Unsupervised Hierarchical Weighted Multi-segmenter

Michal Haindl[1,2], Stanislav Mikeš[1], and Pavel Pudil[1,2]

[1] Institute of Information Theory and Automation
Academy of Sciences CR, Prague, Czech Republic
[2] Faculty of Management, University of Economics
Jindřichův Hradec, Czech Republic
{haindl,xaos}@utia.cz

Abstract. An unsupervised multi-spectral, multi-resolution, multiple-segmenter for textured images with unknown number of classes is presented. 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 four causal directional multi-spectral random field models recursively evaluated for each pixel. 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. The performance of the presented method is extensively tested on the Prague segmentation benchmark using the commonest segmentation criteria and compares favourably with several leading alternative image segmentation methods.

1 Introduction

Segmentation is the fundamental process which partitions a data space into meaningful salient regions. Image segmentation essentially affects the overall performance of any automated image analysis system thus its quality is of the utmost importance. Image regions, homogeneous with respect to some usually textural or colour measure, which result from a segmentation algorithm are analysed in subsequent interpretation steps. Texture-based image segmentation is area of intense research activity in recent years and many algorithms were published in consequence of all this effort. These methods are usually categorised [1] as region-based, boundary-based, or as a hybrid of the two. Different published methods are difficult to compare because of lack of a comprehensive analysis together with accessible experimental data, however available results indicate that the ill-defined texture segmentation problem is still far from being satisfactorily solved. Spatial interaction models and especially Markov random fields-based models are increasingly popular for texture representation [1,2], etc. Several researchers dealt with the difficult problem of unsupervised segmentation using these models see for example [3,4,5] or [6,7,8].

J.A. Benediktsson, J. Kittler, and F. Roli (Eds): MCS 2009, LNCS 5519, pp. 272–282, 2009.

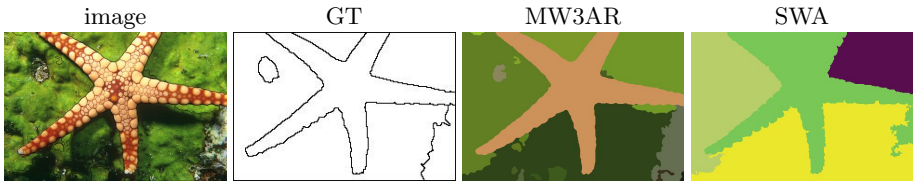

Fig. 1. Selected Berkeley benchmark image, ground truth from the benchmark and the segmentation results from the presented method (MW3AR) and SWA [9]

The concept of decision fusion [10] for high-performance pattern recognition is well known and widely accepted in the area of supervised classification where (often very diverse) classification technologies, each providing complementary sources of information about class membership, can be integrated to provide more accurate, robust and reliable classification decisions than the single classifier applications.

Similar advantages can be expected and achieved [8] also for the unsupervised segmentation applications. However, a direct unsupervised application of the supervised classifiers fusion idea is complicated with unknown number of data hidden classes and consequently a different number of segmented regions in segmentation results to be fused. This paper exploits above advantages by combining several unsupervised segmenters of the same type but with different feature sets.

2 Combination of Multiple Segmenters

The proposed method (MW3AR) combines segmentation results from different resolution. We assume to down-sample input image Y into M different resolutions $Y^{(m)} = \downarrow^{\iota_m} Y$ with sampling factors ι_m $m = 1, \ldots, M$ identical in both horizontal and vertical directions and $Y^{(1)} = Y$. Local texture for each pixel $Y_r^{(m)}$ is represented the 3D simultaneous causal autoregressive random field model (CAR) parameter space Θ_r (4) and modelled by the Gaussian mixture model (5),(6).

