REGION-BASED SEGMENTATION OF IMAGES USING SYNTACTIC VISUAL FEATURES

Tomasz Adamek, Noel E. O'Connor, Noel Murphy

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

ABSTRACT

This paper presents a robust and efficient method for segmentation of images into large regions that reflect the real world objects present in the scene. We propose an extension to the well known Recursive Shortest Spanning Tree (RSST) algorithm based on a new color model and so-called syntactic features [1]. We introduce practical solutions, integrated within the RSST framework, to structure analysis based on the shape and spatial configuration of image regions. We demonstrate that syntactic features provide a reliable basis for region merging criteria which prevent formation of regions spanning more than one semantic object, thereby significantly improving the perceptual quality of the output segmentation. Experiments indicate that the proposed features are generic in nature and allow satisfactory segmentation of real world images from various sources without adjustment to algorithm parameters.

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. Unfortunately, segmentation remains to a large extent an unsolved problem, despite being the subject of study for several decades. Although, it is well recognized that in certain scenarios utilizing segmentation can be very successful, many researchers consider reliable unsupervised general-purpose image segmentation as a hopeless goal. We share the opinion presented in [2], that segmentation, while imperfect, is an essential first step for scene understanding and that a combined architecture for segmentation and recognition is needed. For example, sensible integration of segmentation features in an image indexing context should lead to more efficient object-based indexing solutions [3].

A large number of approaches to image or video segmentation has been proposed in the literature [4, 2, 5, 6]. They can be broadly classified as either *Top-Down* (model-based) or *Bottom-Up* (visual feature-based) approaches. Many approaches aim to create large regions using simple homogeneity criteria based only on color, texture or motion. However, applications for such approaches are limited as they often fail to create meaningful partitions due to either the complexity of the scene or difficult lighting conditions.

In [1], Ferran and Casas present a new source of potentially important information to a wide range of segmentation applications which they term *Syntactic Features*. Syntactic features represent geometric properties of regions and their spatial configurations. Examples of such features include homogeneity, compactness, regularity, inclusion or symmetry. They advocate the use of the above features in bottom-up approaches as a way of partitioning images into more meaningful objects without assuming any application dependent semantic models. The have carried out proof of concept of this idea using quasi-inclusion criterion.

Syntactic features can serve as an important source of information that helps to close the gap between low level features and semantic interpretation of the scene. Of course, they do not provide evidence "strong enough" to reliably group regions with very different colour, texture and motion into complex semantic objects. However, we argue that they can significantly improve the performance of automatic bottom-up segmentation techniques bringing them closer to the semantic level by allowing creation of large meaningful regions. Furthermore, syntactic and semantic sources of information are not mutually exclusive, and if available they should be both used to complement the segmentation procedure. Syntactic features can be useful even in the case of supervised scenarios where they can facilitate more intuitive user interactions e.g. by allowing the segmentation process make more "intelligent" decisions whenever the information provided by the user is not sufficient.

In this paper, we focus on the integration of syntactic features into the image segmentation process and propose a practical solution for fully automatic, robust and fast segmentation that is potentially useful for object-based multimedia applications. Our objective is to automatically segment images into large regions generally corresponding to objects or parts of objects whilst avoiding creation of regions spanning more than one semantic object. We believe that incorporation of syntactic features into the segmentation process is a key element to meet this objective and allow satisfactory results with challenging "Real World" images.

We adopt the RSST region merging algorithm as a convenient framework for the introduction of syntactic features. Syntactic features are extracted by structure analysis, performed within the RSST framework, and are based on the shapes and spatial configuration of image regions. The initial partition, which is required for initial structure analysis, is obtained by the RSST algorithm implemented as in [4]. Then, region features are extended by a new color model and region boundary criterion. Subsequently, the merging order and merging criteria are re-defined based on the extracted syntactic features and the merging process continues until there are no remaining region pairs fulfilling the merging criteria.

