USING DEMPSTER-SHAFER THEORY TO FUSE MULTIPLE INFORMATION SOURCES IN REGION-BASED SEGMENTATION

Tomasz Adamek, Noel E. O'Connor

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

ABSTRACT

This paper presents a new method for segmentation of images into large regions that reflect the real world objects present in a scene. It explores the feasibility of utilizing spatial configuration of regions and their geometric properties (the so-called *Syntactic Visual Features* [1]) for improving the correspondence of segmentation results produced by the well-known *Recursive Shortest Spanning Tree* (RSST) algorithm [2] to semantic objects present in the scene. The main contribution of this paper is a novel framework for integration of evidence from multiple sources with the region merging process based on the *Dempster-Shafer* (DS) theory [3] that allows integration of sources providing evidence with different accuracy and reliability. Extensive experiments indicate that the proposed solution limits formation of regions spanning more than one semantic object.

***Index Terms*—** Image segmentation, Dempster-Shafer theory.

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 and is the first step of many object-based applications. A large number of approaches to image or video segmentation have been proposed in the past [4, 5, 6, 7]. This paper focuses on automatic *visual feature-based* segmentation of images, which does not require developing models of individual objects.

Many existing approaches aim to create large regions using (relatively simple) homogeneity criteria based on colour or texture. However, their application is often limited as they fail to create meaningful partitions due to either the complexity of the scene or difficult lighting conditions. In [1], Ferran and Casas propose a new source of evidence for region merging which they term *Syntactic Visual Features*, representing geometric properties of regions and their spatial configurations, e.g. homogeneity, compactness, regularity, inclusion or symmetry. They advocate their use in bottom-up approaches as a way of partitioning images into more meaningful entities without assuming any application dependent semantic models. Although the idea of using geometric properties within a region merging process to create more meaningful partitions is not new in itself, meaningful integration of evidence from different sources into a single merging cost is a non-trivial task and the approaches currently available in the literature are often ad-hoc in nature. Clearly, it is necessary to take into account the reliability of different sources of information when integrating them within the region merging algorithm. Moreover, certain geometric properties in some cases cannot be measured and therefore cannot provide any evidence supporting the region merging hypothesis, i.e. the framework should be able to accept that certain measurements may not be precise (doubtful) or even "unknown" in some cases. Also, to the best of the author's knowledge, currently there is no objective evaluation available in the literature to support the usefulness of any such features.

In this paper several extensions to the the well-known *Recursive Shortest Spanning Tree* (RSST) algorithm [2]RSST [2] are proposed among which the most important is a novel framework for integration of evidence from multiple sources of information within the region merging process. The RSST algorithm is selected as it provides a convenient framework for integration of a region's geometric properties, and also due to its speed. The integration framework is based on the *Dempster-Shafer* (DS) theory [3] and allows integration of sources providing evidence with different accuracy and reliability. DS theory is selected as it appears more general and more flexible than the commonly used probability model and allows representation of states of knowledge and opinions not possible with probability models [8]. Other contributions include a new colour model and colour homogeneity criteria, and practical solutions to analysis of region's geometric properties and their spatial configuration.

The 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 [5]) 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. In certain applications stopping criterion may not be required, e.g. when revealing hierarchical structure of the scene by creating a *Binary Partition Tree* (BPT) [6].

Since at each iteration only two regions connected by the least cost link are merged the merging order is exclusively controlled by the function used to compute the merging cost (often referred as merging criterion [6, 1]). The most commonly used criterion is based on the colour homogeneity computed based on the regions' average colours [5]. The RSST algorithm is fast and capable of producing meaningful partitions containing large number of regions (typically more than 100) useful in many applications [5, 6]. However, the commonly used colour homogeneity criteria are insufficient to produce meaningful results in applications where further simplification of the scene is required, i.e. it is incapable of producing large regions with a high probability of corresponding to semantic notions [6]. One of the main reasons for poor performance of the original algo-

This research was supported by the European Commission under contract FP6-027026, Knowledge Space of semantic inference for automatic annotation and retrieval of multimedia content - K-Space

rithm is its utilization of identical simple homogeneity criterion at all stages of the merging process, i.e. the same merging criterion is used at the level of pixels and at stages where regions contain thousands of pixels. We believe that the limitations of the RSST algorithm can be overcame by modifying and extending both the merging cost measure and the stopping criterion. In this paper we focus on developing new merging criteria based on evidence provided by different features (extended colour representation and geometric properties) fused using an integration framework based on DS theory. A new stopping criterion aimed at producing partitions containing the most salient objects present in the scene will be reported elsewhere.

