Unsupervised Image Segmentation Contest

Michal Haindl and Stanislav Mikeš
Institute of Information Theory and Automation of the ASCR
Prague, Czech Republic 182 08
Email: {haindl,xaos}@utia.cas.cz

***Abstract*—The unsupervised color image segmentation competition is taking place in conjunction with the ICPR 2014 conference. It aims to promote evaluation of unsupervised color image segmentation algorithms using publicly available data sets. The results evaluation is based on the standard performance assessment methodology using the online web verification server. We present in this paper the top six preliminary results submitted to this contest till the ICPR paper submission deadline.**

I. Introduction and Related Work

Unsupervised or supervised texture segmentation is the prerequisite for successful content-based image retrieval, scene analysis, automatic acquisition of virtual models, quality control, security, medical applications and many others. Although more than 1000 different methods were already published [1], [2], [3], [4], [5], [6], [7], [8], this ill-defined problem is still far from being solved and even cannot be solved in its full generality. In addition to that, very little is known about properties and behaviour of already published segmentation methods and their potential user is left to randomly select one due to absence of any counselling. This is among others due to missing reliable performance comparison between different techniques because very limited effort was spent to develop suitable quantitative measures of segmentation quality that can be used to evaluate and compare segmentation algorithms. Rather than advancing the most promising image segmentation approaches novel algorithms are often satisfied just being sufficiently different from the previously published ones and tested only on a few carefully selected positive examples.

The contest aim is to overcome these problems by suggesting the most promising approaches to the unsupervised learning and image segmentation and to unify the verification methodology used in the image segmentation research. The performance of all submitted algorithms will be summarised in a presentation given at the conference.

II. Contest Benchmark

The contest uses the Prague texture segmentation datagenerator and benchmark [9], [10], [11] which is web based (`http://mosaic.utia.cas.cz`) service designed to mutually compare, validate, and rank different texture or image segmenters and to support new segmentation and classification methods development. The benchmark verifies their performance characteristics in either supervised or unsupervised mode on potentially unlimited image / frame sets of monospectral, multispectral, bidirectional texture function (BTF), satellite, and dynamic textures using extensive set of prevalent numerical criteria. It enables to test their noise robustness, scale, rotation or illumination invariance, etc.

III. Contest Data

Benchmark contest data sets are computer generated 512×512 mosaics using the Voronoi polygon random generator filled with randomly selected natural color textures (see Fig.1). The contest uses the large size (80 textural mosaics) unsupervised *Colour* benchmark without noise degradation. Linear region borders are chosen for the contest but the benchmark allows various border types.

Fig. 1. Texture mosaics generating scheme.

IV. Performace Evaluation

The benchmark has implemented the twenty seven most frequented evaluation criteria categorized (see detailed specification in the benchmark) into four groups: region-based [12] (5 criteria with the standard threshold + 5 performance curves – Fig.2 – with their performance integrals over all threshold settings), pixel-wise (12 + F–measure curve), consistency measures (2) [13] and clustering comparison criteria (3) [14]. The performance criteria mutually compare ground truth image regions with the corresponding machine segmented regions. The contest criterion is the average rank over all benchmark criteria. The top methods will be verified by organizers using submitted codes. During the contest submission period all participants can see only their results and the non-contest results in the benchmark. The detailed mutual comparison table will be publicized after the contest workshop.

V. Submitted Methods

A. VRA-PMCFA

The Voting Representativeness – Priority Multi-Class Flooding Algorithm [15] is an unsupervised texture image segmentation framework with unknown number of regions, which involves feature extraction and classification in feature space, followed by flooding and merging in spatial domain. The distribution of the features for the different classes are obtained by a block-wise unsupervised voting framework using the blocks grid graph or its minimum spanning tree and the Mallows distance. The final clustering is obtained by using the k-centroids algorithm. An efficient flooding algorithm is used, namely, Priority Multi-Class Flooding Algorithm, that assign pixels to labels using Bayesian dissimilarity criteria. Finally,

1051-4651/14 $31.00 © 2014 IEEE
DOI 10.1109/ICPR.2014.264

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

TABLE I. Colour benchmark results for VRA-PMCFA, texNCUT, FSEG, MW3AR8, Deep Brain Model, CGCHI; (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 curves integrals; $\bar{F}$ = F–measure curve; small numbers are the corresponding measure rank over the listed methods).

a region merging method, which incorporates boundary information, is introduced for obtaining the final segmentation map. The proposed scheme is executed for several number of 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.

