation Criterion

The evaluation measure proposed in [23] assesses the quality of partitions in terms of spatial accuracy error. The method allows evaluation of segmentation in cases where both the evaluated segmentation and the ground-truth mask might contain several regions. Special consideration is devoted not only to the accuracy of boundary localization but also to over- and under-segmentation. To tackle such issues, the evaluation starts by establishing exclusive correspondence between regions in the reference and evaluated masks. Then, three different types of errors

Table 1. Results of cross-validation

STOPPING CRITE-RION	Original colour homogeneity criterion [3]	Homogeneity based on ADCS and the Syntactic Visual Features [7]
"Manual"	0.99	0.50
Number of regions	0.93	0.72
Merging cost	0.77	0.69
PSNR	0.84	0.82
Accum. merging cost	0.81	**0.63**

are taken into account: (i) accuracy errors for the associated pairs of regions from both reference and evaluated masks, (ii) errors due to under-segmentation computed based on non-associated regions from the reference mask, and finally (iii) errors due to over-segmentation computed from non-associated regions from the evaluated mask. Although the method is quite simple, a set of convincing evaluation results was provided in [23] indicating that the method correlates well with subjective evaluation. In this paper, to enable comparison of results for images with different sizes, the segmentation accuracy measure is normalized by the size of the image [7].

5.3 Results

All parameters (thresholds) were found during the training phase by an optimization process. Table 1 shows the average spatial accuracy error for all evaluated stopping criteria. Additionally, the first column contains results of a "manual" criterion where the merging process is stopped based on the known number of regions in the ground-truth.

The first observation to be made is that utilization of the merging cost in the enhanced RSST always results in a lower value of the average spatial error, irrespective of the stopping criterion used. Secondly, stopping criteria based on the merging cost usually perform better that the PSNR stopping criterion. Finally, the proposed stopping criterion significantly outperforms all other criteria when used with the merging cost based on ADCS and the syntactic visual features. It should be noted that this combination of merging cost together with the new stopping criterion leads to significantly lower average spatial segmentation error than the original RSST approach even with the "manual" stopping criterion based on the number of regions in the ground-truth mask.

Figure 2 shows selected segmentation results obtained by the original RSST algorithm with the PSNR based stopping criterion and segmentations produced by the merging criterion integrating the extended colour representation combined with the syntactic visual features, together with the new stopping criterion. The first five rows show examples where the new approach improves the results compared to the original RSST and the last five rows show examples where the proposed approach obtained higher spatial segmentation error than the original

RSST approach. However, it should be stressed that the latter cases are very rare (31 from 120 images) and in fact even in such cases the partitions appear somewhat intuitive.

For comparison, the last column from Figure 2 shows the results of the Blobworld segmentation algorithm[2]. The Blobworld algorithm has been extensively tested and used as a segmentation processing stage in the literature [24].

6 Conclusions

This paper presented an automatic stopping criterion for image segmentation algorithms based on iterative region merging. The approach is particularly applicable to segmentation approaches that integrate syntactic visual features. Whilst such approaches potentially offer the possibility of segmentation results that reflect real-world structure, they still need an appropriate stopping criterion to determine when such a result has been reached. Unlike other approaches, the proposed approach is not based on a single pre-defined threshold, but rather evaluates the 'goodness' of a segmentation result at a single iteration and makes a final decision based on the merging cost accumulated during the overall process. This makes the overall segmentation approach broadly applicable and experimental results show strong performance against a manually generated ground truth for a heterogeneous image collection.

Although, in many cases the proposed stopping criterion does not produce a perfect segmentation, i.e. ideally aligned with the manual segmentation of semantic regions, it successfully identifies the most visually salient regions in almost all evaluated images. This presents a tremendous opportunity for utilizing such salient regions in CBIR systems. It should be stressed that the method performs well on images presenting very different challenges with fixed value of the parameter T_{cum} and unlike other stopping criteria, e.g. based on the value of merging cost or PSNR, very erroneous segmentations are extremely rare.

Acknowledgments

The research leading to this paper was supported by the European Commission under contract FP6- 027026 (K-Space) and by Science Foundation Ireland under grant 03/IN.3/I361.

