



# 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.

© 2016 Elsevier Ltd. All rights reserved.

## 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](mailto:haindl@utia.cas.cz) (M. Haindl), [xaos@utia.cas.cz](mailto:xaos@utia.cas.cz) (S. Mikeš).

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  $512 \times 512$  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-PMCF, 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.

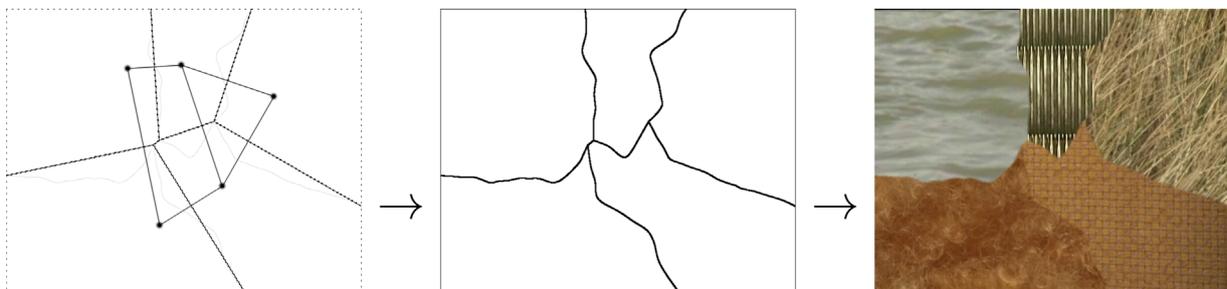
#### 5.1. VRA-PMCF

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

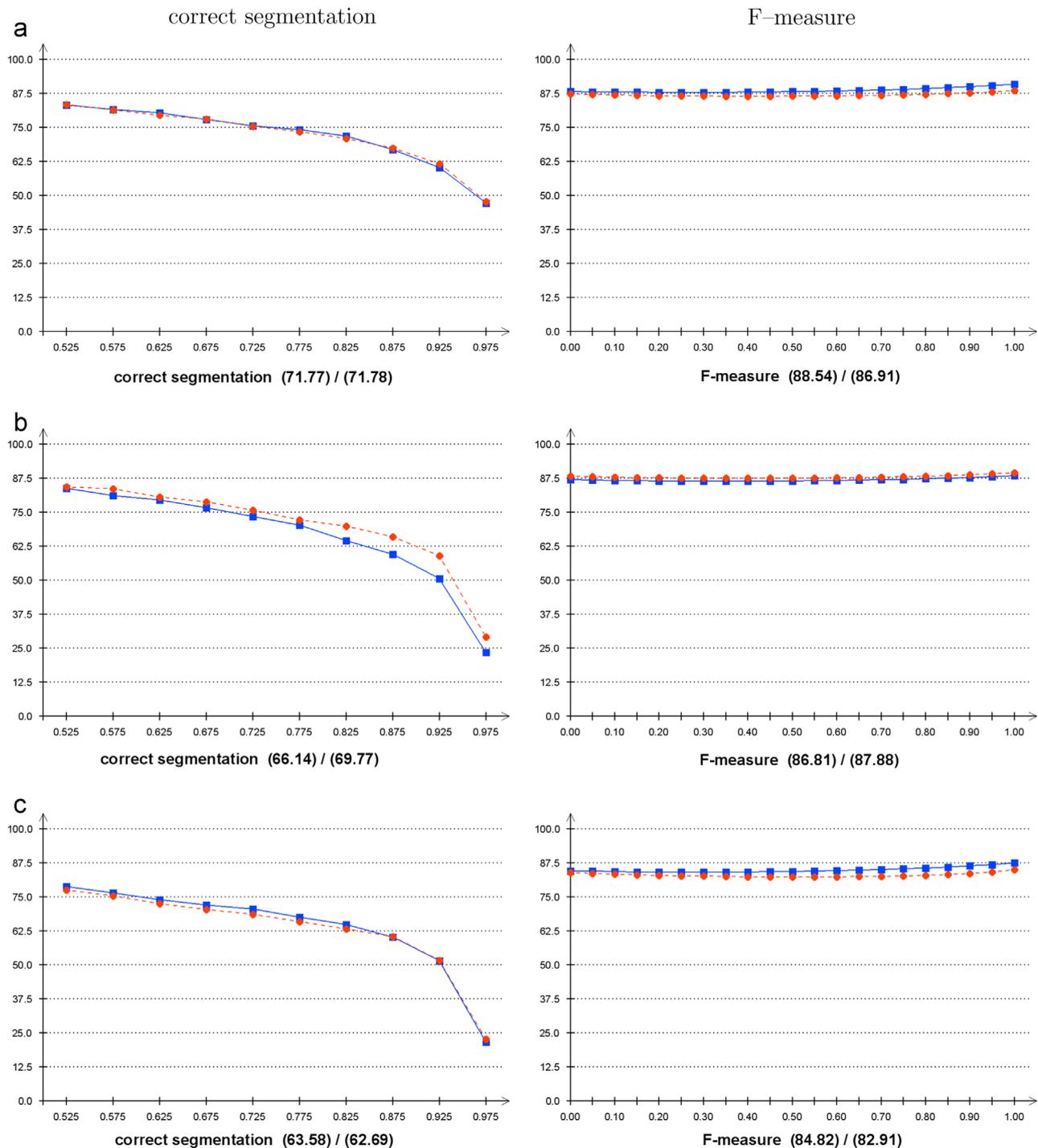
**Table 1**

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.



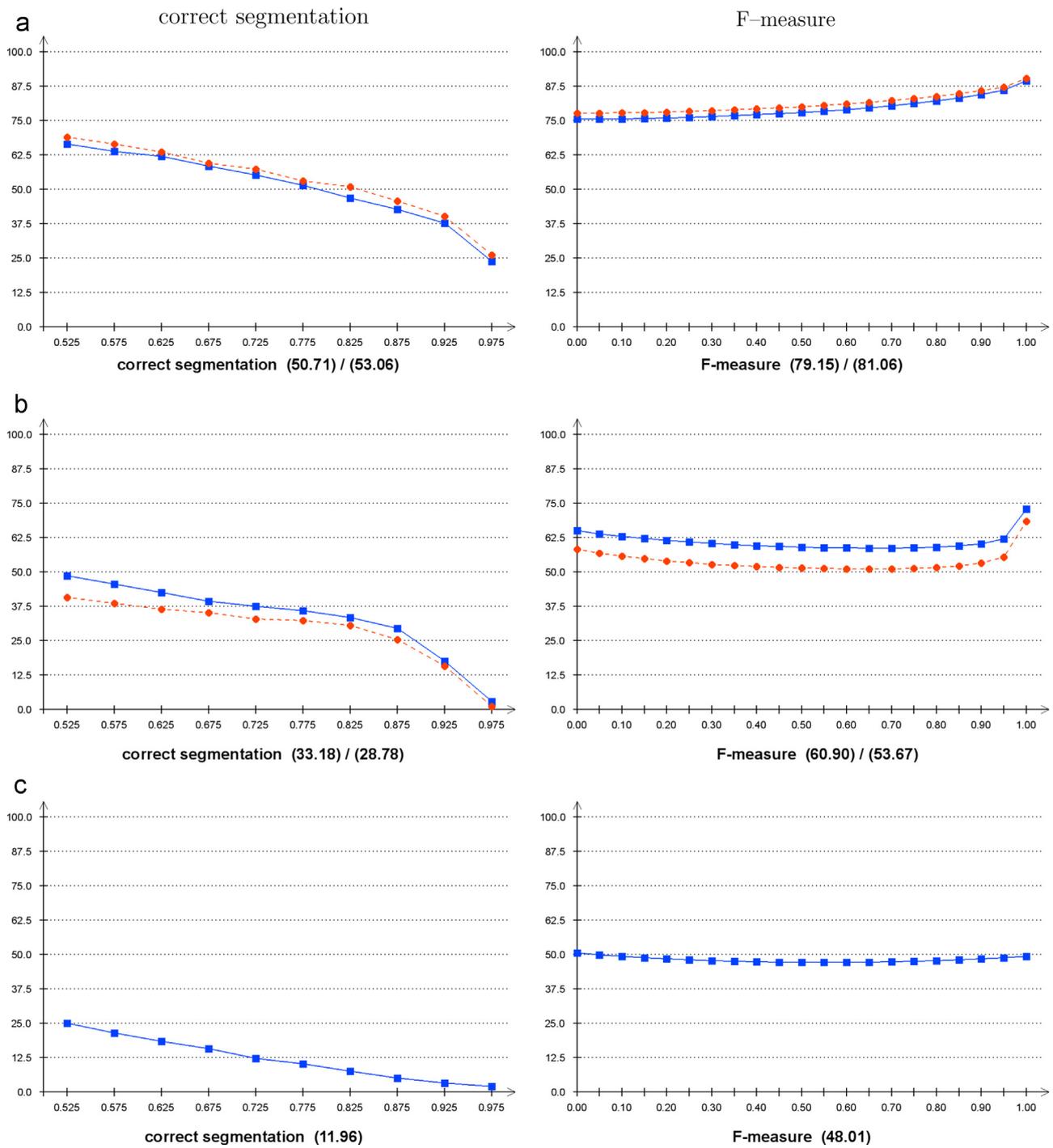
**Fig. 2.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: VRA-PMCFA (a), texNCUT (b), FSEG (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

method. The distribution of the features for the different classes is obtained by a block-wise unsupervised voting framework using the block's grid graph or its minimum spanning tree and the Mallows distance. The final clustering is obtained by using the k-centroids algorithm with the Bhattacharyya distance. The flooding algorithm used is the Priority Multi-Class Flooding Algorithm [25], which assigns pixels to labels using the Bayesian dissimilarity criteria. Finally, a region-merging method, which incorporates boundary information, is introduced for obtaining the final segmentation map. This scheme is executed for several regions, and the number of regions is selected to minimize a criterion that takes into account the average likelihood per pixel of the

classification map and penalizes the complexity of the regions boundaries.