2.1 Single-Resolution Texture Model

Static smooth multi-spectral textures require three dimensional models for adequate representation. We assume that single multi-spectral textures can be locally modelled using a 3D simultaneous causal autoregressive random field model (CAR). This model can be expressed as a stationary causal uncorrelated noise driven 3D autoregressive process [11]:

$$Y_r = \gamma X_r + e_r \ , \tag{1}$$

where $\gamma = [A_1, \ldots, A_\eta]$ is the $d \times d\eta$ parameter matrix, d is the number of spectral bands, 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 Y_{r-s} and $r, r-1, \ldots$ is a chosen direction of movement on the image index lattice I. The selection of an appropriate CAR model support (I_r^c) is important to obtain good texture representation but less important for segmentation. The optimal neighbourhood as well as the Bayesian parameters estimation of a CAR model can be found analytically under few additional and acceptable assumptions using the Bayesian approach (see details in [11]). The recursive Bayesian parameter estimation of the CAR model is [11]:

$$\hat{\gamma}_{r-1}^T = \hat{\gamma}_{r-2}^T + \frac{V_{x(r-2)}^{-1} X_{r-1}(Y_{r-1} - \hat{\gamma}_{r-2} X_{r-1})^T}{(1 + X_{r-1}^T V_{x(r-2)}^{-1} X_{r-1})} \ , \tag{2}$$

where $V_{x(r-1)} = \sum_{k=1}^{r-1} X_k X_k^T + V_{x(0)}$. Local texture for each pixel is represented by four parametric vectors. Each vector contains local estimations of the CAR model parameters. These models have identical contextual neighbourhood I_r^c but they differ in their major movement direction (top-down, bottom-up, rightward, leftward), i.e.,

$$\tilde{\gamma}_r^T = \{\hat{\gamma}_r^t, \hat{\gamma}_r^b, \hat{\gamma}_r^r, \hat{\gamma}_r^l\}^T \ . \tag{3}$$

The parametric space $\tilde{\gamma}$ is subsequently smooth out, rearranged into a vector and its dimensionality is reduced using the Karhunen-Loeve feature extraction $(\bar{\gamma})$. Finally we add the average local spectral values ζ_r to the resulting feature vector (Θ_r).

2.2 Mixture Based Segmentation

Multi-spectral texture segmentation is done by clustering in the CAR parameter space Θ defined on the lattice I where

$$\Theta_r = [\bar{\gamma}_r, \zeta_r]^T \tag{4}$$

is the modified local parameter vector (3) computed for the lattice location r. We assume that this parametric space can be represented using the Gaussian mixture model (GM) with diagonal covariance matrices due to the previous CAR parametric space decorrelation. The Gaussian mixture model for CAR parametric representation at the m-th resolution $(m = 1, \ldots, M)$ is as follows:

$$p(\Theta_r^{(m)}) = \sum_{i=1}^{K^{(m)}} p_i^{(m)} \, p(\Theta_r^{(m)} \,|\, \nu_i^{(m)}, \Sigma_i^{(m)}) \ , \tag{5}$$

$$p(\Theta_r^{(m)} \,|\, \nu_i^{(m)}, \Sigma_i^{(m)}) = \frac{|\Sigma_i^{(m)}|^{-\frac{1}{2}}}{(2\pi)^{\frac{d}{2}}} \, e^{-\frac{(\Theta_r^{(m)} - \nu_i^{(m)})^T (\Sigma_i^{(m)})^{-1} (\Theta_r^{(m)} - \nu_i^{(m)})}{2}} \ . \tag{6}$$

The mixture model equations (5),(6) are solved using a modified EM algorithm.