The remaining of the paper is organized as follows: The next section describes briefly the RSST framework and the proposed extensions. The syntactic features extracted and solutions for fast structure analysis are described in section 3. Section 4 briefly discusses the integration of the new homogeneity criteria into the merging framework. Selected results for a variety of challenging images are presented in section 5 and conclusions are formulated in section 6.

2. RSST SEGMENTATION FRAMEWORK

The RSST segmentation starts by mapping the input image into a weighted graph [7]. The regions form the nodes of the graph and the links between neighboring regions represent the merging cost. Merging is performed iteratively. At each iteration two regions connected by the least cost link are merged. Merging two regions involves creation of a joint representation for the new region and updating its links with its neighbors. The process continues until either a desired number of regions or a minimum link cost is reached. In the original approach [4], the merging cost is based on two features: region average color and size.

The geometric properties proposed in this paper are integrated in the RSST framework by employing new measures for calculating the cost of merging two neighboring regions. The introduced features control the merging order and provide a basis for a reliable stopping criteria. To facilitate fast analysis of shape and spatial configuration the region model is extended by boundary representation. It should be noted that the proposed analysis does not require any parameterization or ordering of the contour pixels leading to fast contour merging even in the case of complex region topology.

3. SYNTACTIC VISUAL FEATURES

Syntactic features advocated by Ferran and Casas include homogeneity, compactness, regularity, inclusion and symmetry [1]. A subset of the above features will be integrated into the segmentation process by performing structure analysis providing basis for complex homogeneity criteria used for quantifying cost of merging neighboring regions. Each homogeneity criterion will be evaluated for pairs of neighboring regions denoted as R_i and R_j and values of the corresponding cost measure will be mapped into the range $[0, 1]$.

3.1. Homogeneity

In the current approach we consider only spatial color homogeneity, but in the future, we plan to ntegrate texture features [5] and shape homogeneity. The color is represented using the CIE LUV color space, which is approximatively perceptually uniform. The initial segmentation is based on the regions' average color and the homogeneity criterion is calculated using the Euclidean distance. However, as the regions grow, average values become insufficient for effective characterization of their color. Therefore, when the total number of regions falls below a pre-defined value (typically 255) an extended color model is used instead. In this model, each region contains a list of pairs of color/population. After the initial segmentation, each list contains only one such pair representing each region's average color and size. Subsequenty, whenever two regions are merged, the joint representation is created by concatenating lists from both regions. This extended color representation has the advantage of low requirements for both storage and updating the merging cost. Also, no prior assumption need be made about the underlying color distribution. It has been shown by Fauqueur and Boujemaa [6] that such representations can be efficiently compared using quadratic distance. In our current implementation, the creation of the joint color representation is followed by density-based outlier detection and pairs of colors marked as outliers are excluded from the computation of the quadratic distance.

In addition to the above extended colour model, over-segmentation of objects with slow gradual color changes is prevented by adopting the so-called *Boundary Melting* approach [8] which favors merging of regions with low magnitude of color gradient along their common boundary. The merging cost is computed from the following formula:

$$c_g(i,j) = 1 - \frac{W_{ij}}{\min\{l_i, l_j\}} \qquad (1)$$

where W_{ij} is the number of weak edges on the common boundary and l_i,l_j are the perimeter lengths of regions R_i,R_j.

3.2. Regularity (low complexity)

Further evidence for region merging is provided by the fact that the shape complexity of most objects tends to be rather low. In the proposed approach, both global and local shape complexities are explored.

Shape global complexity x_i of an region R_i can be defined as the ratio between its perimeter length l_i and the square root of its area a_i. Changes in shape complexity for pair (R_i, R_j) caused by their merging are measured as: $h(i,j) = \min\{x_i, x_j\} \big/ x_{ij}$, where x_{ij} denotes shape complexity of a hypothetical region formed by merging R_i and R_j. The min function was chosen to prevent biasing of the measure by very complex regions. Values of $h(i,j)$ are mapped into range $[0, 1]$ using the following formula:

$$c_x(i,j) = clip\{1.1 - 0.69 \cdot h(i,j)\} \qquad (2)$$

where $clip$ denotes a clipping operation of values from outside the range $[0, 1]$. Low values of c_x provide strong evidence for merging the two regions e.g. values below 0.4 indicate that the joint region will have lower shape complexity than either of the two original regions.