## 5.2. FSEG

This factorization-based texture segmenter [16] uses local spectral histograms as features. The local histogram is concatenated from histograms of filter responses computed through convolution with a chosen Gabor and Laplacian of Gaussian filter bank over a square window centered in each pixel's location. The method constructs a  $M \times N$  feature matrix using  $M$ -dimensional feature vectors in an  $N$ -pixel image. Based on the fact that each

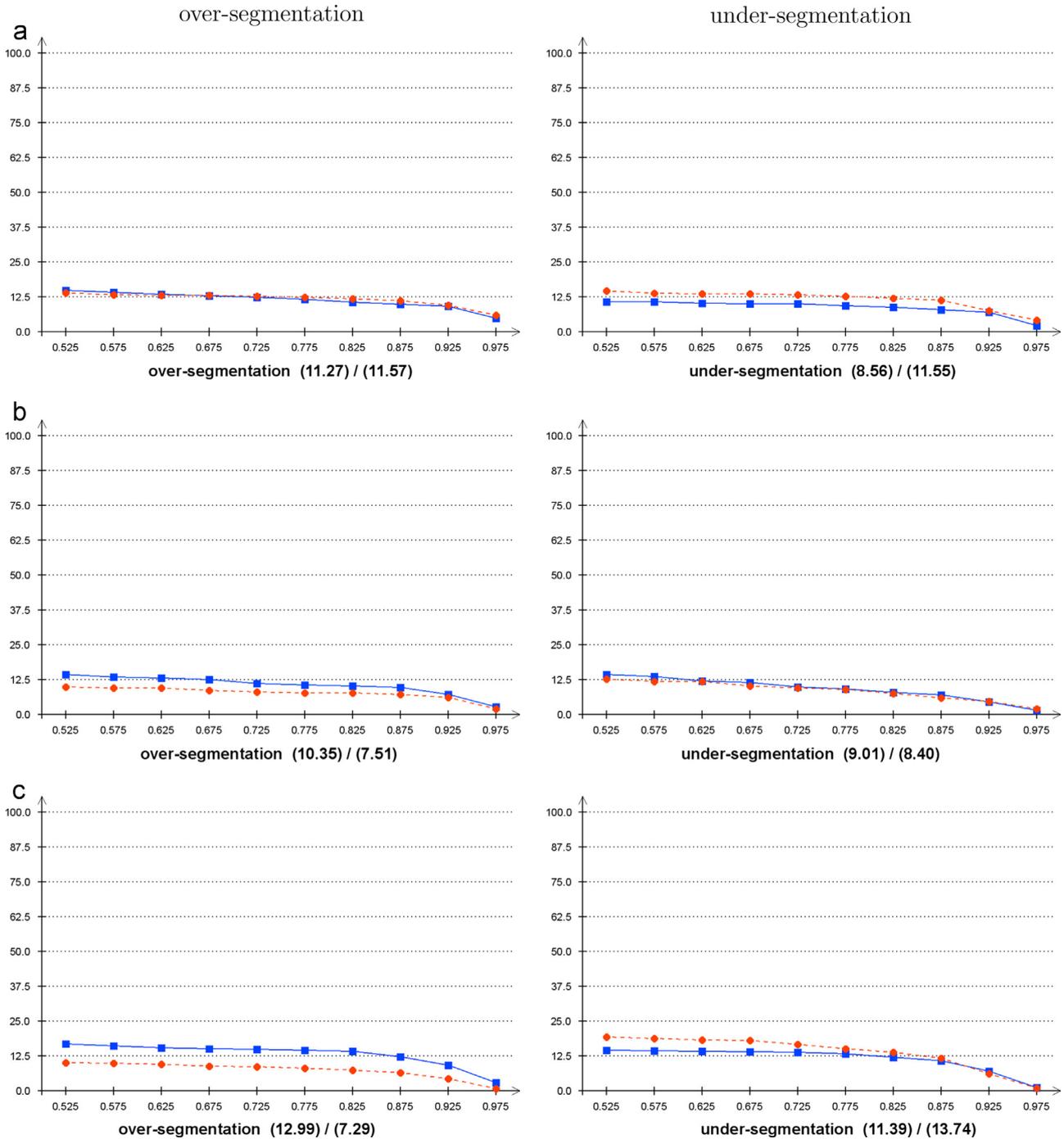


**Fig. 3.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: MW3AR8 (a), Deep Brain Model (b), and CGCHI (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

feature can be approximated by a linear combination of several representative features, this method factorizes the feature matrix into two matrices: one consisting of the representative features, and the other containing weights of the representative features used for a linear combination at each pixel. The factorization method is based on a singular-value decomposition and non-negative matrix factorization. The method uses local spectral histograms to discriminate between region appearances in a computationally efficient way, and at the same time accurately localizes the region boundaries.

### 5.3. MW3AR8

This unsupervised multi-spectral, multi-resolution, multiple-segmenter for textured images with an unknown number of classes is based on [26]. The segmenter utilizes a weighted combination of several unsupervised segmentation results, each in a different resolution, using the modified sum rule. Multi-spectral textured image mosaics are locally represented by eight causal directional multi-spectral random field models, recursively evaluated for each pixel. Each single local texture model is expressed



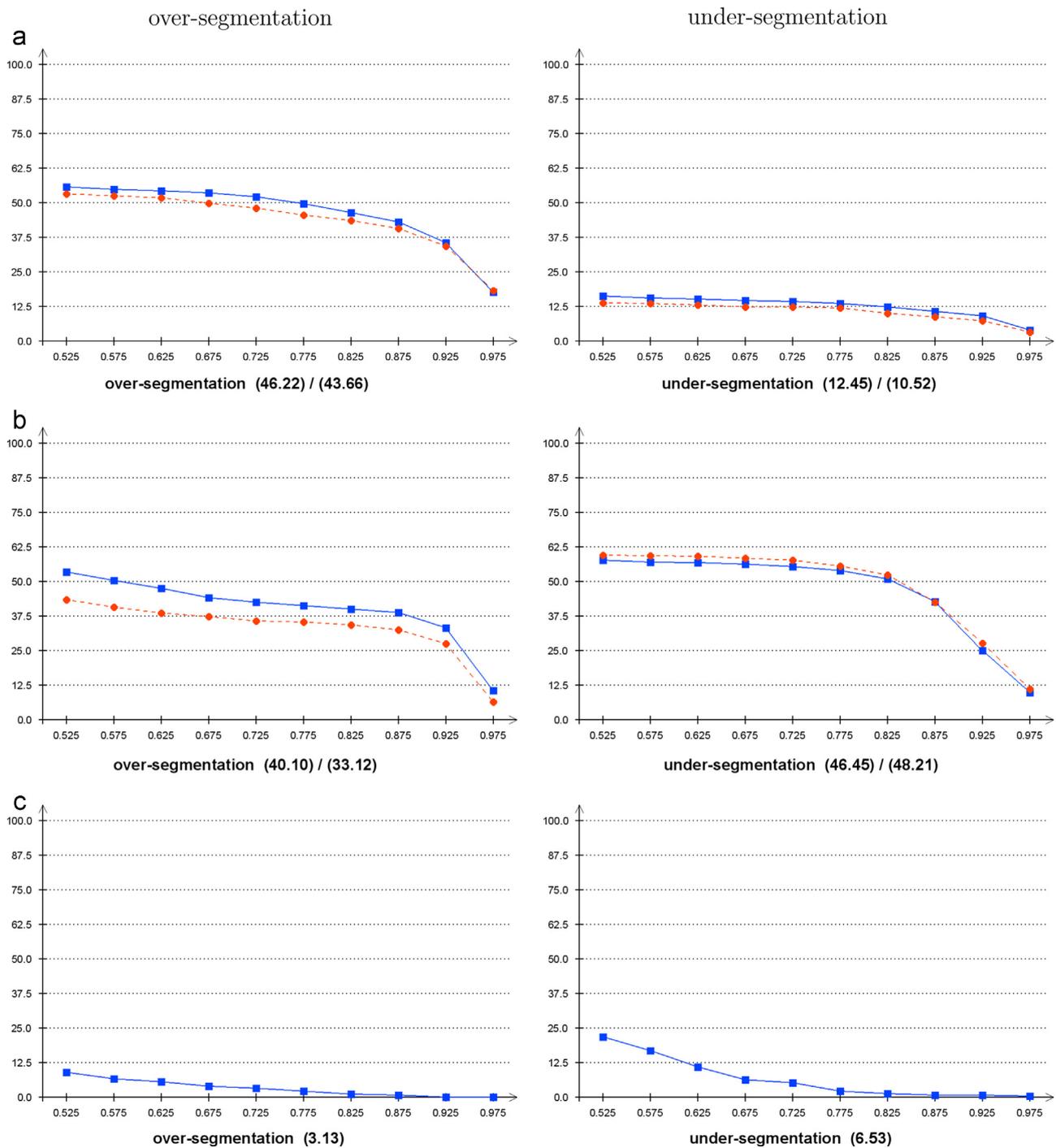
**Fig. 4.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: VRA-PMCFA (a), texNCUT (b), FSEG (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

as a stationary causal uncorrelated noise-driven 3D autoregressive process [27]:  $Y_r = \gamma X_r + e_r$ , where  $\gamma = [A_1, \dots, A_\eta]$  is the parameter matrix,  $r = [r_1, r_2]$  is the regular lattice multiindex,  $I_r^c$  is a causal neighborhood index set with  $\eta = \text{card}(I_r^c)$  and  $e_r$  is a Gaussian white noise vector with zero mean and a constant but unknown covariance, and  $X_r$  is the corresponding vector of the contextual neighbors  $Y_{r-s}$ . The single-resolution segmentation part of the algorithm is based on the underlying Gaussian mixture model and starts with an over-segmented initial estimation which is adaptively modified until the optimal number of homogeneous texture segments is reached. This method did not officially

participate in the contest because it was developed by the organizers. It was used for comparison only.