B. FSEG

A factorizaton-based texture segmenter [16] uses local spectral histograms as features. It constructs an $M \times N$ feature matrix using M–dimensional feature vectors in an N–pixel image. Based on the observation that each feature can be approximated by a linear combination of several representative features, the method factors the feature matrix into two matrices - one consisting of the representative features, and the other containing weights of representative features at each pixel used for linear combination. The factorization method is based on singular value decomposition and nonnegative matrix factorization. The method uses local spectral histograms to discriminate region appearances in a computationally efficient way and at the same time accurately localizes region boundaries.

C. MW3AR8

An unsupervised multi-spectral, multi-resolution, multiple-segmenter for textured images with unknown number of classes is based on [17]. The segmenter is based on a weighted combination of several unsupervised segmentation results, each in 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. Single local texture model is expressed as a stationary causal uncorrelated noise driven 3D autoregressive process [18]:

$$Y_r = \gamma X_r + e_r \ ,$$

where $\gamma = [A_1, \ldots, A_\eta]$ is the parameter matrix, $r = [r_1, r_2]$ is the regular lattice multiindex, I_r^c is a causal neighborhood index set with $\eta = card(I_r^c)$ and e_r is a white Gaussian noise vector with zero mean and a constant but unknown covariance, X_r is a corresponding vector of the contextual neighbours

