Pattern Recognition 57 (2016) 136–151

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

A competition in unsupervised color image segmentation

Michal Haindl *, Stanislav Mikeš

The Institute of Information Theory and Automation of the Czech Academy of Sciences, 182 08 Prague, Czech Republic

ARTICLE INFO

Article history:
Received 30 September 2015
Received in revised form
13 January 2016
Accepted 2 March 2016
Available online 11 March 2016

Keywords:
Unsupervised image segmentation
Segmentation contest
Texture analysis

ABSTRACT

A competition in unsupervised color image segmentation took place in conjunction with the 22nd International Conference on Pattern Recognition (ICPR 2014). It aimed to promote evaluation of unsupervised color image segmentation algorithms using publicly available data sets, and to allow for any subsequent methods to be easily evaluated and compared with the results of the contested methods under identical conditions. Our comparison of different methods is based on the standard methodology of performance assessment using an on-line verification server. We present in this paper the evaluation of the top six results submitted to the ICPR 2014 contest in unsupervised color image segmentation and compare them with 11 other state-of-the-art unsupervised image segmenters.

1. Introduction and related work

Unsupervised or supervised texture segmentation is a prerequisite for numerous applications useful for image understanding, such as the content-based image retrieval, scene analysis, automatic acquisition of virtual models, quality control, security, medical applications, and many others. Although a large number of more or less different methods have already been published [1–17], and other novel algorithms are continually appearing, this ill-defined problem is still far from having been satisfactorily solved, and cannot even be solved in its full generality, i.e., to perform optimally for any and all image segmentation tasks. Visual scenes are highly variable and each method's performance also depends on a visual scene category and on image parameters, such as resolution, illumination and viewing conditions. In addition to that, not much is known about behavior of the already published segmentation methods, including appropriate setting of their parameters; their potential user is left to randomly select one. One of the reasons for this situation is the lack of sufficient empirical data and, consequently, the absence of any counseling. This is, among other reasons, due to a lack of a reliable performance comparison between different techniques because very limited effort has been spent to develop suitable quantitative measures of segmentation quality that could be used for evaluating and comparing segmentation algorithms. Rather than advancing the most promising image segmentation approaches, novel algorithms are often introduced merely on the basis of being sufficiently different from those already described in the literature, even if they have dubious performance and have only been tested on a few carefully selected favorable examples.

The unsupervised image segmentation contest, which took place in conjunction with the ICPR 2014 Conference, aimed at overcoming these problems by suggesting the most promising approaches to the unsupervised learning and image segmentation and at unifying the verification methodology used in the image segmentation research. The contest requirements were to submit segmentation results on the generated large color texture mosaics set, a brief description of the unsupervised segmentation method, and its code or binaries and the required parameters. None of the methods was allowed to utilize user interaction or knowledge about the number of regions in the mosaic.

Although, the performance assessment of all submitted contest algorithms was briefly summarized in the presentation given at the conference, the contest framework has a much broader applicability. It can guide and inspire development of new methods and serve as a reliable and efficient means of progress checking during such an effort.

2. Contest benchmark

The contest uses the Prague texture segmentation data-generator and benchmark [18–20], which is a web-based (http://mosaic.utia.cas.cz) service designed to mutually compare, validate, and rank different texture or image segmenters – supervised or unsupervised – and to support development of new segmentation and classification methods. Although this benchmark has already been serving the community for ten years, it is being permanently upgraded while maintaining the backward compatibility of the accumulated results during its decade in service. The benchmark

* Corresponding author. Tel.: +420 266052350.
E-mail addresses: haindl@utia.cas.cz (M. Haindl), xaos@utia.cas.cz (S. Mikeš).

http://dx.doi.org/10.1016/j.patcog.2016.03.003
0031-3203/

verifies the performance characteristics of the submitted image segmenters in either supervised or unsupervised mode on potentially unlimited image/frame sets of mono-spectral, multi-spectral, bidirectional texture function (BTF), satellite, and dynamic textures using extensive sets of prevalent numerical criteria. It enables us to test their noise robustness, scale, rotation or illumination invariance, select several types of region borders, etc.