#### 5.4. Deep Brain Model

The Deep Brain Model is an unsupervised segmentation framework with an unknown number of classes simulating the deep structure of the primate visual cortex. This model is based on a deep scale space, in which a pool of receptive field models in pre-cortical processing and early vision is applied in each scale to produce feature maps. The graph-based image segmentation [28]



**Fig. 5.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: MW3AR8 (a), Deep Brain Model (b), and CGCHI (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

is then employed to select object boundaries among the edges of super-pixels.

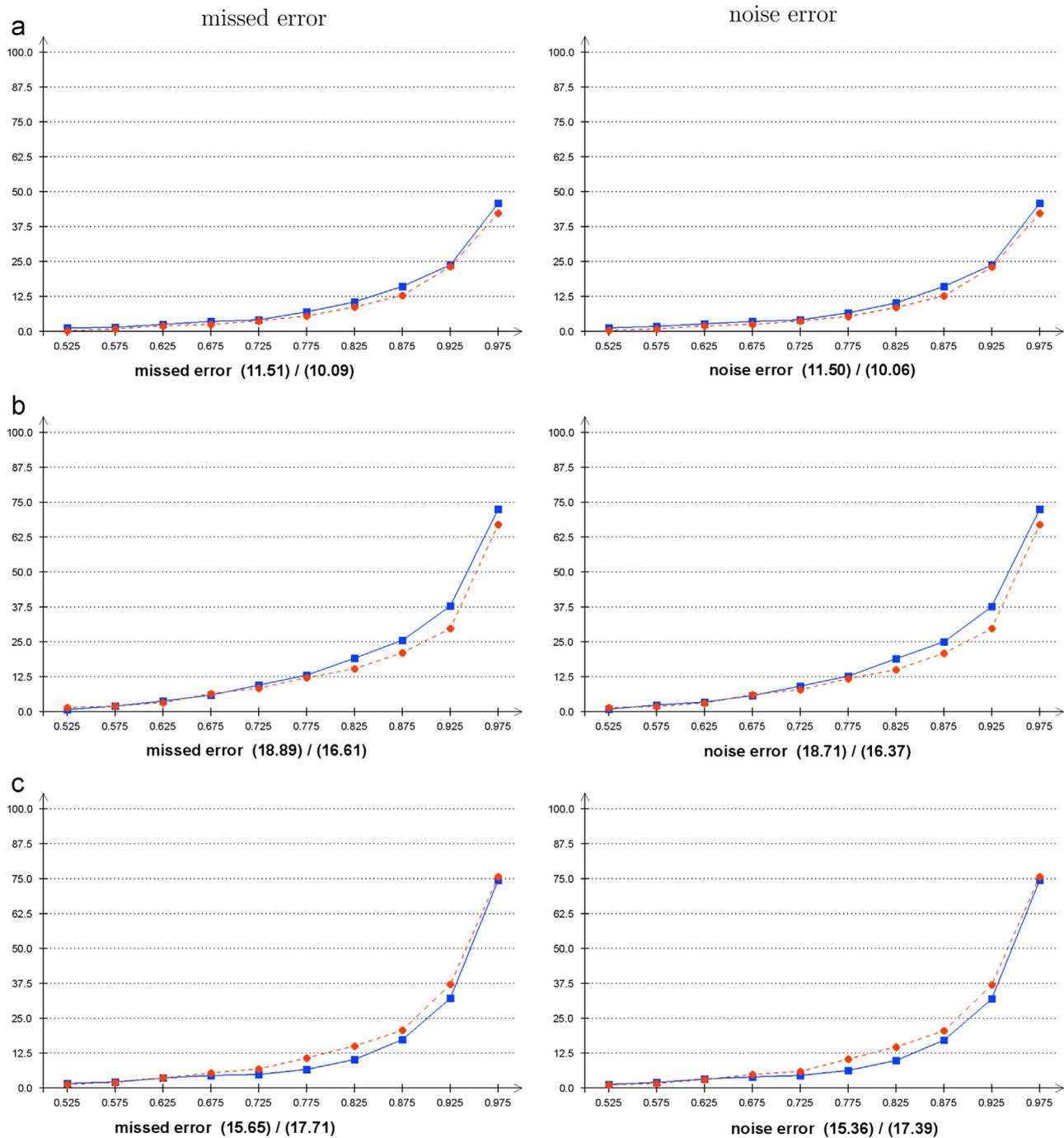
### 5.5. CGCHI

The Combined Graph Cut [9] segmentation with histogram information [29] on regions method is a combination of global and local coherent information. The graph cut method [9] performs the image partitioning via kernel mapping into data of a higher dimension so that the piecewise constant model, and the unsupervised graph cut formulation thereof, becomes applicable. It finds a sufficient number of clusters by using histograms and

probability theory. Subsequently, this method uses metric-space strategies to model local intensity features of the input image. Some problems in this method are from two main sources: the wrong estimation of the number of clusters, and modeling method failures. This method was excluded from the contest due to a missing validation code.

### 5.6. texNCUT

The texNCUT submission is a modification of the normalized cut method [4], which uses textural features based on super-pixels [30]. The number of regions for the evaluated partitions was



**Fig. 6.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: VRA-PMCFa (a), texNCUT (b), FSEG (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

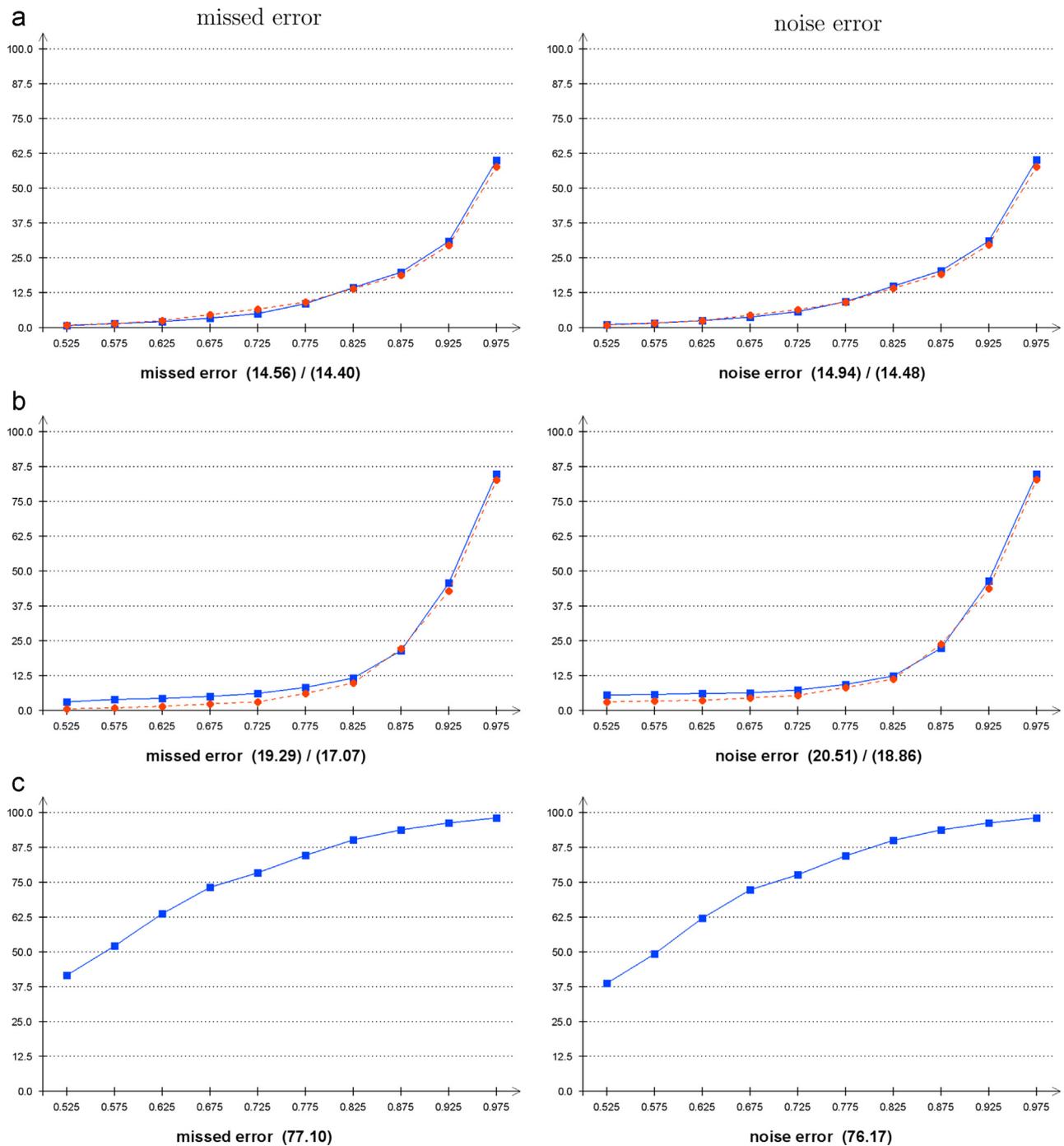
manually set to the number of regions in the ground-truth partitions, which violated the contest conditions. This method was therefore excluded from the final set of contestants.

## 6. Results' evaluation

The selected (12 of 80) test images for visual comparison of the top six methods (VRA-PMCFa, texNCUT, FSEG, MW3AR8, Deep Brain Model, CGCHI) submitted to the contest are shown in Figs. 8 and 9. Nevertheless, the main benefit of the benchmark is the numerical performance criteria evaluated for each tested method. Integrated numerical results of these six methods are

shown in Tables 2 and 3, where  $\uparrow$ –/– $\downarrow$  denotes the required criterion direction and numbers in bold face marks the best criterion value achieved among all six compared methods.