Local shape complexity or boundary jaggedness $J(p)$ at pixel p belonging to the contour of region R_i is estimated using an adaptation of the corner detector based on the Moravec operator [9]. Using $J(p)$ we can easily compute average jaggedness along the whole boundary of R_i and jaggedness of the common boundary between two neighboring regions.

3.3. Compactness

It is well known that "real world" objects tend to be compact (i.e they exhibit adjacency of their constituent parts). In the case of the RSST algorithm, adjacency of constituent parts is imposed intrinsically during link initialization where regions are considered adjacent when they have at least one adjacent pixel. We extend this binary notion of adjacency by defining an adjacency merging cost measure:

$$c_a(i,j) = 1 - \frac{L_{ij}}{min\{l_i, l_j\}} \qquad (3)$$

where L_{ij} is the length of the common boundary between regions R_i and R_j.

However, favoring the creation of compact and simple regions also discourages formation of large regions corresponding to occluded background - see for example Figure 1. This problem can be avoided, without affecting the main central objects, by favoring merging of regions fulfilling condition $B_{ij} \cap F \neq \emptyset$ where B_{ij}

Fig. 1. Formation of occluded background. Points $B_{ij} \cap F$ are marked in red.

denotes the common boundary between the two regions and F is the image boundary.

3.4. Inclusion

Real world objects may contain holes, but very often smaller and (usually) less significant parts are included inside larger objects. Therefore, total inclusion of a small region inside another provides very strong evidence for merging the two regions. In the merging procedure described in the next section such evidence overrides all other merging criteria. Total inclusion can be seen as a special case of adjacency and can be detected either by checking condition $c_a(i,j) = 0$ or ,in the case of the RSST algorithm, by checking if the link between the two regions is the only link for either.

4. MERGING PROCEDURE

Geometric properties and merging costs defined in the previous section have to be somehow integrated in the single measure used to calculate link weights. Our observations indicate that strong evidence from just a subset of features indicating either a ”very good” or ”very bad” merge is usually more reliable than for example their weighted sum. Moreover, the merging criteria should somehow adapt as the segmentation progresses. Therefore, we divide the merging process into stages. In a given stage, links corresponding to pairs of neighboring regions, are allowed to compete for merging only if they fulfil the merging criteria defined for this stage. The merging order is controlled by cost $c_o(i,j)$ calculated as a weighted sum of costs related to color homogeneity, changes in shape complexity and adjacency: $c_o(i,j) = 3 \cdot c_c(i,j) + c_a(i,j) + c_x(i,j)$. If a link does not fulfill the merging criteria specific to the current stage its cost is set to infinity. Each stage continues until there are no more links with finite cost associated with it. When a new stage starts, the merging criteria are redefined and costs for all links are recalculated.

The function used to calculate link weights can be expressed in the following general form:

$$c(i,j) = \begin{cases} c_o(i,j) & \text{if } H_s(i,j) \\ \infty & \text{otherwise} \end{cases} \tag{4}$$

where $H_s(i,j)$ denotes boolean operator testing if a link corresponding to pair (R_i, R_j) fulfills the merging criteria defined for stage s. The criteria are defined by a set of thresholds controlling the values of merging costs measures and region properties described in the previous section. To avoid over-segmentation and at the same time ensure formation of meaningful regions, the strengths of color and geometric criteria should be carefully combined, i.e. weak color criteria should be used in combination with strong geometric criteria and vice versa.