Initialization. The algorithm is initialised using $\nu_i^{(m)}, \Sigma_i^{(m)}$ statistics for each resolution m estimated from the corresponding thematic maps in two subsequent steps:

1. refining direction
   $\nu_i^{(m-1)}\left(\forall \Theta_r^{(m-1)} : r \in \uparrow \Xi_i^{(m)}\right), \quad \Sigma_i^{(m-1)}\left(\forall \Theta_r^{(m-1)} : r \in \uparrow \Xi_i^{(m)}\right)$
   $m = M+1, M, \ldots, 2 \quad i = 1, \ldots, K^{(m)}$,
2. coarsening direction
   $\nu_i^{(m)}\left(\forall \Theta_r^{(m)} : r \in \downarrow \Xi_i^{(m-1)}\right), \quad \Sigma_i^{(m)}\left(\forall \Theta_r^{(m)} : r \in \downarrow \Xi_i^{(m-1)}\right)$
   $m = 2, 3, \ldots, M \quad i = 1, \ldots, K^{(m)}$,

where $\Xi_i^{(m)} \subset I \ \forall m, i$, and the first initialisation thematic map $\Xi_i^{(M+1)}$ is approximated by the rectangular subimages obtained by regular division of the input texture mosaic. All the subsequent refining step are initialised from the preceding coarser resolution upsampled thematic maps. The final initialisation results from the second coarsening direction where the gradually coarsening segmentations are initialised using the preceding downsampled thematic maps. For each possible couple of components the Kullback Leibler divergence

$$D\left(p(\Theta_r \mid \nu_i, \Sigma_i) \,||\, p(\Theta_r \mid \nu_j, \Sigma_j)\right) = \int_\Omega p(\Theta_r \mid \nu_i, \Sigma_i) \log\left(\frac{p(\Theta_r \mid \nu_i, \Sigma_i)}{p(\Theta_r \mid \nu_j, \Sigma_j)}\right) d\Theta_r$$

is evaluated and the most similar components, i.e.,

$$\{i, j\} = \arg\min_{k,l} D\left(p(\Theta_r \mid \nu_l, \Sigma_l) \,||\, p(\Theta_r \mid \nu_k, \Sigma_k)\right)$$

are merged together in each initialisation step. This initialisation results in K_{ini} subimages and recomputed statistics ν_i, Σ_i . $K_{ini} > K$ where K is the optimal number of textured segments to be found by the algorithm.

Two steps of the EM algorithm are repeating after initialisation. The components with smaller weights than a fixed threshold $(p_j < \frac{0.1}{K_{ini}})$ are eliminated. For every pair of components we estimate their Kullback Leibler divergence. From the most similar couple, the component with the weight smaller than the threshold is merged to its stronger partner and all statistics are actualised using the EM algorithm. The algorithm stops when either the likelihood function has negligible increase $(\mathcal{L}_t - \mathcal{L}_{t-1} < 0.05)$ or the maximum iteration number threshold is reached.

2.3 Resulting Mixture Probabilities

Resulting mixture model probabilities are mapped to the original fine resolution image space for all $m = 1, \ldots, M$ mixture submodels ((5)(6)). The M cooperating segmenters deliver their class response in the form of conditional probabilities. Each segmenter produces a preference list based on the mixture component probabilities of a particular pixel belonging a particular class, together with a set of confidence measurement values generated in the original decision-making process.

Single-segmenters Correspondence. Single-resolution segmentation results cannot be combined without knowledge of the mutual correspondence between regions in all different-resolution segmentation probabilistic mixture component maps ($K^1 \times \sum_{m=2}^{M} K^m$ combinations). Mutual assignments of two probabilistic maps are solved by using the Munkre's assignment algorithm [8] which finds the minimal cost assignment

$$g : A \mapsto B, \ \sum_{\alpha \in A} f(\alpha, g(\alpha))$$

between sets A, B, $|A| = |B| = n$ given the cost function $f(\alpha, \beta)$, $\alpha \in A, \beta \in B$. α corresponds to the fine resolution probabilistic maps, β corresponds to downsampled probabilistic maps and $f(\alpha, \beta)$ is the Kullback Leibler divergence between probabilistic maps. The algorithm has polynomial complexity instead of exponential for the exhaustive search.