It shows a qualitative gap between the VRA-PMCFa method and the remaining ones on the submission set. It has the average rank difference of 18% compared with the second texNCUT method. The VRA-PMCFa method scores best in all except five criteria (*OS*, *US*, *C*, *RM*, *dVI*). This method is very robust and excellent in the correct segmentation criterion (both in its average *CS* and integral  $\overline{CS}$  forms), which is demonstrated by the flat curve (Fig. 2) of this criterion. However, the method is sensitive to optimal parameter setting. It was obviously carefully tuned for the submission where it received 20 winning contest criteria, but it



**Fig. 7.** Benchmark criteria curves (blue line – submitted, red dashed line – validation) for the methods: MW3AR8 (a), Deep Brain Model (b), and CGCHI (c), respectively. (For interpretation of the references to color in this figure caption, the reader is referred to the web version of this paper.)

decreases to ten on the validation set and also the average rank difference from the texNCUT method drops to only 3%. The under-segmentation error worsens by 3%, and similarly the precision and omission errors, which take over the former commission error pixels. The segmentation error of VRA-PMCFSA increases as the number of regions and their size variability become higher. Naturally, the second method, texNCUT, thus gets improved from 11% to 48%. Both methods accumulate nearly all winning criteria, as can be seen in Tables 2 and 3. VRA-PMCFSA performs worse on the over- and under-segmentation (*OS*, *US*) criteria but these erroneous tendencies are mutually balanced and uniform over the whole threshold range (Fig. 4). The precision measure (*CC*) dropped to the third rank on the validation set, which suggests some

loss of region details. The performance decrease between both sets is between 1% and 3% for single measures.

The second method, texNCUT, performs well except for its region border localization, which is rather poor, as can be seen in Figs. 8d–11d. It is also indicated by numerical results of the global and local consistency error criteria (see *LCE* and *GCE* rows in Tables 2 and 3). The lowest ranking criteria for texNCUT are the *ME* and *NE* error measures. This fact suggests possible improvement using a connected component approach to eliminate small area outliers.

The third method, FSEG, performs solidly on all criteria (ranked between 2 and 4) but does not win in any specific criterion. It has the worst validation performance for the local consistency error

**Table 2**

Color benchmark results (submitted) for VRA-PMCFA, texNCUT, FSEG, MW3AR8, Deep Brain Model, and CGCHI methods.

| critierion        | VRA-PMCFA (1.33)          | texNCUT (2.38)            | FSEG (3.05)        | MW3AR8 (3.90)      | Deep Brain Model (5.10) | CGCHI (5.24)             |
|-------------------|---------------------------|---------------------------|--------------------|--------------------|-------------------------|--------------------------|
| ↑CS               | <b>75.14</b> <sup>1</sup> | 72.54 <sup>2</sup>        | 69.18 <sup>3</sup> | 53.66 <sup>4</sup> | 36.54 <sup>5</sup>      | 10.95 <sup>6</sup>       |
| ↓OS               | 12.13 <sup>3</sup>        | 10.92 <sup>2</sup>        | 14.69 <sup>4</sup> | 51.40 <sup>6</sup> | 41.63 <sup>5</sup>      | <b>2.19</b> <sup>1</sup> |
| ↓US               | 9.85 <sup>3</sup>         | 9.61 <sup>2</sup>         | 13.64 <sup>4</sup> | 14.21 <sup>5</sup> | 55.02 <sup>6</sup>      | <b>3.96</b> <sup>1</sup> |
| ↓ME               | <b>4.38</b> <sup>1</sup>  | 10.25 <sup>5</sup>        | 5.13 <sup>2</sup>  | 5.54 <sup>3</sup>  | 6.71 <sup>4</sup>       | 81.91 <sup>6</sup>       |
| ↓NE               | <b>4.37</b> <sup>1</sup>  | 9.83 <sup>5</sup>         | 4.62 <sup>2</sup>  | 6.33 <sup>3</sup>  | 7.87 <sup>4</sup>       | 81.39 <sup>6</sup>       |
| ↓O                | <b>4.51</b> <sup>1</sup>  | 7.33 <sup>2</sup>         | 9.18 <sup>3</sup>  | 19.86 <sup>4</sup> | 47.36 <sup>5</sup>      | 59.33 <sup>6</sup>       |
| ↓C                | 8.89 <sup>2</sup>         | <b>8.17</b> <sup>1</sup>  | 12.54 <sup>3</sup> | 84.27 <sup>5</sup> | 99.63 <sup>6</sup>      | 51.77 <sup>4</sup>       |
| ↑CA               | <b>83.45</b> <sup>1</sup> | 80.58 <sup>2</sup>        | 78.23 <sup>3</sup> | 70.15 <sup>4</sup> | 49.82 <sup>5</sup>      | 35.62 <sup>6</sup>       |
| ↑CO               | <b>88.12</b> <sup>1</sup> | 86.89 <sup>2</sup>        | 84.45 <sup>3</sup> | 75.41 <sup>4</sup> | 62.63 <sup>5</sup>      | 50.50 <sup>6</sup>       |
| ↑CC               | <b>90.73</b> <sup>1</sup> | 88.28 <sup>3</sup>        | 87.38 <sup>4</sup> | 89.36 <sup>2</sup> | 70.34 <sup>5</sup>      | 49.27 <sup>6</sup>       |
| ↓I.               | <b>11.88</b> <sup>1</sup> | 13.11 <sup>2</sup>        | 15.55 <sup>3</sup> | 24.59 <sup>4</sup> | 37.37 <sup>5</sup>      | 49.50 <sup>6</sup>       |
| ↓II.              | <b>1.48</b> <sup>1</sup>  | 2.36 <sup>2</sup>         | 2.52 <sup>3</sup>  | 2.63 <sup>4</sup>  | 12.39 <sup>6</sup>      | 10.69 <sup>5</sup>       |
| ↑EA               | <b>88.07</b> <sup>1</sup> | 86.39 <sup>2</sup>        | 84.25 <sup>3</sup> | 77.82 <sup>4</sup> | 56.56 <sup>5</sup>      | 47.04 <sup>6</sup>       |
| ↑MS               | <b>83.92</b> <sup>1</sup> | 80.33 <sup>2</sup>        | 78.83 <sup>3</sup> | 70.25 <sup>4</sup> | 46.01 <sup>5</sup>      | 26.89 <sup>6</sup>       |
| ↓RM               | 3.75 <sup>2</sup>         | <b>3.69</b> <sup>1</sup>  | 4.73 <sup>4</sup>  | 3.77 <sup>3</sup>  | 5.38 <sup>5</sup>       | 10.28 <sup>6</sup>       |
| ↑CI               | <b>88.72</b> <sup>1</sup> | 86.97 <sup>2</sup>        | 85.04 <sup>3</sup> | 79.67 <sup>4</sup> | 59.27 <sup>5</sup>      | 48.39 <sup>6</sup>       |
| ↓GCE              | <b>6.55</b> <sup>1</sup>  | 11.92 <sup>4</sup>        | 9.34 <sup>2</sup>  | 9.58 <sup>3</sup>  | 13.03 <sup>5</sup>      | 42.35 <sup>6</sup>       |
| ↓LCE              | <b>3.90</b> <sup>1</sup>  | 6.85 <sup>4</sup>         | 6.08 <sup>3</sup>  | 5.07 <sup>2</sup>  | 7.56 <sup>5</sup>       | 38.59 <sup>6</sup>       |
| ↓dD               | <b>7.59</b> <sup>1</sup>  | 9.18 <sup>2</sup>         | 10.01 <sup>3</sup> | 14.15 <sup>4</sup> | 21.44 <sup>5</sup>      | 40.15 <sup>6</sup>       |
| ↓dM               | <b>4.76</b> <sup>1</sup>  | 6.03 <sup>2</sup>         | 6.99 <sup>3</sup>  | 10.00 <sup>4</sup> | 25.35 <sup>6</sup>      | 25.32 <sup>5</sup>       |
| ↓dVI              | 14.22 <sup>2</sup>        | <b>14.19</b> <sup>1</sup> | 14.33 <sup>3</sup> | 15.90 <sup>6</sup> | 15.49 <sup>5</sup>      | 14.47 <sup>4</sup>       |
| ↑ $\overline{CS}$ | <b>71.77</b>              | 66.14                     | 63.58              | 50.71              | 33.18                   | 11.96                    |
| ↓ $\overline{OS}$ | 11.27                     | 10.35                     | 12.99              | 46.22              | 40.10                   | <b>3.13</b>              |
| ↓ $\overline{US}$ | 8.56                      | 9.01                      | 11.39              | 12.45              | 46.45                   | <b>6.53</b>              |
| ↓ $\overline{ME}$ | <b>11.51</b>              | 18.89                     | 15.65              | 14.56              | 19.29                   | 77.10                    |
| ↓ $\overline{NE}$ | <b>11.50</b>              | 18.71                     | 15.36              | 14.94              | 20.51                   | 76.17                    |
| ↑ $\overline{F}$  | <b>88.54</b>              | 86.81                     | 84.82              | 79.15              | 60.90                   | 48.01                    |

Benchmark criteria: CS = correct segmentation; OS = over-segmentation; US = under-segmentation; ME = missed error; NE = noise error; O = omission error; C = commission error; CA = class accuracy; CO = recall – correct assignment; CC = precision – object accuracy; I = type I error; II = type II error; EA = mean class accuracy estimate; MS = mapping score; RM = root mean square proportion estimation error; CI = comparison index; GCE = Global Consistency Error; LCE = Local Consistency Error; dD = Van Dongen metric; dM = Mirkin metric; dVI = variation of information;  $\overline{f}$  are the performance curve integrals;  $\overline{F}$  = F-measure curve; small numbers are the corresponding measure rank values over the listed methods; numbers in the method's panel are the average ranks.

(LCE) out of all compared methods, which suggests poor local refinement of the reference ground truth.

The fourth method, MW3AR8, is slightly worse than the third one but has very good precision capability (CC). Its current version has a large commission error and strong over-segmentation tendency (OS), which is also demonstrated in Fig. 5, and suggests the direction for its possible improvement. Both properties also hold for its simplified version MW3AR in comparison with alternative segmenters (Table 4). The MW3AR8 method slightly improve all benchmark criteria on the validation set except for ME, NE, and dVI.