2. OVERVIEW OF THE APPROACH

The segmentation process is divided into two stages. The initial partition, required for structure analysis, is obtained by the RSST algorithm with the original merging criterion implemented as in [5] 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, region representation is extended by region boundary and a new colour model. Subsequently, homogeneity criteria are re-defined based on colour and *Syntactic Visual Features* and the merging process continues until a certain stopping criterion is fulfilled. The new merging order is based on evidence provided by different features (colour and geometric properties) fused using the proposed integration framework. It should be stressed that all extensions proposed in the following sections consider the second stage of merging when regions already have meaningful shape.

The merging cost utilizing evidence provided by multiple sources such as colour homogeneity and the *Syntactic Visual Features* is computed based on the DS theory [3] which allows taking into account the reliability of different sources of information, e.g. colour homogeneity provides much more reliable evidence than any geometric properties, and also taking into account that certain measurements may not be precise (doubtful) or even "unknown" in some cases.

3. COLOUR HOMOGENEITY

As segmentation progresses and regions grow, average values may become unsuitable for effective characterization of their colour. To ensure that this simplification does not compromise the merging order, a fine and compact representation of a region's colour, motivated by the so-called *Adaptive Distribution of Colour Shades* (ADCS) proposed in [9], is used during the second merging stage. In this model, 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) which can more precisely represent its complex colour variations [7]. This extended colour representation has the advantage of low requirements for both storage and for updating the merging cost. Also, no prior assumption needs to be made about the underlying colour distribution – see Figure 1. Two ADCS representations can be efficiently compared using *Quadratic Distance* [9]. The proposed merging cost based on the colour difference between regions r_i and r_j is calculated as [7]:

$$C_{ext}(i,j) = \left[d^2_{quad}(\mathbf{I},\mathbf{J})\right]^{q_{ext}} \cdot \frac{1}{a_{img}} \cdot \frac{a_i a_j}{a_i + a_j} \qquad (1)$$

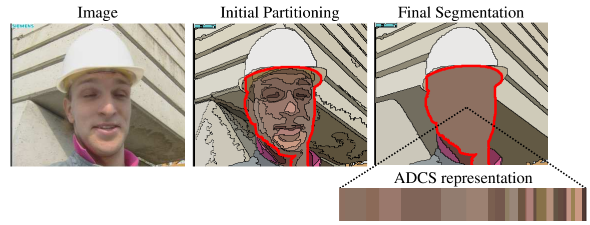


Fig. 1. Example of ADCS representation.

where $d^2_{quad}(\mathbf{I},\mathbf{J})$ is the quadratic distance between regions' colour distributions characterized by two ADCS representations $\mathbf{I}$ and $\mathbf{J}$, a_i and a_j denote region sizes, a_{img} is the size of the entire image (such normalization by a_{img} does not affect the merging order but simplifies the integration making C_{ext} size independent), and parameter q_{ext} controls the balance between colour distance and the size dependent scaling factor and can be found experimentally during the training.

4. SYNTACTIC VISUAL FEATURES

This section discuses two region geometric properties, adjacency and changes in global shape complexity.

4.1. Adjacency

It is a well known fact that "real world" objects tend to be compact, i.e. they exhibit adjacency of their constituent parts [1]. In the case of the RSST algorithm, adjacency of parts is imposed intrinsically during link initialization where regions are considered adjacent when they have at least one adjacent pixel. It is proposed here that this binary notion of adjacency be extended by defining an adjacency measure:

$$C_{adj}(i,j) = 1.0 - \frac{l_{ij}}{\min\{l_i, l_j\}} \qquad (2)$$

where l_{ij} is the length of the common boundary between a pair of neighbouring regions r_i and r_j, and l_i and l_j are their perimeter lengths.