References

1. H.D. Cheng, X. H. Jiang, Y. Sun, and Jing Li Wang, "Color image segmentation: Advances & prospects," *Pattern Recognition*, vol. 34, no. 12, pp. 2259–2281, 2001.
2. O.J. Morris, M.J. Lee, and A.G. Constantinides, "Graph theory for image analysis: an approach based on the shortest spanning tree," in *IEE Proceedings*, Apr. 1986, vol. 133, pp. 146–152.

[2] Source code obtained from http://elib.cs.berkeley.edu/src/blobworld/

3. A.A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services - the European COST 211 Framework," *IEEE Trans. CSVT*, vol. 8, no. 7, pp. 802–813, Nov. 1998.
4. P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Trans. on Image Processing*, vol. 9, no. 4, pp. 561–576, Apr. 2000.
5. J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
6. R. Nock and F. Nielsen, "Statistical region merging," *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 26, no. 11, pp. 1452–1458, Nov. 2004.
7. T. Adamek, *Using Contour Information and Segmentation for Object Registration, Modeling and Retrieval*, Ph.D. thesis, School of Electronic Engineering, Dublin City University, June 2006.
8. P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for filtering, segmentation, and information retrieval," in *Proc. IEEE Int'l Conf. on Image Processing (ICIP'98), Chicago (IL), USA*, Oct. 1998.
9. Sai Ho Kwok, A. G. Constantinides, and Wan-Chi Siu, "An efficient recursive shortest spanning tree algorithm using linking properties," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 852–863, June 2004.
10. Kobus Barnard, Pinar Duygulu, Raghavendra Guru, Prasad Gabbur, and David Forsyth, "The effects of segmentation and feature choice in a translation model of object recognition," in *Proc. IEEE Conf. On Computer Vision and Pattern Recognition (CVPR'03)*, 2003.
11. C. Ferran Bennstrom and J. R. Casas, "Binary-partition-tree creation using a quasi-inclusion criterion," in *Proc. 8th Int'l Conf. on Information Visualization (IV'04), London, UK*, 2004.
12. P. Vasseur, C. Pégard, E. M. Mouaddib, and L. Delahoche, "Perceptual organization approach based on dempster-shafer theory," *Pattern Recognition*, vol. 32, no. 8, pp. 1449–1462, 1999.
13. N. Zlatoff, B. Tellez, and A. Baskurt, "Region-based perceptual grouping: a cooperative approach based on dempster-shafer theory," in *Proc. of the SPIE, Volume 6064*, 2006, pp. 244–254.
14. Tomasz Adamek, Noel E. O'Connor, and Noel Murphy, "Region-based segmentation of images using syntactic visual features," in *Proc. 6th Int'l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'05), Montreux, Switzerland*, Apr. 2005.
15. O. Salerno, M. Pardas, V. Vilaplana, and F. Marqués, "Object recognition based on binary partition trees," in *Proc. Int'l Conf. on Image Processing (ICIP '04)*, Oct. 2004, vol. 2, pp. 929–932.
16. J. H. Ward, "Hierarchical grouping to optimize an objective function," *American Stat. Assoc.*, vol. 58, pp. 236–245, 1963.
17. Saman Cooray, Noel Edward O'Connor, Sean Marlow, Noel Murphy, and Thomas Curran, "Semi-automatic video object segmentation using recursive shortest spanning tree and binary partition tree," in *Proc. 3rd Int'l Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS'01), Tampere, Finland*, 2001.
18. J. Fauqueur and N. Boujemaa, "Region-based image retrieval: Fast coarse segmentation and fine color description," *Journal of Visual Languages and Computing, special issue on Visual Information Systems*, vol. 15, pp. 69–95, 2004.
19. P. Smets, E.H. Mamdami, D. Dubois, and H. Prade, *Non-Standard Logics for Automated Reasoning*, ISBN 0126495203. Academic Press, Harcourt Brace Jovanovich Publisher, 1988.