The Deep Brain Model performs steadily worse than all four validated methods except the variation of information (dVI) measure, which improves from the worst to the best position between both data sets.

The CGCHI method has the worst performance among all of the submitted methods; it stands as the worst method in 19 criteria. CGCHI suffers from poor correct segmentation (CS), as well as huge missed and noise errors. The localized segment borders are mostly distinctively wrong; thus the global and local consistency error criteria (LCE and GCE) are also very high.

**Table 3**

Alternative benchmark results (validation) for VRA-PMCFA, texNCUT, FSEG, MW3AR8, and Deep Brain Model methods.

| critierion        | VRA-PMCFA (1.86)          | texNCUT (2.05)            | FSEG (3.33)        | MW3AR8 (3.43)             | Deep Brain Model (4.33)   |
|-------------------|---------------------------|---------------------------|--------------------|---------------------------|---------------------------|
| ↑CS               | <b>74.13</b> <sup>1</sup> | 73.90 <sup>2</sup>        | 67.44 <sup>3</sup> | 55.01 <sup>4</sup>        | 32.54 <sup>5</sup>        |
| ↓OS               | 12.24 <sup>3</sup>        | <b>7.89</b> <sup>1</sup>  | 8.01 <sup>2</sup>  | 47.19 <sup>5</sup>        | 35.33 <sup>4</sup>        |
| ↓US               | 12.83 <sup>3</sup>        | <b>9.15</b> <sup>1</sup>  | 15.52 <sup>4</sup> | 12.03 <sup>2</sup>        | 56.97 <sup>5</sup>        |
| ↓ME               | 4.69 <sup>2</sup>         | 10.11 <sup>5</sup>        | 9.36 <sup>4</sup>  | 7.78 <sup>3</sup>         | <b>4.07</b> <sup>1</sup>  |
| ↓NE               | <b>4.60</b> <sup>1</sup>  | 9.61 <sup>5</sup>         | 8.75 <sup>4</sup>  | 7.72 <sup>3</sup>         | 6.14 <sup>2</sup>         |
| ↓O                | 7.59 <sup>2</sup>         | <b>6.68</b> <sup>1</sup>  | 9.71 <sup>3</sup>  | 16.18 <sup>4</sup>        | 53.19 <sup>5</sup>        |
| ↓C                | <b>5.70</b> <sup>1</sup>  | 6.35 <sup>2</sup>         | 10.75 <sup>3</sup> | 81.02 <sup>4</sup>        | 96.98 <sup>5</sup>        |
| ↑CA               | <b>82.21</b> <sup>1</sup> | 82.19 <sup>2</sup>        | 76.46 <sup>3</sup> | 72.44 <sup>4</sup>        | 44.69 <sup>5</sup>        |
| ↑CO               | 87.29 <sup>2</sup>        | <b>88.08</b> <sup>1</sup> | 83.90 <sup>3</sup> | 77.53 <sup>4</sup>        | 58.13 <sup>5</sup>        |
| ↑CC               | 88.50 <sup>3</sup>        | 89.46 <sup>2</sup>        | 84.83 <sup>4</sup> | <b>90.34</b> <sup>1</sup> | 68.25 <sup>5</sup>        |
| ↓I.               | 12.71 <sup>2</sup>        | <b>11.92</b> <sup>1</sup> | 16.10 <sup>3</sup> | 22.47 <sup>4</sup>        | 41.87 <sup>5</sup>        |
| ↓II.              | 2.03 <sup>2</sup>         | <b>1.95</b> <sup>1</sup>  | 2.75 <sup>4</sup>  | 2.55 <sup>3</sup>         | 11.08 <sup>5</sup>        |
| ↑EA               | 86.43 <sup>2</sup>        | <b>87.44</b> <sup>1</sup> | 82.23 <sup>3</sup> | 79.93 <sup>4</sup>        | 51.26 <sup>5</sup>        |
| ↑MS               | 82.10 <sup>2</sup>        | <b>82.12</b> <sup>1</sup> | 76.92 <sup>3</sup> | 72.65 <sup>4</sup>        | 39.50 <sup>5</sup>        |
| ↓RM               | 4.35 <sup>3</sup>         | <b>3.33</b> <sup>1</sup>  | 5.07 <sup>4</sup>  | 3.63 <sup>2</sup>         | 5.53 <sup>5</sup>         |
| ↑CI               | 87.10 <sup>2</sup>        | <b>88.05</b> <sup>1</sup> | 83.18 <sup>3</sup> | 81.49 <sup>4</sup>        | 54.60 <sup>5</sup>        |
| ↓GCE              | <b>6.09</b> <sup>1</sup>  | 10.92 <sup>4</sup>        | 10.20 <sup>3</sup> | 8.60 <sup>2</sup>         | 11.38 <sup>5</sup>        |
| ↓LCE              | <b>3.45</b> <sup>1</sup>  | 6.38 <sup>4</sup>         | 6.55 <sup>5</sup>  | 5.05 <sup>2</sup>         | 5.42 <sup>3</sup>         |
| ↓dD               | <b>7.79</b> <sup>1</sup>  | 8.47 <sup>2</sup>         | 10.51 <sup>3</sup> | 13.10 <sup>4</sup>        | 22.51 <sup>5</sup>        |
| ↓dM               | 5.51 <sup>2</sup>         | <b>5.03</b> <sup>1</sup>  | 6.96 <sup>3</sup>  | 8.48 <sup>4</sup>         | 29.21 <sup>5</sup>        |
| ↓dVI              | 14.32 <sup>2</sup>        | 14.49 <sup>4</sup>        | 14.36 <sup>3</sup> | 16.06 <sup>5</sup>        | <b>13.39</b> <sup>1</sup> |
| ↑ $\overline{CS}$ | <b>71.78</b>              | 69.77                     | 62.69              | 53.06                     | 28.78                     |
| ↓ $\overline{OS}$ | 11.57                     | 7.51                      | <b>7.29</b>        | 43.66                     | 33.12                     |
| ↓ $\overline{US}$ | 11.55                     | <b>8.40</b>               | 13.74              | 10.52                     | 48.21                     |
| ↓ $\overline{ME}$ | <b>10.09</b>              | 16.61                     | 17.71              | 14.40                     | 17.07                     |
| ↓ $\overline{NE}$ | <b>10.06</b>              | 16.37                     | 17.39              | 14.48                     | 18.86                     |
| ↑ $\overline{F}$  | 86.91                     | <b>87.88</b>              | 82.91              | 81.06                     | 53.67                     |

Benchmark criteria: CS = correct segmentation; OS = over-segmentation; US = under-segmentation; ME = missed error; NE = noise error; O = omission error; C = commission error; CA = class accuracy; CO = recall – correct assignment; CC = precision – object accuracy; I = type I error; II = type II error; EA = mean class accuracy estimate; MS = mapping score; RM = root mean square proportion estimation error; CI = comparison index; GCE = Global Consistency Error; LCE = Local Consistency Error; dD = Van Dongen metric; dM = Mirkin metric; dVI = variation of information;  $\overline{f}$  are the performance curve integrals;  $\overline{F}$  = F-measure curve; small numbers are the corresponding measure rank values over the listed methods; numbers in the method's panel are the average ranks.

The evaluated methods can be clustered into three qualitative performance groups – the best (VRA-PMCFA, texNCUT), medium (FSEG, MW3AR8), and the worst (Deep Brain Model, CGCHI).

A more detailed insight into the behavior of single methods can be obtained by consulting the corresponding criteria description [19] and their achieved values shown in Tables 2 and 3.

### 6.1. Alternative segmenters

Table 4 shows numerical results of other well-known, published unsupervised segmenters, sorted from left to right according to the average rank results. This table contains the following 11 methods:

MW3AR is an unsupervised multi-spectral, multi-resolution, multiple-segmenter [26] for textured images with an unknown number of classes. The segmenter is based on a weighted combination of several unsupervised segmentation results, each in a different resolution, using the modified sum rule.

The TEX-ROI-SEG [31] segmenter combines Bhattacharyya distances and a modified version of the maximally stable extremal region detector.

GSRM [32] is a region-merging technique based on a size-weighted/unweighted direct statistical measure of the empirical distributions of the regions, using the Kullback–Leibler divergence or Bhattacharyya coefficient.