Although simple, the above measure provides quite rich information about geometrical relations between both regions, i.e values of $C_{adj}(i,j)$ close to zero indicate almost total inclusion whereas values close to one point to weak adjacency and therefore provide strong evidence against merging.

4.2. Regularity (low complexity)

Intuitively, more complex shapes have a longer perimeter in relation to their area compared to simple shapes. Therefore, global shape complexity x_i of a region r_i can be measured as the ratio between its perimeter length l_i and the square root of its area a_i: $x_i = l_i/\sqrt{a_i}$. Finally, average changes of global shape complexity for a pair of neighbouring regions r_i and r_j caused by their merging are measured as:

$$C_{cpx}(i,j) = \frac{x_{ij}}{\left[\frac{a_i x_i + a_j x_j}{a_i + a_j}\right]} \qquad (3)$$

where x_{ij} denotes shape complexity of a hypothetical region formed by merging r_i and r_j. Low values of $C_{cpx}(i,j)$ provide evidence for merging the two regions e.g. values below 1.0 indicate that the complexity of the joint region is lower than the average of the complexities of the two original regions.

5. INTEGRATION OF EVIDENCE

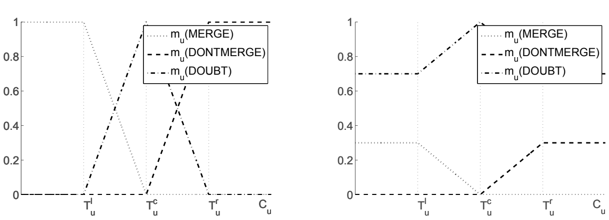
5.1. Application of the DS theory to Region Merging

Details regarding the DS theory can be found in [3]. This section describes the major aspects of the application of the DS theory to the problem of integration of multiple features within the region merging process such as: (i) definition of the frame of discernment, (ii) general form of the belief structure used to model the way a piece of evidence is brought by a source to a proposition, (iii) taking into account source reliability, (iv) combination of beliefs from multiple sources and (v) a new merging cost formula based on the combination of beliefs.

Let the frame of discernment Ω represent a set of two hypothesis, namely $MERGE$ and $DONTMERGE$, which are exhaustive and exclusive $\Omega = \{MERGE, DONTMERGE\}$. The power set 2^Ω of Ω is composed of 4 propositions A: $2^\Omega = \{\, \emptyset, \{MERGE\}, \{DONTMERGE\}, \{MERGE \cup DONTMERGE\}\, \}$. The proposition $\{MERGE \cup DONTMERGE\}$ will be also referred to in this paper as doubt $\{DOUBT\}$.

Let us define a set of U independent evidence sources $S_1, \ldots, S_U$ where S_1 always denotes the colour homogeneity criterion and $S_2, \ldots, S_U$ denote a set of other properties associated with a link between two regions. A piece of evidence (measurement) C_u brought by a source S_u to a proposition A is modelled by the belief structure m_u called the *Basic Belief Assignment* (BBA) formally defined as: $m_u : 2^\Omega \longrightarrow [0,1]$ with $m_u(\emptyset) = 0$ and $\sum_{A \subseteq \Omega} m_u(A) = 1$. In other words, for each measure C_u (e.g. a colour homogeneity or a syntactic feature) model m_u maps each value of C_u into a belief on a proposition (singleton or composed hypothesis) of 2^Ω.

In this work the models are based on both expert knowledge and statistical knowledge (i.e. parameter estimation) obtained from the training collection similar to the methodologies proposed in [8] for training of facial emotions classifiers. The general form of such models is shown in Figure 2a. Let us denote the neutral value of measure C_u as T_u^c, i.e. the value of C_u which does not provide evidence either for merging or not merging and for which the entire mass of belief is associated with doubt ($MERGE \cup DONTMERGE$) while the masses of belief associated with propositions $MERGE$ or $DONTMERGE$ are equal to zero. As the value of C_u decreases(increases) from T_u^c to $T_u^l(T_u^r)$ the mass of belief associated with doubt decreases while the mass of belief associated with $MERGE(DONTMERGE)$ increases.



(a) BBA for a reliable source (b) Discounted BBA.