20. P.L. Rosin, “Unimodal thresholding,” *Pattern Recognition*, vol. 34, no. 11, pp. 2083–2096, Nov. 2001.
21. S. Horowitz and T. Pavlidis, “Picture segmentation by a tree traversal algorithm,” *J. Assoc. Compt. Math.*, vol. 23, no. 2, pp. 368–388, 1976.
22. U. Ramer, “An iterative procedure for the polygonal approximation of plane curves,” *Computer, Graphics and Image Processing*, vol. 1, pp. 244–256, 1972.
23. V. Mezaris, I. Kompatsiaris, and M.G. Strintzis, “Still image objective segmentation evaluation using ground truth,” in *Proc. 5th COST 276 Workshop (2003), Berlin*, 2003, pp. 9–14.
24. Chad Carson, Serge Belongie, Hayit Greenspan, and Jitendra Malik, “Blobworld: Color- and texture-based image segmentation using EM and its application to image querying and classification,” *IEEE Trans. Pattern Anal. and Machine Intell.*, vol. 24, no. 8, pp. 1026–1037, Aug. 2002.

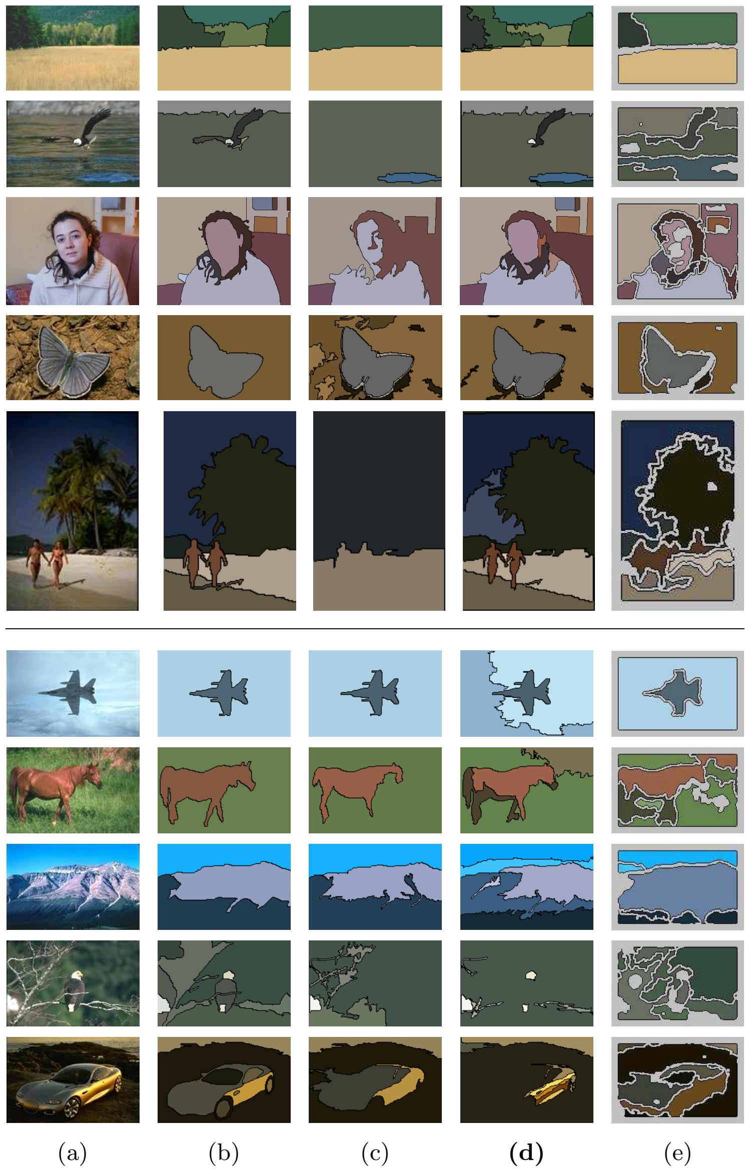


Fig. 2. Selected segmentation results. Column: (a) Original image; (b) Reference mask created by manual annotation; (c) Original RSST [3] with PSNR Stoping Criterion; **(d) RSST extended with syntactic features [7] with the proposed stopping criterion**; (e) Blobworld.