Final Parametric Space. The parametric vectors representing texture mosaic pixels are assigned to the clusters based on our modification of the sum rule according to the highest component probabilities, i.e., Y_r is assigned to the cluster ω_{j^*} if

$$\pi_{r,j^*} = max_j \sum_{s \in I_r} w_s \left(\sum_{m=1}^{M} \frac{p^2(\Theta_{r-s}^{(m)} \mid \nu_j^{(m)}, \Sigma_j^{(m)})}{\sum_{i=1}^{M} p(\Theta_{r-s}^{(i)} \mid \nu_j^{(i)}, \Sigma_j^{(i)})} \right) ,$$

where w_s are fixed distance-based weights, I_r is a rectangular neighbourhood and $\pi_{r,j^*} > \pi_{thre}$ (otherwise the pixel is unclassified). The area of single cluster blobs is evaluated in the post-processing thematic map filtration step. Regions with similar statistics are merged. Thematic map blobs with area smaller than a given threshold are attached to its neighbour with the highest similarity value.

3 Experimental Results

The algorithm was tested on natural colour textures mosaics from the Prague Texture Segmentation Data-Generator and Benchmark (http://mosaic.utia.cas.cz) [12]. The benchmark test mosaics layouts and each cell texture membership are randomly generated and filled with colour textures from the large (more than 1000 high resolution colour textures) Prague colour texture database. The benchmark ranks segmentation algorithms according to a chosen criterion. There are implemented twenty seven most frequented evaluation criteria categorised into four criteria groups – region-based [12], pixel-wise [12], clustering comparison criteria, and consistency measures [12]. The region-based [12] performance criteria mutually compare ground truth (GT) image regions with the corresponding machine segmented regions (MS). The pixel-wise criteria group contains the most frequented classification criteria such as the omission and commission errors, class accuracy, recall, precision, etc. Finally the last two criteria sets incorporate the global and local consistency errors [12] and three clustering comparison criteria.

Table 1. 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; dM = Mirkin metric; dD = Van Dongen metric; dVI = variation of information;

	Benchmark – Colour											
	MW3AR	TFR / KLD [13]	TFR [14]	AR3D + EM multi [8]	AR3D + EM [7]	GMRF + EM [6]	HGSE [15]	EGBIS [16]	JSEG [17]	SWA def_par [9]	Blobworld [18]	EDISON [19]
CS	53.04	51.25	46.13	43.22	37.42	31.93	29.81	28.78	27.47	27.06	21.01	12.68
OS	59.53	5.84	2.37	49.27	59.53	53.27	10.69	19.69	38.62	50.21	7.33	86.91
US	3.20	7.16	23.99	16.55	8.86	11.24	33.76	39.15	5.04	4.53	9.30	0.00
ME	5.63	31.64	26.70	10.30	12.54	14.97	26.89	20.42	35.00	25.76	59.55	2.48
NE	6.96	31.38	25.23	12.56	13.14	16.91	25.04	21.54	35.50	27.50	61.68	4.68
O	19.32	19.65	28.73	21.99	34.32	33.61	48.94	44.35	37.94	33.01	41.45	73.17
C	86.19	9.67	12.50	87.38	100.00	100.00	32.39	82.87	92.77	85.19	58.94	100.00
CA	71.89	67.45	61.32	64.51	59.46	57.91	49.60	51.10	55.29	54.84	46.23	31.19
CO	74.66	76.40	73.00	71.00	64.81	63.51	63.37	64.12	61.81	60.67	56.04	31.55
CC	95.04	81.12	68.91	90.14	91.79	89.26	66.09	72.73	87.70	88.17	73.62	98.09
I.	25.34	23.60	27.00	29.00	35.19	36.49	36.63	35.88	38.19	39.33	43.96	68.45
II.	0.74	4.09	8.56	3.79	3.39	3.14	13.51	7.59	3.66	2.11	6.72	0.24
EA	80.43	75.80	68.62	73.90	69.60	68.41	58.74	59.88	66.74	66.94	58.37	41.29
MS	71.78	65.19	59.76	64.47	58.89	57.42	46.63	49.03	55.14	53.71	40.36	31.13
RM	3.09	7.21	8.61	4.55	4.88	4.86	13.31	8.38	4.96	6.11	7.96	3.21
CI	82.43	77.21	69.73	76.51	73.15	71.80	61.17	63.11	70.27	70.32	61.31	50.29
GCE	8.17	20.35	15.52	15.31	12.13	16.03	16.75	16.64	18.45	17.27	31.16	3.55
LCE	5.78	14.36	12.03	7.97	6.69	7.31	10.46	8.97	11.64	11.49	23.19	3.44
dM	8.97	12.64	17.47	13.51	15.43	15.27	27.95	19.72	15.19	13.68	20.03	16.84
dD	14.78	18.01	18.21	16.87	19.76	20.63	22.90	21.29	23.38	24.20	31.11	35.37
dVI	16.67	14.06	13.04	16.11	17.10	17.32	12.83	13.79	17.37	17.16	15.84	25.65