Fig. 2. General form of BBA functions.

When a source is considered not completely reliable the confidence in this source can be attenuated [3]. Figure 2b shows an example of the BBA from Figure 2a discounted by factor $\alpha_u = 0.3$. It should be observed that discounting simply transfers the belief from propositions $MERGE$ and $DONTMERGE$ to $DOUBT$.

BBA for two or more information sources are combined using the orthogonal sum also called the Dempster's rule of combination [3]. According to this operator, which is commutative and associative, the combined belief structure $m^\oplus$ is defined as: $m^\oplus = m_1 \oplus m_2 \oplus \ldots \oplus m_U$, where for two sources of information S_u and S_v the combined belief structure $m^\oplus$ is defined as:

$$\forall A \subseteq \Omega \qquad m^\oplus(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_u(B) \cdot m_v(C) \qquad (4)$$

where K can be interpreted as a measure of the conflict between the sources to be combined ($m^\oplus(\emptyset)$) and is defined as: $K = \sum_{B \cap C = \emptyset} m_u(B) \cdot m_v(C)$.

Finally, the new cost of merging of two neighbouring regions r_i and r_j based on evidence from one or a combination of sources is defined using an empirically derived formula:

$$C_{Total}(i,j) = m^\oplus_{(i,j)}(DONTMERGE) - m^\oplus_{(i,j)}(MERGE) \qquad (5)$$

where $m^\oplus_{(i,j)}(DONTMERGE)$ and $m^\oplus_{(i,j)}(MERGE)$ are combined beliefs that measurements obtained for the pair r_i and r_j from all integrated sources of evidence support the hypothesis $MERGE$ and $DONTMERGE$ respectively.

5.2. Designing Belief Structures for each Source of Information

The design of BBAs for each source of information S_u involves selection of the thresholds T_u^l, T_u^c, T_u^r and the discounting factor α_u all of which were estimated using a training collection of manually segmented images. First, the training collection of images with ground-truth was used to automatically generate a training collection containing examples of links with an associated set of measurements (i.e. colour homogeneity and geometric properties) and classified as either "link to be merged" or "link which should not be merged". Then, the thresholds T_u^l (T_u^r) were estimated by finding the minimum (maximum) value of C_u in the training population of links which "should" ("should not") be merged. In fact, to avoid bias caused by a single link (e.g. generated from an "accidental" manual segmentation) the above measures were found individually for each image in the training collection and the final values of parameters T_u^l and T_u^r were computed as their average values. The value of T_u^c was computed simply as mid point of T_u^l and T_u^r.

For simplicity, the discounting factor corresponding to colour homogeneity α_1 is always set to one, since intuitively the evidence provided by colour homogeneity is the most reliable of all sources. For the remaining sources discounting factors are found by combining them with the colour homogeneity criterion using the proposed framework and searching for discounting factor/factors minimizing the average spatial segmentation error on the training collection [10].

Additionally, we observed that the measure C_{cpx} requires a more flexible form of BBA than the general one due to the fact that whenever the hypothetical region r_{ij} (created by merging two neighbouring regions r_i and r_j) shares a significant part of its contour with the image border the measure C_{cpx} does not provide any useful evidence, i.e. reliability of evidence provided by the measure C_{cpx} depends on the length of the contour of r_{ij} common with the image border. This is taken into account by additional transfer of the belief from propositions $MERGE$ and $DONTMERGE$ to the proposition $DOUBT$ depending on the adjacency of r_{ij} with the border of the entire image by applying an additional discounting operation for each pair of regions r_i and r_j using an additional discounting factor defined as: $\alpha'_{cpx}(i,j) = 1.0 - l_{Iij}/l_{ij}$, where l_{ij} is the perimeter length of r_{ij} and l_{Iij} denotes the length of the common boundary between r_{ij} and the entire image. Note that whenever $l_{Iij} = l_{ij}$ the total mass of belief is assigned to proposition $DOUBT$. This example is a good illustration of the ignorance management capabilities of DS theory.