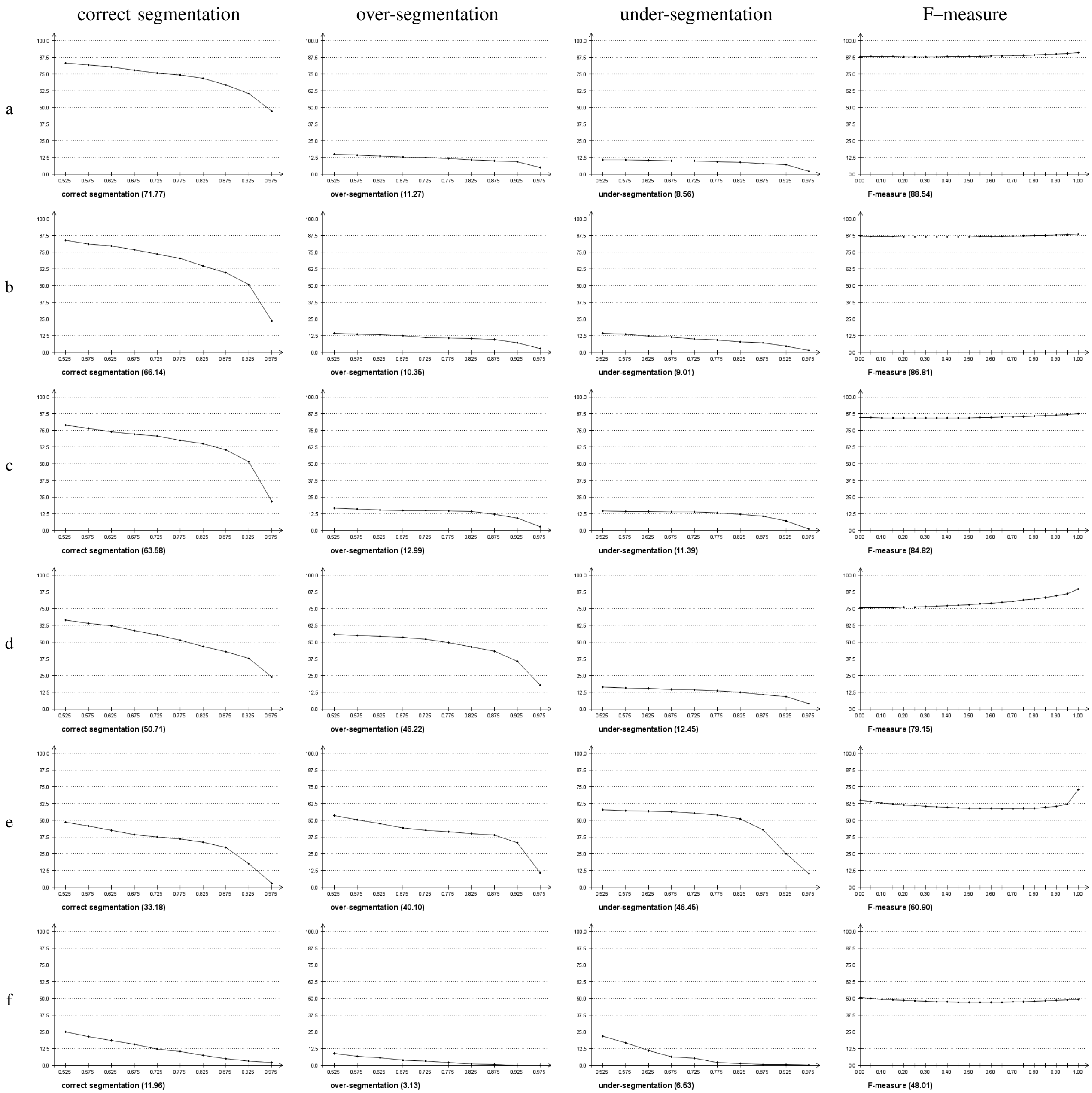

Fig. 2. Benchmark criteria curves for method: VRA-PMCFA (a), texNCUT (b), FSEG (c), MW3AR8 (d), Deep Brain Model (e), and CGCHI (f), respectively.

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.

D. Deep Brain Model

Deep brain model is an unsupervised segmentation framework with 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 [19] is then employed to select object boundaries among the edges of superpixels.

E. CGCHI

The Combined Graph Cut [20] based segmentation with histogram information [21] on regions method is a combination of global and local coherent information. It finds the sufficient number of clusters by using histograms and probability theory. Subsequently, the method uses metric space strategies to model local intensity features of input image. Some problems in this method are from two main sources, wrong number of cluster estimation and the other one is the modeling method failures.

F. *texNCUT*

A modification of the NCUT method [4] which is using textural features.

VI. Results Evaluation

Selected (twelve of eighty) test images for visual comparison of the recent top six methods (VRA-PMCFA, texNCUT, FSEG, MW3AR8, Deep Brain Model, CGCHI) submitted to the contest are shown in Figs.3,4.

However, the main benefit of the benchmark are the numerical performance criteria evaluated for each tested method. Integrated numerical results of these six methods are reported in Tab.I, where $\uparrow$ / $\downarrow$ denotes the required criterion direction and bold numbers the best criterion value achieved from all six compared methods.

It shows a qualitative gap between the VRA-PMCFA method and the remaining ones. The VRA-PMCFA method scores best in all criteria except five criteria (OS, US, C, RM, dVI). The method is very robust in the correct segmentation criterion which is demonstrated on the flat curve (Fig.2-a) of this criterion. The second texNCUT method performs well except its region border localization which is rather poor as can be seen on Figs.3,4-d. It is indicated also by numerical results of the global and local consistency error criteria (see LCE and GCE rows in Tab.I). The third FSEG method has solid performance on all criteria (ranked between 2–4) but does not win any specific criterion. The fourth MW3AR8 method is slightly worse than the third one. Its current version has strong over-segmentation tendency which is also demonstrated on Fig.2-d.

More detailed insight into the behaviour of single methods can be obtained by consulting corresponding criteria description [10] and their achieved values in Tab.I.

VII. Conclusion

Unusually extensive benchmarking of the contest methods allows to get deep and reliable insight into their properties. The presented results are still preliminary, especially the last two methods (Deep Brain Model, CGCHi) where authors just submitted their very early results. Some contest authors are continuing development of their algorithms and the final top methods ordering and their performance might change.

Acknowledgment

This research was supported by grant GAČR 14-10911S.

References

[1] Y. J. Zhang, "Evaluation and comparison of different segmentation algorithms," *Pattern Recognition Letters*, vol. 18, pp. 963–974, 1997.

[2] K. Fu and J. Mui, "A survey on image segmentation," *Pattern Recognition*, vol. 13, pp. 3–16, 1981.

[3] N. Pal and S. Pal, "A review on image segmentation techniques," *Pattern Recognition*, vol. 26, no. 9, pp. 1277–1294, 1993.