The segmentation process consist of the following main stages:

- Initial segmentation into N regions based on color and size,
- Merging based on c_c,c_x and c_a
- Boundary melting,
- Reduction of shape complexity ($c_x < 0.4$), weak color crit.,
- Removal of small regions, weak color criterion,
- Merging very complex regions, weak color criterion,
- Merging jagged regions, weak color criterion,
- Merging reg. with very similar color, weak geometric crit.,
- Creation of simple regions, weak color criterion,
- Creation of occluded background based on condition $B_{ij} \cap F \neq \emptyset$ and weak color criterion.

The above multistage approach has proven to be particulary convenient for experimentation and development purposes however some stages could be amalgamated to further reduce the computational cost. We are also conscious that the above approach requires a certain number of thresholds which were adjusted based on empirical evidence. However, as is demonstrated in the next section, the syntactic features are generic in nature and once defined, they significantly reduce dependency on parameters controlling other features, e.g. color.

5. RESULTS

The proposed algorithm was applied to images from various sources and the results are presented in Figure 2. This figure also shows the results of the Blobworld segmentation algorithm [1] for the same images for comparison. The Blobworld algorithm has been extensively tested and produced very satisfactory results [2] in the past.

The results indicate that the incorporation of the syntactic features in the RSST algorithm not only provides a reliable stopping criteria for merging but also results in the creation of large regions which rarely span more than one semantic object. The proposed method avoids over-segmentation and in contrast to the Blobworld algorithm, produces accurate region boundaries.

However, perhaps the most important property of our algorithm is that it works well on images from different collections, such as standard test images, professional databases, digital photographs and video keyframes from broadcast TV, all *using a fixed set of parameter values*. This, along with the fact that it took less than 3 seconds on Pentium III 600 MHz to segment an image of CIF size (352x288)[2], makes the proposed algorithm ideal for object-based applications such as content-based image indexing and retrieval in large collections or definition of interest regions for content-based coding of still images in the context of the JPEG2000 [5], for example.

6. CONCLUSIONS

A new method was presented for the segmentation of color images into large regions corresponding to the objects present in an image. To the best of our knowledge this is the first comprehensive and practical solution for integration of syntactic features into an

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

[2] Blobworld ,using mostly Matlab code, required over 30 minutes for this task.

automatic image segmentation process. It was shown that the geometric features can significantly improve the perceptual quality of segmentation. The results indicate that the proposed method allows meaningful partitioning of real world images from various sources without any adjustment to the parameter values.

The proposed list of merging cost components is by no means complete and could be further extended – by adding symmetry analysis, for example. It should also be noted that the proposed set of features is not functionally orthogonal and there are some overlaps between different syntactic features, e.g. between global shape complexity and compactness. Addressing such issues will be the focus of our future research.

7. REFERENCES

[1] C. F. Bennstrom and J. R. Casas, "Binary-partition-tree creation using a quasi-inclusion criterion," in *IEEE Computer Society Press, in the proceedings of the Eighth International Conference on Information Visualization (IV). London, UK*, 2004.

[2] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Color- and texture-based image segmentation using em and its application to image querying and classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 8, pp. 1026–1037, August 2002.

[3] T. Y. Lui and E. Izquierdo, "Scalable object-based image retrieval," in *IEEE International Conference on Image Processing, ICIP 2003, Barcelona, Spain*, 2003.

[4] 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 Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 7, pp. 802–813, Nov 1998.

[5] V. Mezaris, I.Kompatsiaris, and M.G.Strintzis, "Still image segmentation tools for object-based multimedia applications," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 18, no. 4, pp. 701–725, June 2004.

[6] J. Fauqueur and N. Boujemaa, "Region-based image retrieval: Fast coarse segmentation and fine color description," *Journal of Visual Languages and Computing (JVLC), special issue on Visual Information Systems*, vol. 15, pp. 69–95, 2004.

[7] O. Morris, M. Lee, and A. Constantinides, "Graph theory for image analysis: an approach based on the shortest spanning tree," in *IEE Proceedings*, vol. 133, April 1986, pp. 146–152.