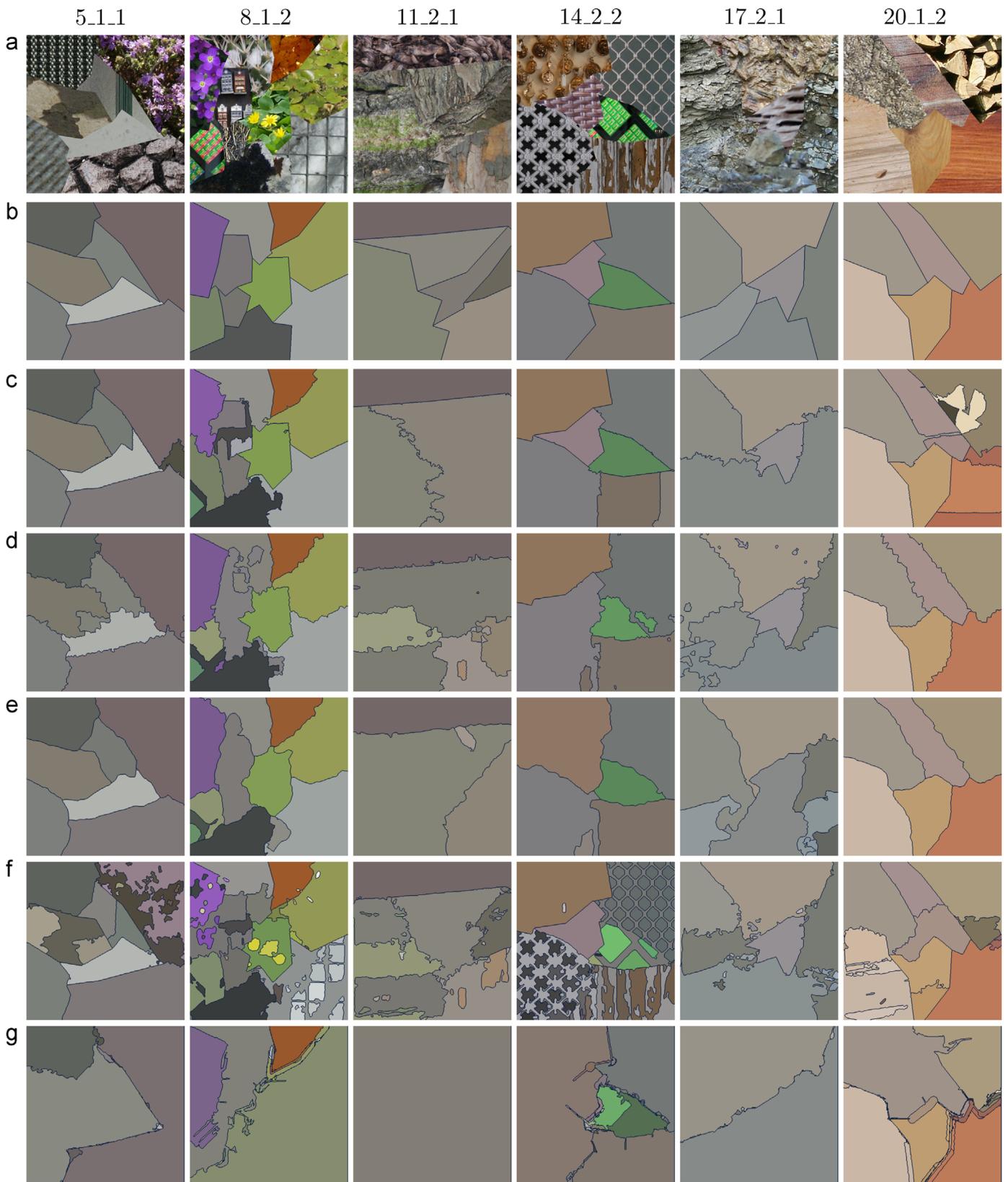
**Fig. 8.** Selected color benchmark mosaics (a), ground-truth (b), VRA-PMCF (c), texNCUT (d), FSEG (e), MW3AR8 (f), Deep Brain Model (g), and CGCHI (h) submitted segmentation results, respectively.



**Fig. 9.** Selected color benchmark mosaics (a), ground-truth (b), VRA-PMCF A (c), texNCUT (d), FSEG (e), MW3AR8 (f), Deep Brain Model (g), and CGCHI (h) submitted segmentation results, respectively.



**Fig. 10.** Selected validation benchmark mosaics (a), ground-truth (b), VRA-PMCF A (c), texNCUT (d), FSEG (e), MW3AR8 (f), and Deep Brain Model (g) segmentation results, respectively.



**Fig. 11.** Selected validation benchmark mosaics (a), ground-truth (b), VRA-PMCF A (c), texNCUT (d), FSEG (e), MW3AR8 (f), and Deep Brain Model (g) segmentation results, respectively.

**Table 4**

Color benchmark results for MW3AR, TEX-ROI-SEG, GSRM unsup. (KL sw), TBES, TFR/KLD, SWA, EDISON, JSEG, EGBIS, HGS (E), and Blobworld methods.

| critereon | MW3AR<br>(2.90)           | TEX-ROI-SEG<br>(3.48)     | GSRM unsup.<br>(4.43) | TBES<br>(4.57)      | TFR/KLD<br>(5.10)        | SWA<br>(6.90)      | EDISON<br>(7.10)            | JSEG<br>(7.33)      | EGBIS<br>(7.38)     | HGS<br>(7.86)             | Blobworld<br>(8.95) |
|-----------|---------------------------|---------------------------|-----------------------|---------------------|--------------------------|--------------------|-----------------------------|---------------------|---------------------|---------------------------|---------------------|
| ↑CS       | 53.04 <sup>3</sup>        | <b>56.37</b> <sup>1</sup> | 54.05 <sup>2</sup>    | 37.72 <sup>5</sup>  | 51.25 <sup>4</sup>       | 27.06 <sup>9</sup> | 12.68 <sup>11</sup>         | 27.47 <sup>8</sup>  | 28.78 <sup>7</sup>  | 29.81 <sup>6</sup>        | 21.01 <sup>10</sup> |
| ↓OS       | 59.53 <sup>9</sup>        | 11.93 <sup>4</sup>        | 15.88 <sup>5</sup>    | 59.77 <sup>10</sup> | <b>5.84</b> <sup>1</sup> | 50.21 <sup>8</sup> | 86.91 <sup>11</sup>         | 38.62 <sup>7</sup>  | 19.69 <sup>6</sup>  | 10.69 <sup>3</sup>        | 7.33 <sup>2</sup>   |
| ↓US       | 3.20 <sup>3</sup>         | 19.79 <sup>9</sup>        | 14.45 <sup>8</sup>    | 1.25 <sup>2</sup>   | 7.16 <sup>6</sup>        | 4.53 <sup>4</sup>  | <b>0.00</b> <sup>1</sup>    | 5.04 <sup>5</sup>   | 39.15 <sup>11</sup> | 33.76 <sup>10</sup>       | 9.30 <sup>7</sup>   |
| ↓ME       | 5.63 <sup>2</sup>         | 11.55 <sup>4</sup>        | 13.62 <sup>5</sup>    | 7.24 <sup>3</sup>   | 31.64 <sup>9</sup>       | 25.76 <sup>7</sup> | <b>2.48</b> <sup>1</sup>    | 35.00 <sup>10</sup> | 20.42 <sup>6</sup>  | 26.89 <sup>8</sup>        | 59.55 <sup>11</sup> |
| ↓NE       | 6.96 <sup>2</sup>         | 10.29 <sup>4</sup>        | 14.03 <sup>5</sup>    | 8.22 <sup>3</sup>   | 31.38 <sup>9</sup>       | 27.50 <sup>8</sup> | <b>4.68</b> <sup>1</sup>    | 35.50 <sup>10</sup> | 21.54 <sup>6</sup>  | 25.04 <sup>7</sup>        | 61.68 <sup>11</sup> |
| ↓O        | 19.32 <sup>2</sup>        | <b>18.21</b> <sup>1</sup> | 23.44 <sup>4</sup>    | 27.55 <sup>5</sup>  | 19.65 <sup>3</sup>       | 33.01 <sup>6</sup> | 73.17 <sup>11</sup>         | 37.94 <sup>7</sup>  | 44.35 <sup>9</sup>  | 48.94 <sup>10</sup>       | 41.45 <sup>8</sup>  |
| ↓C        | 86.19 <sup>8</sup>        | <b>9.63</b> <sup>1</sup>  | 26.47 <sup>3</sup>    | 95.08 <sup>10</sup> | 9.67 <sup>2</sup>        | 85.19 <sup>7</sup> | <b>100.00</b> <sup>11</sup> | 92.77 <sup>9</sup>  | 82.87 <sup>6</sup>  | 32.39 <sup>4</sup>        | 58.94 <sup>5</sup>  |
| ↑CA       | <b>71.89</b> <sup>1</sup> | 69.45 <sup>2</sup>        | 66.75 <sup>4</sup>    | 63.30 <sup>5</sup>  | 67.45 <sup>3</sup>       | 54.84 <sup>7</sup> | 31.19 <sup>11</sup>         | 55.29 <sup>6</sup>  | 51.10 <sup>8</sup>  | 49.60 <sup>9</sup>        | 46.23 <sup>10</sup> |
| ↑CO       | 74.66 <sup>3</sup>        | <b>78.26</b> <sup>1</sup> | 73.67 <sup>4</sup>    | 65.66 <sup>5</sup>  | 76.40 <sup>2</sup>       | 60.67 <sup>9</sup> | 31.55 <sup>11</sup>         | 61.81 <sup>8</sup>  | 64.12 <sup>6</sup>  | 63.37 <sup>7</sup>        | 56.04 <sup>10</sup> |
| ↑CC       | 95.04 <sup>3</sup>        | 81.24 <sup>7</sup>        | 81.41 <sup>6</sup>    | 96.20 <sup>2</sup>  | 81.12 <sup>8</sup>       | 88.17 <sup>4</sup> | <b>98.09</b> <sup>1</sup>   | 87.70 <sup>5</sup>  | 72.73 <sup>10</sup> | 66.09 <sup>11</sup>       | 73.62 <sup>9</sup>  |
| ↓I        | 25.34 <sup>3</sup>        | <b>21.74</b> <sup>1</sup> | 26.33 <sup>4</sup>    | 34.34 <sup>5</sup>  | 23.60 <sup>2</sup>       | 39.33 <sup>9</sup> | 68.45 <sup>11</sup>         | 38.19 <sup>8</sup>  | 35.88 <sup>6</sup>  | 36.63 <sup>7</sup>        | 43.96 <sup>10</sup> |
| ↓II       | 0.74 <sup>3</sup>         | 4.16 <sup>8</sup>         | 3.35 <sup>5</sup>     | 0.60 <sup>2</sup>   | 4.09 <sup>7</sup>        | 2.11 <sup>4</sup>  | <b>0.24</b> <sup>1</sup>    | 3.66 <sup>6</sup>   | 7.59 <sup>10</sup>  | 13.51 <sup>11</sup>       | 6.72 <sup>9</sup>   |
| ↑EA       | <b>80.43</b> <sup>1</sup> | 76.31 <sup>2</sup>        | 73.14 <sup>5</sup>    | 74.19 <sup>4</sup>  | 75.80 <sup>3</sup>       | 66.94 <sup>6</sup> | 41.29 <sup>11</sup>         | 66.74 <sup>7</sup>  | 59.88 <sup>8</sup>  | 58.74 <sup>9</sup>        | 58.37 <sup>10</sup> |
| ↑MS       | <b>71.78</b> <sup>1</sup> | 68.88 <sup>2</sup>        | 65.75 <sup>3</sup>    | 63.73 <sup>5</sup>  | 65.19 <sup>4</sup>       | 53.71 <sup>7</sup> | 31.13 <sup>11</sup>         | 55.14 <sup>6</sup>  | 49.03 <sup>8</sup>  | 46.63 <sup>9</sup>        | 40.36 <sup>10</sup> |
| ↓RM       | <b>3.09</b> <sup>1</sup>  | 7.37 <sup>8</sup>         | 6.76 <sup>6</sup>     | 5.27 <sup>4</sup>   | 7.21 <sup>7</sup>        | 6.11 <sup>5</sup>  | 3.21 <sup>2</sup>           | 4.96 <sup>3</sup>   | 8.38 <sup>10</sup>  | 13.31 <sup>11</sup>       | 7.96 <sup>9</sup>   |
| ↑CI       | <b>82.43</b> <sup>1</sup> | 77.86 <sup>2</sup>        | 75.09 <sup>5</sup>    | 77.29 <sup>3</sup>  | 77.21 <sup>4</sup>       | 70.32 <sup>6</sup> | 50.29 <sup>11</sup>         | 70.27 <sup>7</sup>  | 63.11 <sup>8</sup>  | 61.17 <sup>10</sup>       | 61.31 <sup>9</sup>  |
| ↓GCE      | 8.17 <sup>3</sup>         | 11.98 <sup>5</sup>        | 8.87 <sup>4</sup>     | 7.57 <sup>2</sup>   | 20.35 <sup>10</sup>      | 17.27 <sup>8</sup> | <b>3.55</b> <sup>1</sup>    | 18.45 <sup>9</sup>  | 16.64 <sup>6</sup>  | 16.75 <sup>7</sup>        | 31.16 <sup>11</sup> |
| ↓LCE      | 5.78 <sup>2</sup>         | 6.71 <sup>5</sup>         | 5.98 <sup>3</sup>     | 6.02 <sup>4</sup>   | 14.36 <sup>10</sup>      | 11.49 <sup>8</sup> | <b>3.44</b> <sup>1</sup>    | 11.64 <sup>9</sup>  | 8.97 <sup>6</sup>   | 10.46 <sup>7</sup>        | 23.19 <sup>11</sup> |
| ↓dD       | 14.78 <sup>2</sup>        | <b>13.66</b> <sup>1</sup> | 15.95 <sup>3</sup>    | 19.40 <sup>5</sup>  | 18.01 <sup>4</sup>       | 24.20 <sup>9</sup> | 35.37 <sup>11</sup>         | 23.38 <sup>8</sup>  | 21.29 <sup>6</sup>  | 22.90 <sup>7</sup>        | 31.11 <sup>10</sup> |
| ↓dM       | <b>8.97</b> <sup>1</sup>  | 11.74 <sup>3</sup>        | 12.36 <sup>4</sup>    | 10.93 <sup>2</sup>  | 12.64 <sup>5</sup>       | 13.68 <sup>6</sup> | 16.84 <sup>8</sup>          | 15.19 <sup>7</sup>  | 19.72 <sup>9</sup>  | 27.95 <sup>11</sup>       | 20.03 <sup>10</sup> |
| ↓dVI      | 16.67 <sup>7</sup>        | 13.74 <sup>2</sup>        | 14.87 <sup>5</sup>    | 17.76 <sup>10</sup> | 14.06 <sup>4</sup>       | 17.16 <sup>8</sup> | 25.65 <sup>11</sup>         | 17.37 <sup>9</sup>  | 13.79 <sup>3</sup>  | <b>12.83</b> <sup>1</sup> | 15.84 <sup>6</sup>  |
| ↑CS̄      | 49.60                     | <b>52.98</b>              | 51.95                 | 36.49               | 47.58                    | 26.42              | 12.95                       | 29.13               | 30.69               | 27.82                     | 19.10               |
| ↓OS̄      | 51.08                     | 11.54                     | 14.66                 | 51.90               | <b>5.27</b>              | 44.49              | 76.33                       | 37.70               | 19.86               | 9.70                      | 10.81               |
| ↓US̄      | 2.93                      | 17.94                     | 14.06                 | 0.85                | 7.11                     | 5.26               | <b>0.00</b>                 | 6.38                | 33.66               | 31.62                     | 8.35                |
| ↓MĒ      | 16.05                     | 19.19                     | 18.19                 | 17.54               | 37.14                    | 33.36              | <b>13.92</b>                | 34.72               | 28.07               | 32.86                     | 58.54               |
| ↓NĒ      | 16.87                     | 18.68                     | 18.36                 | 18.02               | 37.29                    | 33.72              | <b>15.30</b>                | 35.38               | 28.74               | 32.47                     | 61.24               |
| ↓F̄       | <b>81.85</b>              | 77.41                     | 74.52                 | 76.41               | 76.81                    | 69.35              | 47.42                       | 69.23               | 62.12               | 60.51                     | 60.46               |