[4] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, 2000.

[5] L. Lucchese and S. Mitra, "Color image segmentation: A state-of-the-art survey," *Proc. of the Indian National Science Academy (INSA-A)*, vol. 67, no. 2, pp. 207–221, 2001.

[6] H. Cheng, X. Jiang, Y. Sun, and J. Wang, "Color image segmentation: advances and prospects," *Pattern Recognition*, vol. 34, pp. 2259–2281, 2001.

[7] J. Freixenet, X. Munoz, D. Raba, J. Marti, and X. Cufi, "Yet another survey on image segmentation: Region and boundary information integration," in *European Conference on Computer Vision*, 2002, p. III: 408 ff. [Online]. Available: http://link.springer-ny.com/link/service/series/0558/bibs/2352/23520408.htm

[8] M. Galun, E. Sharon, R. Basri, and A. Brandt, "Texture segmentation by multiscale aggregation of filter responses and shape elements," in *International Conference on Computer Vision*, 2003, pp. 716–723. [Online]. Available: http://dx.doi.org/10.1109/ICCV.2003.1238418

[9] S. Mikeš and M. Haindl, "Prague texture segmentation data generator and benchmark," *ERCIM News*, no. 64, pp. 67–68, 2006.

[10] M. Haindl and S. Mikeš, "Texture segmentation benchmark," in *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008*, B. Lovell, D. Laurendeau, and R. Duin, Eds. Los Alamitos: IEEE Computer Society, December 2008, pp. 1–4. [Online]. Available: http://doi.ieeecomputersociety.org/

[11] M. Haindl and S. Mikeš, *Pattern Recognition; Techniques, Technology and Applications*. Vienna: I-Tech Education and Publishing, 2008, ch. Unsupervised Texture Segmentation, pp. 227–248. [Online]. Available: http://www.intechweb.org/book.php?id=88&content=subject&sid=1

[12] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, D. B. Goldgof, K. Bowyer, D. W. Eggert, A. Fitzgibbon, and R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 673–689, July 1996.

[13] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, vol. 2, July 2001, pp. 416–423. [Online]. Available: http://www.cs.berkeley.edu/projects/vision/grouping/segbench/

[14] M. Meila, "Comparing clusterings – an axiomatic view," in *ICML: Machine Learning: Proceedings of the Seventh International Conference*, 2005, pp. 577 – 584.

[15] C. Panagiotakis, L. Grinias, and G. Tziritas, "Texture Segmentation Based on Voting of Blocks, Bayesian Flooding and Region Merging," in *submitted to Proceedings of the 22nd International Conference on Pattern Recognition, ICPR 2014*, 2014.

[16] J. Yuan and D. Wang, "Factorization-based texture segmentation," The Ohio State University, Columbus, USA, Tech. Rep. OSU-CISRC-1/13-TR0, 2013.

[17] M. Haindl, S. Mikeš, and P. Pudil, "Unsupervised hierarchical weighted multi-segmenter," in *Lecture Notes in Computer Science, MCS 2009*, J. Benediktsson, J.Kittler, and F. Roli, Eds., vol. 5519 (2009). Springer, 2009, pp. 272–282.

[18] M. Haindl, "Visual data recognition and modeling based on local markovian models," in *Mathematical Methods for Signal and Image Analysis and Representation*, ser. Computational Imaging and Vision, L. Florack, R. Duits, G. Jongbloed, M.-C. Lieshout, and L. Davies, Eds. Springer London, 2012, vol. 41, ch. 14, pp. 241–259, 10.1007/978-1-4471-2353-8_14. [Online]. Available: http://dx.doi.org/10.1007/978-1-4471-2353-8_14

[19] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," *IJCV*, vol. 59, no. 2, pp. 167–181, September 2004.

[20] M. Salah, A. Mitiche, and I. Ayed, "Multiregion image segmentation by parametric kernel graph cuts," *Image Processing, IEEE Transactions on*, vol. 20, no. 2, pp. 545–557, 2011.

[21] J. Puzicha, T. Hofmann, and J. Buhmann, "Histogram clustering for unsupervised image segmentation," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, 1999, pp. –608 Vol. 2.