Tab.1 compares the overall benchmark performance of the proposed algorithm MW3AR ($M = 5, \iota_1 = 1, \iota_2 = 1.5, \iota_3 = 2$, with the Blobworld [18], JSEG [17], Edison [19], TFR/KLD [14], SWA [9], EGBIS [16], HGS [15], and our previously published methods AR3D-multi [8], GMRF [6], AR3D [7], respectively. MW3AR demonstrates a significant improvement (e.g. 23 % for the correct segmentation CS criterion) over our previously published unsupervised multi-segmenter AR3D-multi [8].

These results illustrated in Figs.1,2,3 and Tab.1 demonstrate very good pixel-wise, correct region segmentation, missed error, noise error, and undersegmentation properties of our method while the oversegmentation results are slightly

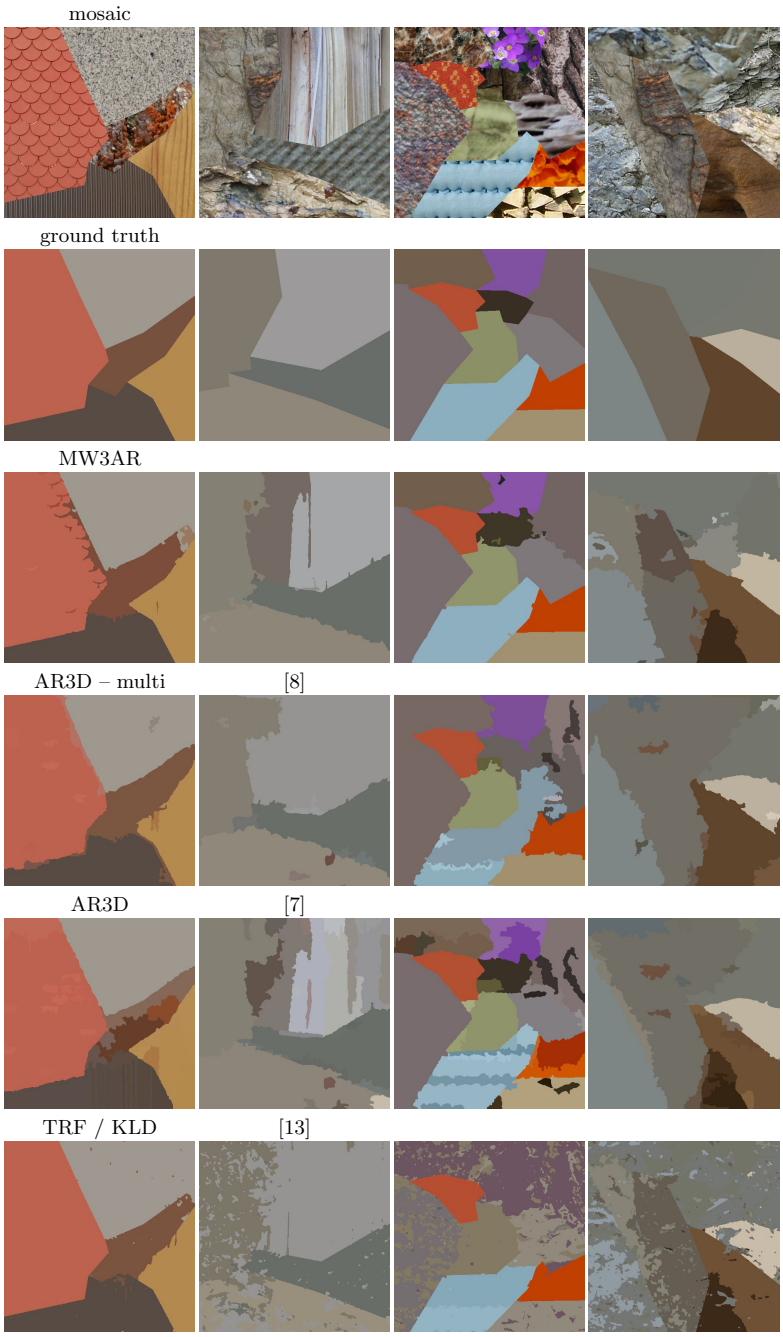

Fig. 2. Selected experimental texture mosaics, ground truth from the benchmark and the corresponding segmentation results