[8] M. Sonka, V. Hlavac, and R. Boyle, *Image Processing, Analysis, and Machine Vision*, 2nd ed. PWS, 1998, ch. 5.

[9] H. Moravec, "Towards authomatic visual obstacle avoidance," in *In Proceedings of the 5th International Joint Conference on Artificial Inteligence, Caregie-Mellon University, Pittsburg, PA*, August 1977.

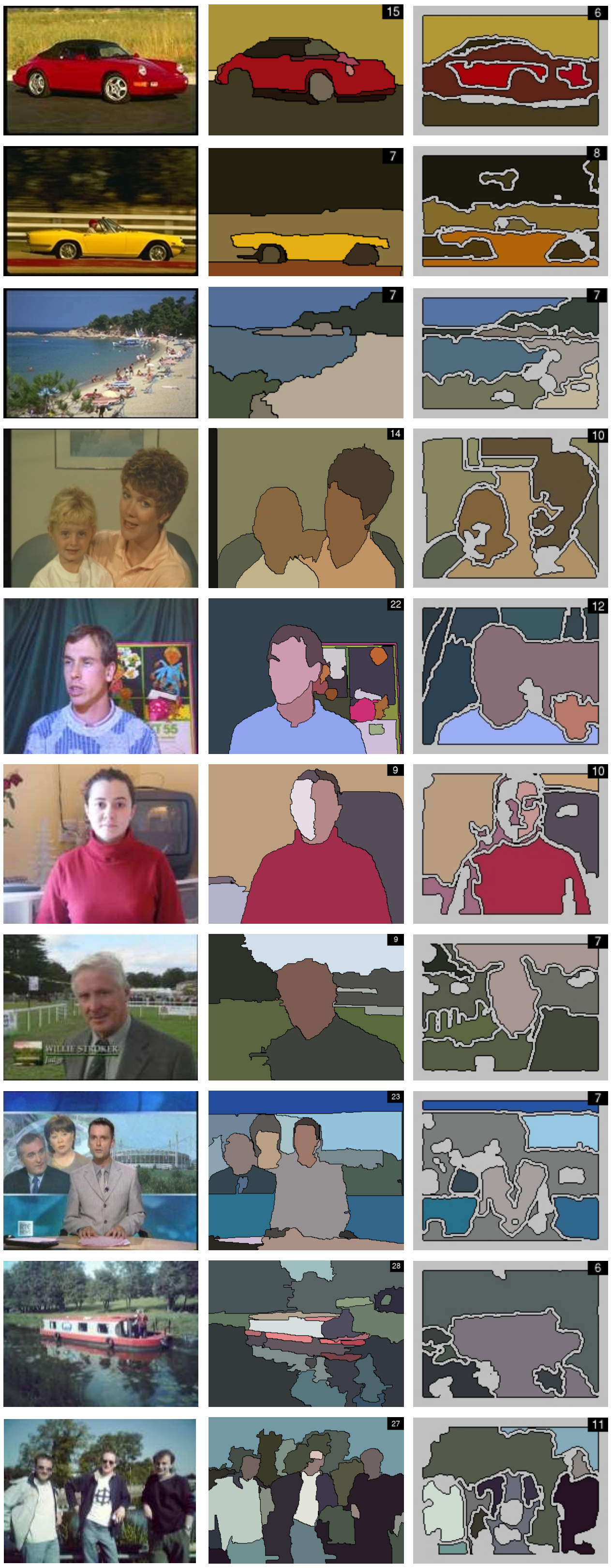

Fig. 2. Segmentation results for images from various sources including: Corel gallery (rows 1-3), key frames from well known MPEG4 test seq. (rows 4-5), digital camera (row 6), key frames from news broadcast (rows 7-8), mobile phone camera (rows 9-10). Results of the proposed algorithm and Blobworld are shown in the second and third column respectively. Labels in top-right corners represent numbers of regions.