Fig. 3. Selected benchmark mosaics (a), ground-truth (b), VRA-PMCFA (c), texNCUT (d), FSEG (e), MW3AR8 (f), Deep Brain Model (g), CGCHI (h) segmentation results, respectively.

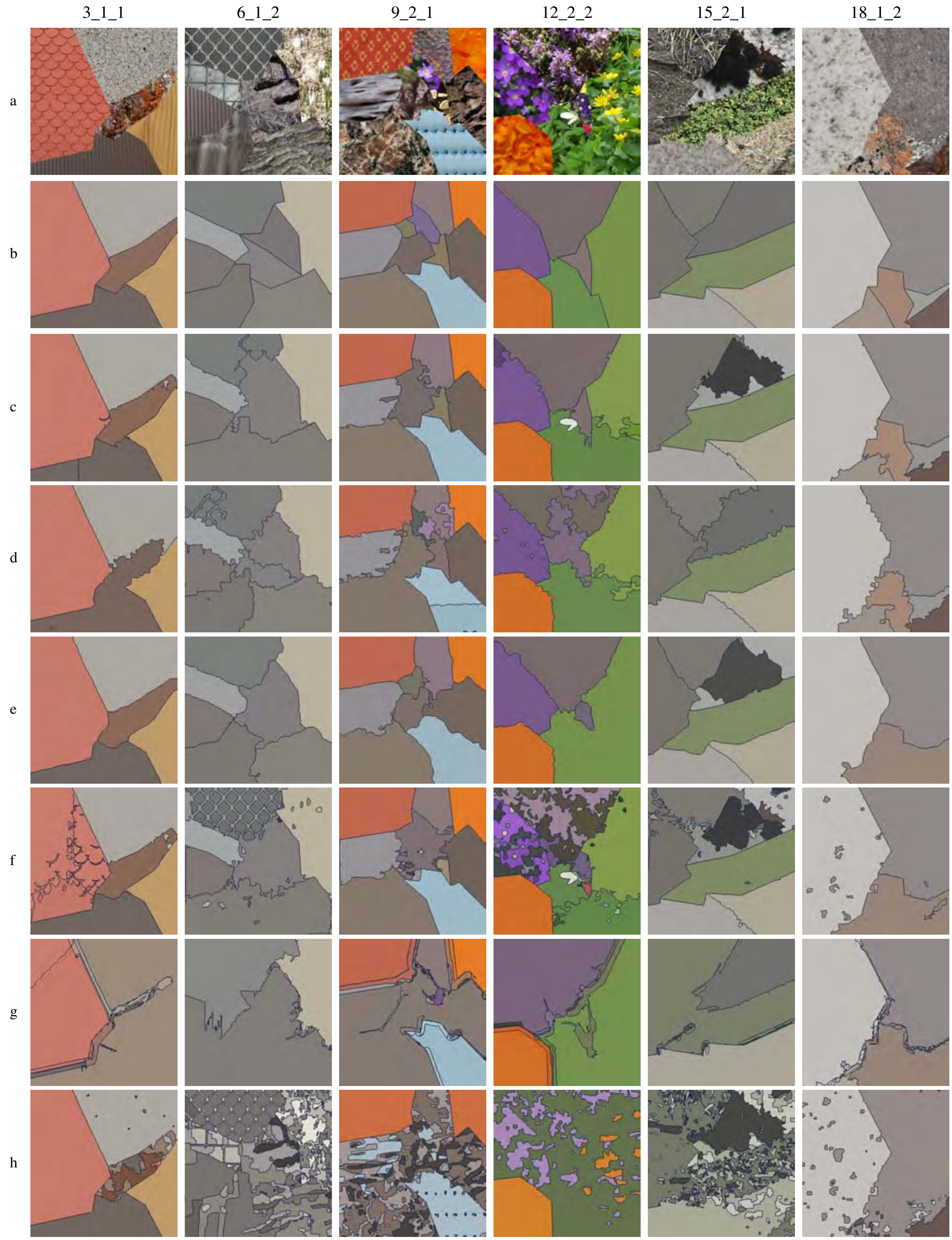

Fig. 4. Selected benchmark mosaics (a), ground-truth (b), VRA-PMCFA (c), texNCUT (d), FSEG (e), MW3AR8 (f), Deep Brain Model (g), CGCHI (h) segmentation results, respectively.