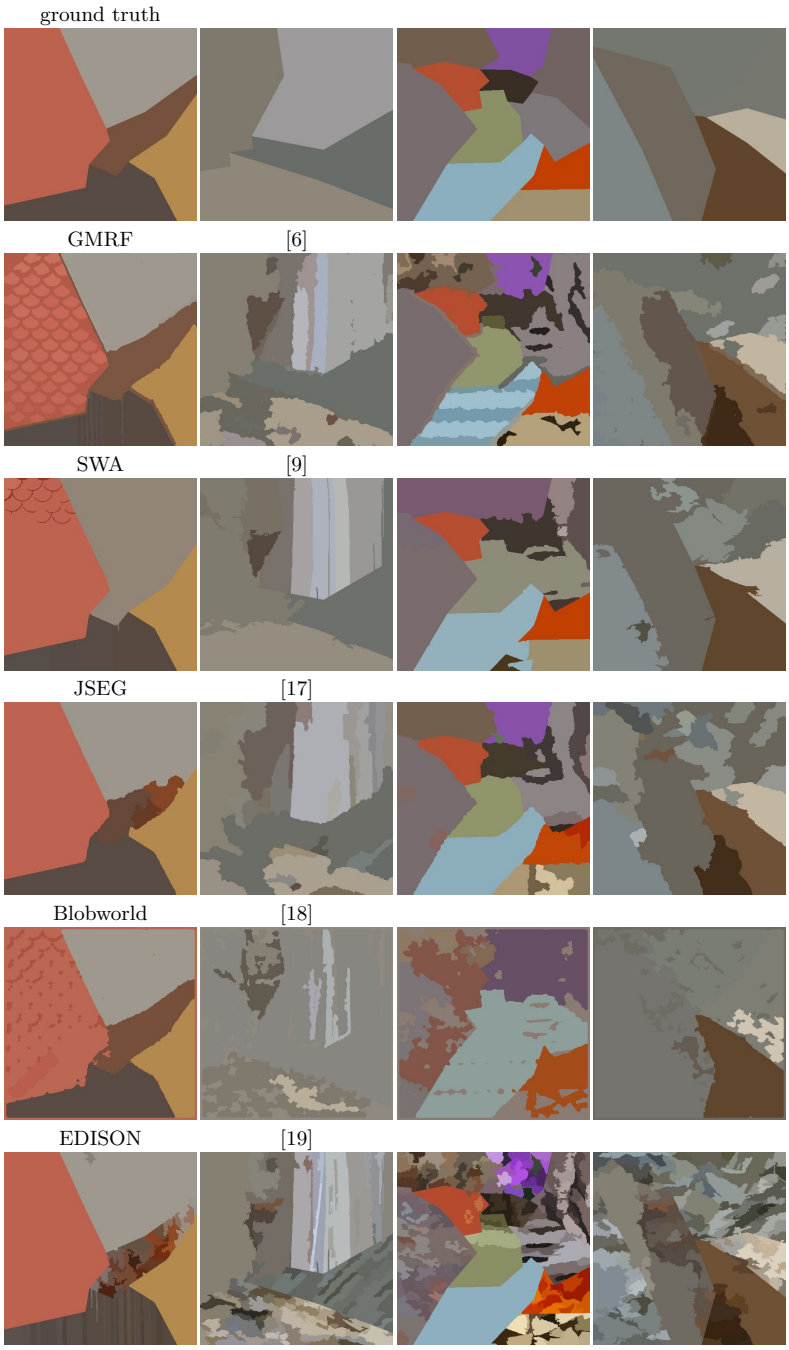

Fig. 3. Selected ground truth from the benchmark and the corresponding segmentation results

worse and dVI results are only average. For all the pixel-wise criteria or the consistency measures our method is among the best ones. The table demonstrates improvement of the presented multi-segmenter method over our previous multi-segmenter [8] and its single-segmenter version published earlier [7] in most benchmark criteria.

Figs.2,3 and show four selected 512×512 experimental benchmark mosaics created from four to eleven natural colour textures. The last four or five rows on these figures demonstrate comparative results from the eight alternative leading algorithms. Hard natural textures were chosen rather than synthesised (for example using Markov random field models) ones because they are expected to be more difficult for the underlying segmentation model. The third row on Fig.2 demonstrates robust behaviour of our CAR3D-multi algorithm but also infrequent algorithm failures producing the oversegmented thematic map for some textures. Such failures can be reduced by a more elaborate postprocessing step. The TFR/KLD [14], AR3D [7], GMRF [6], SWA [9], EGBIS [16], JSEG [17], Blobworld [18], HGS [15], and Edison [19], algorithms on these data performed mostly worse as can be seen in their corresponding rows on Figs.2,3 some areas are undersegmented while other parts of the mosaics are oversegmented. Fig.2 illustrates also the improvement of the multi-segmenter version of the algorithm at the cost of slight increase of computational complexity. These results can be further improved by sophisticated postprocessing and by the optimisation of the directional models contextual neighbourhoods.

4 Conclusions

We proposed a significant improvement of our previously published unsupervised multi-segmenter. The MW3AR segmenter is computationally efficient, noise resilient and robust method for unsupervised textured image segmentation with unknown number of classes based on the underlying CAR and GM texture models. The algorithm is very fast, despite of using the random field type data representation, due to its efficient recursive parameter estimation of the underlying models and therefore is much faster than the usual Markov chain Monte Carlo estimation approach required for these image representations. Usual drawback of most segmentation methods is their application dependent parameters to be experimentally estimated. Our method requires only a contextual neighbourhood selection and two additional thresholds. The method's performance is demonstrated on the extensive benchmark tests on natural texture mosaics. It performs favourably compared with eight alternative leading segmentation algorithms. Our method accomplishes very good segmentation results also on natural images from the Berkeley segmentation benchmark as well as on remote sensing images.

Acknowledgements

This research was supported by the grant GAČR 102/08/0593 and partially by the MŠMT grants 1M0572 DAR, 2C06019.