Finally, it should be noted that recently two approaches to perceptual grouping incorporating *Gestalt Laws* using the notion of belief have been proposed [11, 12]. Although the approach proposed in this paper shares the same basic idea of using the DS theory for combining evidence obtained by structure analysis as the above methods, all key elements of the approach proposed here are novel such as: (i) the form of *Basic Belief Assignment* functions and their training, (ii) geometric measurements used, and (iii) the grouping framework.

6. RESULTS

The impact of the above extensions on segmentation quality was evaluated in rigorous experimentation using a collection 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. Ground-truth segmentation masks were created manually. The performances were measured in terms of average spatial segmentation error [10] (computed based on the number of misclassified pixels and normalized by the size of the image) using two-fold cross validation. It should be noted that the evaluation method from [10] takes into account over- and under-segmentation.

In order to evaluate solely the merging criteria and avoid the impact of stopping criteria in all experiments the number of regions available from the ground-truth were used as the "optimal stopping criterion".

The best result was obtained for the approach integrating the new colour homogeneity criterion together with the adjacency measure and the measure of shape complexity change which almost halved the average spatial segmentation error compared to the original RSST approach from [5]. Integration of the syntactic features reduced the spatial segmentation error by around 15% compared to the new colour homogeneity criteria used alone (C_{ext}). Also, manual inspection of the results confirmed that the proposed extensions significantly improve correspondence of segmentation to semantic notions. Detail results obtained by several variations of the proposed method can be found in [7].

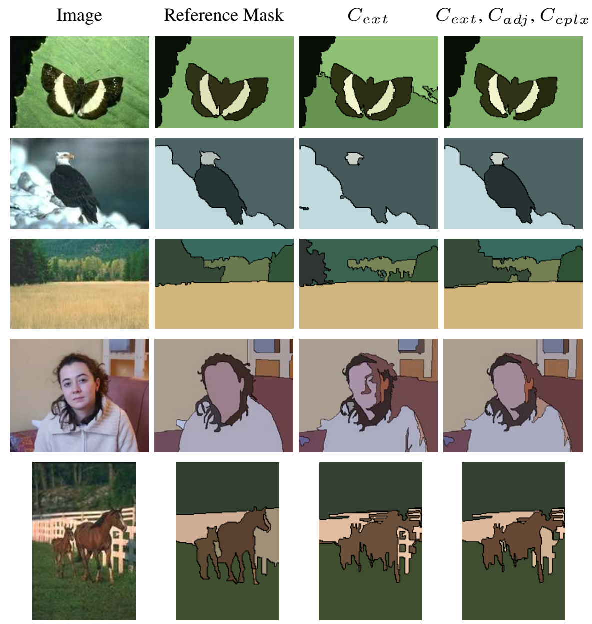


Fig. 3. Segmentations obtained by the proposed merging criterion.

7. CONCLUSIONS

To the best of author's knowledge this paper presents the first comprehensive and practical solution for integration of multiple features, including geometric properties, into a region merging process and demonstrates its performance in extensive experiments. It is shown that syntactic features can provide an additional basis for region merging criteria which limit formation of regions spanning more than one semantic object. The experiments demonstrate that the proposed solutions are generic in nature and allow satisfactory segmentation of real world images without any adjustment to the algorithm's parameters.

8. REFERENCES

[1] 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.

[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.

[3] 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.

[4] F. Morier, J. Benois-Pineau, D. Barba, and H. Sanson, "Robust segmentation of moving image sequences," in *Proc. IEEE Int'l Conf. on Image Processing*, Oct. 1997, vol. 1, pp. 719 – 722.

[5] 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.

[6] 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.

[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] Z. Hammal, A. Caplier, and M. Rombaut, "A fusion process based on belief theory for classification of facial basic emotions," in *Proc. 8th Int'l Conf. on Information Fusion, Philadelphia, USA*, July 2005.

[9] J. Fauqueur and N. Boujemaa, "Region-based image retrieval: Fast coarse segmentation and fine color description," *Journal of Visual Languages and Computing*, vol. 15, pp. 69–95, 2004.

[10] 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.

[11] P. Vasseur, C. Pégard, E. M. Mouaddib, and L. Delahoche, "Perceptual organization approach based on dempster-shafer theory," *Pattern Recognition*, vol. 32, no. 8, 1999.

[12] 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.