3. Contest data

The benchmark contest data sets are computer generated 512×512 pixel mosaics using a Voronoi polygon random generator filled with randomly selected natural color textures (see Fig. 1). Visual scenes contain objects from various materials; these materials are typically represented as visual textures [10,21] mapped on the corresponding object shapes. A material's appearance predominantly depends on the viewing, illumination, and shape properties, among other [21]. The viewing and illumination conditions vary somewhat for each individual texture in the test mosaic, the viewing direction follows the surface normal, and all textures have correct natural illumination. The contest data are roughly planar and as such they only approximate a real visual scene with general object shapes. However, they allow us to know the exact ideal non-subjective segmentation, and to generate test sets of any size we wish, but, most importantly, the ranking of the segmentation methods correlates well with the experiments on real natural scenes, as we have verified on the Berkeley test database [22]. The unlimited size of the test is crucial to obtain stable performance ranking. The contest uses the large size (80 textural mosaics) unsupervised *Color* benchmark without noise degradation. Piecewise linear region borders are chosen for the contest, but the benchmark allows various border types. The participants received the contest data set (Table 1) to be segmented by their methods, and they uploaded the corresponding 80 segmentation results. Another validation set with the same structure (Table 1) was used by the organizers to validate the submitted results.

Table 1 specifies the basic properties of both the contest and validation data. Both sets contain 80 texture mosaics composed of measured color textures. Forty (40) mosaics gradually increase the number of different textures per mosaic from 3 to 12, and the textural fragments are mixtures of ten thematic texture classes. The other 40 mosaics include all six different regions but contain textures from the same thematic class in each mosaic. For any row in Table 1 there are four mosaics with two different mosaic topologies, each with two alternative texture sets.

4. Performance evaluation

The benchmark has implemented the 27 most frequently used evaluation criteria categorized (see the detailed specification in the benchmark) into four groups: region-based [23] (5 criteria with the standard threshold + 5 performance curves – Figs. 2–7 – with their performance integrals over all threshold settings), pixel-wise (12 + *F*–measure curve), consistency measures (2) [22], and clustering comparison criteria (3) [24]. The performance criteria mutually compare ground-truth image regions with the corresponding machine-segmented regions. All criteria are available on two levels – averaged over the corresponding benchmark or computed for every individual test row in Table 1. The contest criterion is the average rank over 21 benchmark criteria. The top methods were verified by the organizers using the submitted codes and the validation data, which were not available to contestants. During the contest submission period, all participants could see only their results and all non-contest results in the benchmark. They could submit an unlimited number of results, and only the best one of those submitted before the deadline was considered.

5. Submitted methods

The following five methods (VRA-PMCFA, FSEG, Deep Brain Model, CGCHI, texNCUT) were submitted to the contest, and the sixth one (MW3AR8) was evaluated outside the contest because it was developed by the organizers. The texNCUT method was excluded from the finals due to a contest condition violation.

5.1. VRA-PMCFA

The Voting Representativeness–Priority Multi-Class Flooding Algorithm (based on [25]) is an unsupervised texture image segmentation framework with an unknown number of regions, which involves feature extraction and classification in the feature space, followed by flooding and merging in the spatial domain. The segmented image is divided into overlapping blocks, whose feature representations are three color Lab components and two wavelet transform components. The block size and the possible range for the number of regions are three parameters of the

Table 1
The 80 mosaics' specification for the contest and validation sets.

Number	No. of regions	Texture class	Number	No. of regions	Texture class
4	3	Mixture	4	6	Bark
4	4	Mixture	4	6	Flowers
4	5	Mixture	4	6	Glass
4	6	Mixture	4	6	Man-made
4	7	Mixture	4	6	Nature
4	8	Mixture	4	6	Plants
4	9	Mixture	4	6	Rock
4	10	Mixture	4	6	Stone
4	11	Mixture	4	6	Textile
4	12	Mixture	4	6	Wood

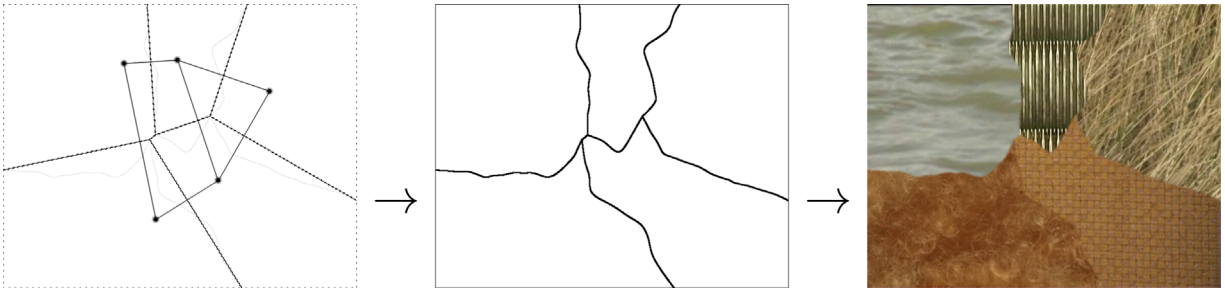

Fig. 1. Texture mosaic generating scheme.