References

1. Reed, T.R., du Buf, J.M.H.: A review of recent texture segmentation and feature extraction techniques. CVGIP–Image Understanding 57(3), 359–372 (1993)
2. Haindl, M.: Texture synthesis. CWI Quarterly 4(4), 305–331 (1991)
3. Panjwani, D., Healey, G.: Markov random field models for unsupervised segmentation of textured color images. IEEE Transactions on Pattern Analysis and Machine Intelligence 17(10), 939–954 (1995)
4. Manjunath, B., Chellapa, R.: Unsupervised texture segmentation using markov random field models. IEEE Transactions on Pattern Analysis and Machine Intelligence 13, 478–482 (1991)
5. Haindl, M.: Texture segmentation using recursive markov random field parameter estimation. In: Proceedings of the 11th Scandinavian Conference on Image Analysis, Lyngby, Denmark, Pattern Recognition Society of Denmark, pp. 771–776 (1999)
6. Haindl, M., Mikeš, S.: Model-based texture segmentation. In: Campilho, A.C., Kamel, M.S. (eds.) ICIAR 2004. LNCS, vol. 3212, pp. 306–313. Springer, Heidelberg (2004)
7. Haindl, M., Mikeš, S.: Unsupervised texture segmentation using multispectral modelling approach. In: Proceedings of the 18th Int. Conf. on Pattern Recognition, ICPR 2006, vol. II, pp. 203–206. IEEE Computer Society, Los Alamitos (2006)
8. Haindl, M., Mikes, S.: Unsupervised texture segmentation using multiple segmenters strategy. In: Haindl, M., Kittler, J., Roli, F. (eds.) MCS 2007. LNCS, vol. 4472, pp. 210–219. Springer, Heidelberg (2007)
9. Sharon, E., Galun, M., Sharon, D., Basri, R., Brandt, A.: Hierarchy and adaptivity in segmenting visual scenes. Nature 442(7104), 719–846 (2006)
10. Kittler, J., Hojjatoleslami, A., Windeatt, T.: Weighting factors in multiple expert fusion. In: Proc. BMVC, BMVA, pp. 41–50 (1997)
11. Haindl, M., Šimberová, S.: A Multispectral Image Line Reconstruction Method. In: Theory & Applications of Image Analysis, pp. 306–315. World Scientific Publishing Co., Singapore (1992)
12. Haindl, M., Mikeš, S.: Texture segmentation benchmark. In: Lovell, B., Laurendeau, D., Duin, R. (eds.) Proceedings of the 19th Int. Conf. on Pattern Recognition, ICPR 2008. IEEE Computer Society, Los Alamitos (2008)
13. Scarpa, G., Haindl, M., Zerubia, J.: A hierarchical finite-state model for texture segmentation. In: ICASSP 2007. IEEE Int. Conf. on Acoustics, Speech and Signal Processing, vol. I, pp. 1209–1212. IEEE, Los Alamitos (2007)
14. Scarpa, G., Haindl, M.: Unsupervised texture segmentation by spectral-spatial-independent clustering. In: Proc. of the 18th Int. Conf. on Pattern Recognition, ICPR 2006, vol. II, pp. 151–154. IEEE Computer Society, Los Alamitos (2006)
15. Hoang, M.A., Geusebroek, J.M., Smeulders, A.W.: Color texture measurement and segmentation. Signal Processing 85(2), 265–275 (2005)
16. Felzenszwalb, P., Huttenlocher, D.: Efficient graph-based image segmentation. IJCV 59(2), 167–181 (2004)

17. Deng, Y., Manjunath, B.: Unsupervised segmentation of color-texture regions in images and video. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(8), 800–810 (2001)
18. Carson, C., Thomas, M., Belongie, S., Hellerstein, J.M., Malik, J.: Blobworld: A system for region-based image indexing and retrieval. In: Third International Conference on Visual Information Systems. Springer, Heidelberg (1999)
19. Christoudias, C., Georgescu, B., Meer, P.: Synergism in low level vision. In: Proceedings of the 16th Int. Conf. on Pattern Recognition, vol. 4, pp. 150–155. IEEE Computer Society, Los Alamitos (2002)