Benchmark criteria: CS = correct segmentation; OS = over-segmentation; US = under-segmentation; ME = missed error; NE = noise error; O = omission error; C = commission error; CA = class accuracy; CO = recall - correct assignment; CC = precision - object accuracy; I = type I error; II = type II error; EA = mean class accuracy estimate; MS = mapping score; RM = root mean square proportion estimation error; CI = comparison index; GCE = Global Consistency Error; LCE = Local Consistency Error; dD = Van Dongen metric; dM = Mirkin metric; dVI = variation of information;  $\bar{f}$  are the performance curve integrals;  $\bar{F}$  = F-measure curve; small numbers are the corresponding measure rank values over the listed methods; numbers in the method's panel are the average ranks.

TBES [12] is an agglomerative clustering process applied to a hierarchy of decreasing window sizes as multi-scale texture features, where the region boundary is coded by an adaptive chain code. The segmentation is performed by minimizing the coding length.

The Texture Fragmentation and Reconstruction segmentation algorithm TFR/KLD [33] is based on a texture modeling where a texture is, for each fixed spatial direction, regarded as a finite-state Markov chain whose states are quantized colors. A simple segmentation algorithm independently processes color and spatial information, by first performing a color-based clustering, which provides the quantized colors, and then by means of a further spatial-based clustering, which separates regions according to their transition probability profiles. Finally, a region-merging algorithm allows us to recover the different textures, that is, to recompose their internal Markov chains.

The SWA [34] algorithm segmentation by weighted aggregation, is derived from algebraic multigrid solvers for physical systems, and consists of fine-to-coarse pixel aggregation. The SWA algorithm starts with a weighted graph representing image pixels, and subsequently creates a hierarchy of coarser and smaller graphs. The edge weights are determined by inheritance from previous levels and regional properties' modification.

Edison [35] is a low-level feature extraction tool that integrates confidence based edge detection and the mean shift based image segmentation.

The EGBIS [28] is a graph-based segmenter using a predicate for measuring the evidence for a boundary between two regions.

The JSEG method [36] consists of two independent steps: color quantization, sampled by a reduced set of significant colors, and

region growing spatial segmentation on multi-scale thematic maps from the first step. Finally, a post-processing technique is applied to merge the adjacent regions based on color similarity and the Euclidean distance.

The HGS segmenter [37] combines the K-means clustering with a region-merging step. It uses a Gabor-Gaussian spatial-color texture representation, and its illumination-invariant C version uses features derived from the Gabor filters applied to log-transformed images and reduced by the principal component analysis.

The Blobworld [38] scheme aims to transform images into a small set of regions coherent in color and texture. This goal is achieved by clustering pixels in a joint color-texture-position eight-dimensional feature space using the EM algorithm. The feature vector containing anisotropy, polarity, and contrast features is represented by a Gaussian mixture model.

Although these results in Table 4 cannot be directly compared with Tables 2 and 3 because they were computed on a smaller evaluation set of 20 mosaics with parameters corresponding to Table 1, i.e., the contest submission set also includes these 20 mosaics, they still significantly indicate the contesting algorithms' standing among the state-of-the-art unsupervised image segmenters. The ranks and average ranks are relative to Table 4 but the numerical values of single criteria can be approximately compared. For example, the contest winner achieved a 20% improvement of the correct segmentation result over the best result in (TEX-ROI-SEG) in this group, and the Mirkin metric, omission error, mapping score, or class accuracy are better for the first four methods in Table 3 than the best alternative methods' results (MW3AR/TEX-ROI-SEG), etc.

These methods can also be roughly clustered into three qualitative performance groups—the best (MW3AR, TEX-ROI-SEG), medium (GSRM, TBES, TFR/KLD), and the worst (SWA, Edison, EGBIS, JSEG, HGS, Blobworld).

All details for these summary results, such as individual mosaics performance, single criteria sensitivity curves, or visual inspection of all results, can be studied in the Prague texture segmentation data-generator and benchmark [18–20] web (<http://mosaic.utia.cas.cz>).

## 7. Conclusions

Unusually extensive benchmarking of the contest methods, evaluated on 160(80+80) different textural mosaics, allows us to get a deep and reliable insight into their properties, advantages, and drawbacks. Half of these test images were available to the competing authors to tune up their methods during their development, but the contest ranking was based on another large set of images, run by the organizers using the submitted segmenters. The contest winner, the VRA-PMCF method, is a high quality unsupervised segmenter, which performs reliably on both test sets. The texNCUT method requires the knowledge of the number of regions in each scene, which is an unjust advantage with respect to all remaining methods, thus preventing this method from participating in their fair ranking. The contest segmenters can also be indirectly compared with 11 alternative, previously published segmenters listed in Table 4, as well as additional segmenters presented in the benchmark (<http://mosaic.utia.cas.cz>). These segmenters were evaluated on a smaller test set of only 20 images. Although their criteria values and mutual ranking of neighboring methods might slightly change if run on a larger contest set, it is obvious that none of these method would beat the winning VRA-PMCF method.

The contest results advice the importance of descriptive, multi-spectral, spatial underlying models and sophisticated post-processing for advanced future unsupervised image segmenters. Such segmenters should avoid artificial color and texture separation, and learn their parameters, as well as the number of region classes, directly from the segmented data.

## Conflict of interest

None declared.

## Acknowledgment

This research was supported by the Czech Science Foundation Project no. GAČR 14-10911S.

## References

- [1] Y.J. Zhang, Evaluation and comparison of different segmentation algorithms, *Pattern Recognit. Lett.* 18 (1997) 963–974.
- [2] K. Fu, J. Mui, A survey on image segmentation, *Pattern Recognit.* 13 (1981) 3–16.
- [3] N. Pal, S. Pal, A review on image segmentation techniques, *Pattern Recognit.* 26 (9) (1993) 1277–1294.
- [4] J. Shi, J. Malik, Normalized cuts and image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8) (2000) 888–905, <http://dx.doi.org/10.1109/34.868688>.
- [5] L. Lucchese, S. Mitra, Color image segmentation: a state-of-the-art survey, *Proc. Indian Natl. Sci. Acad. (INSA-A)* 67 (2) (2001) 207–221.
- [6] H. Cheng, X. Jiang, Y. Sun, J. Wang, Color image segmentation: advances and prospects, *Pattern Recognit.* 34 (2001) 2259–2281.
- [7] J. Freixenet, X. Munoz, D. Raba, J. Marti, X. Cufi, Yet another survey on image segmentation: Region and boundary information integration, in: *European Conference on Computer Vision*, 2002, p. III: 408 ff.
- [8] M. Galun, E. Sharon, R. Basri, A. Brandt, Texture segmentation by multiscale aggregation of filter responses and shape elements, in: *International Conference on Computer Vision*, 2003, pp. 716–723. <http://dx.doi.org/10.1109/ICCV.2003.1238418>.
- [9] M.B. Salah, A. Mitiche, I.B. Ayed, Multiregion image segmentation by parametric kernel graph cuts, *IEEE Trans. Image Process.* 20 (2) (2011) 545–557, <http://dx.doi.org/10.1109/TIP.2010.2066982>.
- [10] D.E. Ilea, P.F. Whelan, Image segmentation based on the integration of colour-texture descriptors—a review, *Pattern Recognit.* 44 (10–11) (2011) 2479–2501, <http://dx.doi.org/10.1016/j.patcog.2011.03.005> (Semi-supervised learning for visual content analysis and understanding).
- [11] A. Sengur, Y. Guo, Color texture image segmentation based on neutrosophic set and wavelet transformation, *Comput. Vis. Image Underst.* 115 (8) (2011) 1134–1144, <http://dx.doi.org/10.1016/j.cviu.2011.04.001>.
- [12] H. Mobahi, S.R. Rao, A.Y. Yang, S.S. Sastry, Y. Ma, Segmentation of natural images by texture and boundary compression, *Int. J. Comput. Vis.* 95 (1) (2011) 86–98.
- [13] S. Alpert, M. Galun, A. Brandt, R. Basri, Image segmentation by probabilistic bottom-up aggregation and cue integration, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (2) (2012) 315–327, <http://dx.doi.org/10.1109/TPAMI.2011.130>.
- [14] J. Kim, K. Grauman, Shape sharing for object segmentation, in: *Computer Vision—ECCV 2012*, Springer, Berlin, Heidelberg 2012, pp. 444–458.
- [15] O. Arandjelovic, Colour invariants under a non-linear photometric camera model and their application to face recognition from video, *Pattern Recognit.* 45 (7) (2012) 2499–2509, <http://dx.doi.org/10.1016/j.patcog.2012.01.013>.
- [16] J. Yuan, D. Wang, Factorization-Based Texture Segmentation, Technical Report OSU-CISRC-1/13-TR0, The Ohio State University, Columbus, USA, 2013.
- [17] H. Zhou, J. Zheng, L. Wei, Texture aware image segmentation using graph cuts and active contours, *Pattern Recognit.* 46 (6) (2013) 1719–1733, <http://dx.doi.org/10.1016/j.patcog.2012.12.005>.
- [18] S. Mikeš, M. Haindl, Prague texture segmentation data generator and benchmark, *ERCIM News* 64 (2006) 67–68.
- [19] M. Haindl, S. Mikeš, Texture segmentation benchmark, in: B. Lovell, D. Laurendeau, R. Duin (Eds.), *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008*, IEEE Computer Society, Los Alamitos, 2008, pp. 1–4. <http://dx.doi.org/10.1109/ICPR.2008.4761118>.
- [20] M. Haindl, S. Mikeš, Pattern Recognition; Techniques, Technology and Applications, I-Tech Education and Publishing, Vienna, 2008, pp. 227–248 (Chapter: Unsupervised Texture Segmentation).
- [21] M. Haindl, J. Filip, Visual Texture, *Advances in Computer Vision and Pattern Recognition*, Springer-Verlag, London (2012) <http://dx.doi.org/10.1007/978-1-4471-4902-6>.
- [22] D. Martin, C. Fowlkes, D. Tal, J. Malik, A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics, in: *Proceedings of 8th International Conference on Computer Vision*, vol. 2, 2001, pp. 416–423.
- [23] A. Hoover, G. Jean-Baptiste, X. Jiang, P.J. Flynn, H. Bunke, D.B. Goldgof, K. Bowyer, D.W. Eggert, A. Fitzgibbon, R.B. Fisher, An experimental comparison of range image segmentation algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 18 (7) (1996) 673–689.
- [24] M. Meila, Comparing clusterings—an axiomatic view, in: *ICML: Machine Learning: Proceedings of the Seventh International Conference*, 2005, pp. 577–584.
- [25] C. Panagiotakis, I. Grinias, G. Tziritas, Natural image segmentation based on tree equipartition, Bayesian flooding and region merging, *IEEE Trans. Image Process.* 20 (8) (2011) 2276–2287, <http://dx.doi.org/10.1109/TIP.2011.2114893>.
- [26] M. Haindl, S. Mikeš, P. Pudil, Unsupervised hierarchical weighted multi-segmenter, in: J. Benediktsson, J. Kittler, F. Roli (Eds.), *Lecture Notes in Computer Science, MCS 2009*, vol. 5519, Springer, Berlin, Heidelberg 2009, pp. 272–282. [http://dx.doi.org/10.1007/978-3-642-02326-2\\_28](http://dx.doi.org/10.1007/978-3-642-02326-2_28).
- [27] M. Haindl, Visual data recognition and modeling based on local Markovian models, in: L. Florack, R. Duits, G. Jongbloed, M.-C. Lieshout, L. Davies (Eds.), *Mathematical Methods for Signal and Image Analysis and Representation, Computational Imaging and Vision*, vol. 41, Springer, London, 2012, pp. 241–259 (Chapter 14), [http://dx.doi.org/10.1007/978-1-4471-2353-8\\_14](http://dx.doi.org/10.1007/978-1-4471-2353-8_14).
- [28] P. Felzenszwalb, D. Huttenlocher, Efficient graph-based image segmentation, *Int. J. Comput. Vis.* 59 (2) (2004) 167–181.
- [29] J. Puzicha, T. Hofmann, J.M. Buhmann, Histogram clustering for unsupervised image segmentation, in: *1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, IEEE, 1999. <http://dx.doi.org/10.1109/CVPR.1999.784981>.
- [30] G. Mori, Guiding model search using segmentation, in: *2005 Tenth IEEE International Conference on Computer Vision*, 2005, ICCV 2005, vol. 2, 2005, pp. 1417–1423. <http://dx.doi.org/10.1109/ICCV.2005.112>.
- [31] M. Donoser, H. Bischof, ROI-SEG: unsupervised color segmentation by combining differently focused sub results, in: *CVPR*, 2007, pp. 1–8. <http://dx.doi.org/10.1109/CVPR.2007.383231>.
- [32] F. Calderero, F. Marques, General region merging approaches based on information theory statistical measures, in: *IEEE International Conference on Image Processing or ICIP*, 2008, pp. 3016–3019. <http://dx.doi.org/10.1109/ICIP.2008.4712430>.
- [33] G. Scarpa, M. Haindl, Unsupervised texture segmentation by spectral-spatial-independent clustering, in: Y. Tang, S. Wang, D. Yeung, H. Yan, G. Lorette (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition, ICPR*

- 2006, vol. II, IEEE Computer Society, Los Alamitos, 2006, pp. 151–154. <http://dx.doi.org/10.1109/ICPR.2006.1147>.
- [34] E. Sharon, M. Galun, D. Sharon, R. Basri, A. Brandt, *Hierarchy and adaptivity in segmenting visual scenes*, *Nature* 442 (7104) (2006) 719–846.
- [35] C. Christoudias, B. Georgescu, P. Meer, *Synergism in low level vision*, in: R. Kasturi, D. Laurendeau, C. Suen (Eds.), *Proceedings of the 16th International Conference on Pattern Recognition*, vol. 4, IEEE Computer Society, Los Alamitos, 2002, pp. 150–155.
- [36] Y. Deng, B. Manjunath, *Unsupervised segmentation of color-texture regions in images and video*, *IEEE Trans. Pattern Anal. Mach. Intell.* 23 (8) (2001) 800–810.
- [37] M.A. Hoang, J.-M. Geusebroek, A.W. Smeulders, *Color texture measurement and segmentation*, *Signal Process.* 85 (2) (2005) 265–275.
- [38] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, J. Malik, *Blobworld: a system for region-based image indexing and retrieval*, in: *Third International Conference on Visual Information Systems*, Springer, Berlin, Heidelberg 1999.

**Michal Haindl** received Ph.D. (1983) and the ScD (DrSc) degree from the Czechoslovak Academy of Sciences. He is a fellow of the IAPR, Professor, head of the Pattern Recognition department of the Institute of Information Theory and Automation, Czech Academy of Sciences, and author of about 300 research papers.

**Stanislav Mikeš** received the M.Sc. and Ph.D. degree (2010) from the Faculty of Mathematics and Physics, Charles University, Prague. He is a researcher at the Pattern Recognition department of the Institute of Information Theory and Automation, Czech Academy of Sciences, focused on unsupervised image and texture segmentation, verification methodology.