# Unsupervised Texture Segmentation

Michal Haindl and Stanislav Mikeš
*Institute of Information Theory and Automation of the ASCR*
*Czech Republic*

## 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 the past thirty years and many algorithms were published in consequence of all this effort, starting from simple thresholding methods up to the most sophisticated random field type methods. Unsupervised methods which do not assume any prior scene knowledge which can be learned to help segmentation process are obviously more challenging than the supervised ones.

Segmentation methods are usually categorized (Reed et al., 1993) 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. The clustering approach resulted in agglomerative and divisive algorithms which were modified for image segmentation as region-based merge and split algorithms. Spatial interaction models and especially Markov random field-based models are increasingly popular for texture representation (Kashyap, 1986; Reed et al., 1993; Haindl, 1991), etc. Several researchers dealt with the difficult problem of unsupervised segmentation using these models see for example (Panjwani et al., 1995; Manjunath et al., 1991; Andrey et al., 1998; Haindl, 1999) or (Haindl et al., 2004, 2005, 2006a).

In this chapter we assume constant illumination and viewing angles for all scene textures, or alternatively that the Lambert law holds for all scene surfaces. If this assumption cannot be assumed than all textures have to be treated either as Bidirectional Texture Functions (BTFs) or some illumination invariant features (Haindl et al., 2006b; Vacha et al., 2007) have too be used.

## 2. Texture segmentation methods

Segmentation methods are based on some pixel or region similarity measure in relation to their local neighbourhood. Boundary-based methods search for the most dissimilar pixels which represent discontinuities in the image, while region based methods on the contrary

search for the most similar areas. These similarity measures in texture segmentation methods use some textural spatial-spectral-temporal features such as Markov random field statistics (MRF) (Haindl et al., 2004, 2005, 2006a), cooccurrence matrix based features, Gabor features, local binary pattern (LBP) (Ojala et al., 1999), autocorrelation features and many others. Segmentation methods can be categorized using various criteria, e.g. region / boundary based, MAP / clustering methods, graph theoretic methods, etc.

## 2.1 Region growing

The basic approach of a region growing algorithm (Pal et al., 1993; Belongie et al., 1998; Deng et al., 2001, 2004; Scarpa et al., 2006, 2007) is to start from a seed regions (mostly one or few pixels) that are assumed to be inside the object to be segmented. The neighbouring pixels to every seed region are evaluated to decide if they should be considered part of the object or not. If they are recognized as similar, they are added to the region and the process continues as long as any undecided pixels remain. Region growing algorithms vary depending on the similarity criteria, seed region selection, the type connectivity used to determine neighbours, and the strategy used to visit neighbouring pixels.

The JSEG method (Deng et al., 2001) consists of two independent steps: colour quantization and region growing spatial segmentation on multiscale thematic maps from the first step.

The Blobworld scheme aims to transform images into a small set of regions which are coherent in colour and texture (Belongie et al., 1998). This is achieved by clustering pixels in a joint colour-texture-position eight-dimensional feature space using the EM algorithm. The feature vector is represented by a Gaussian mixture model.

## 2.2 Split and merge

Split and merge techniques (Pal et al., 1993) start with recursive splitting image into smaller regions until they do not satisfy some homogeneity criterion. The second merging step merges adjacent regions with similar attributes.

## 2.3 Watershed

Watershed segmentation (Shafarenko et al., 1997) classifies pixels into regions using gradient descent on image features and analysis of weak points along region boundaries. The image feature space is treated, using a suitable mapping, as a topological surface where higher values indicate the presence of boundaries in the original image data. It uses analogy with water gradually filling low lying landscape basins. The size of the basins grow with increasing amounts of water until they spill into one another. Small basins (regions) gradually merge together into larger basins. Regions are formed by using local geometric structure to associate the image domain features with local extremes measurement. Watershed techniques produce a hierarchy of segmentations, thus the resulting segmentation has to be selected using either some prior knowledge or manually. These methods are well suited for different measurements fusion and they are less sensitive to user defined thresholds.

## 2.4 Level set segmentation

The paradigm of the level set (Brox et al., 2006) is that it is a numerical method for tracking the evolution of contours and surfaces. Instead of manipulating the contour directly, the contour is embedded as the zero level set of a higher dimensional function called the level-

set function. The level-set function is evolved under the control of a differential equation using some image-based features. At any time, the evolving contour can be obtained by extracting the zero level-set from the output. Level sets allow to model arbitrarily complex shapes and topological changes such as merging and splitting are handled implicitly.

## 2.5 Mean shift segmentation

Edison (Christoudias et al., 2002) is a mean shift based image segmentation with embedded edge information. Its first filtering step uses the mean shift (Comaniciu et al., 2002) segmenter in the combined colour $L*u*v*$ and coordinate feature space. The mean shift weights are derived from the edge confidence measure. The second fusion step recursively fuses the basins of attraction of the modes. The method requires six segmentation parameters to be tuned.

## 2.6 Graph-theoretic segmentation

These methods (Felzenszwalb et al., 1998; Shi et al., 2000; Boykov et al., 2003; Galun et al., 2003; Barbu et al., 2004; Estrada et al., 2005) use graph representation for image pixels or regions where usually small neighbourhood elements are mutually connected with weighted graph edges. These weights indicate pairwise elements similarities. The segmentation is based on finding groups of nodes that are strongly connected to each other, but weakly with the remaining nodes in the graph.

# 3. GMRF-GM segmenter

The adequate representation of general static Lambertian multispectral textures requires three dimensional models. Although full 3D models allows unrestricted spatial-spectral correlation description its main drawback is large amount of parameters to be estimated and in the case of Markov random field based models (MRF) also the necessity to estimate all these parameters simultaneously. Alternatively, it is possible to factorize the 3D static texture space into several (equal to the number $d$ of spectral bands) 2D subspaces. A combination of several simpler 2D data models with less parameters per model allows more compact texture representation and faster estimation algorithms. Natural measured texture data space can be decorrelated only approximately thus the independent spectral component representation suffers with some loss of image information. However, because the segmentation is less demanding application than the texture synthesis, it is sufficient if such a representation maintains discriminative power of the full model even if its visual modeling strength is slightly compromised. The GMRF-GM segmenter (Haindl et al., 2004) uses such 2D generative Gaussian Markov representation.

## 3.1 Spectral factorization

Spectral factorization using the Karhunen-Loeve expansion transforms the original centered data space $\theta$ defined on the rectangular $M \times N$ finite lattice $I$ into a new data space with K-L coordinate axes $\tilde{Y}$. This new basis vectors are the eigenvectors of the second-order statistical moments matrix

$$\Phi = E\{\tilde{Y}_r \tilde{Y}_r^T\} \tag{1}$$

where the multiindex $r$ has two components $r = [r_1, r_2]$, the first component is row and the second one column index, respectively. The projection of the centered random vector $\tilde{Y}_r$ onto the K-L coordinate system uses the transformation matrix $T = [u_1^T, u_2^T, \ldots u_d^T]^T$ which has single rows $u_j$ that are eigenvectors of the matrix $\Phi$.

$$\bar{Y}_r = T\tilde{Y}_r \tag{2}$$

Components of the transformed vector $\bar{Y}_r$ (2) are mutually uncorrelated. If we assume further on Gaussian vectors $\bar{Y}_r$ then they are also independent, i.e.,

$$p(\bar{Y}_r) = \prod_{k=1}^{d} p(\bar{Y}_{r,k})$$

and single monospectral random fields can be modeled independently.

## 3.2 GMRF texture model

We assume that single monospectral texture factors $(Y_r = \bar{Y}_{r,k})$ can be modeled using a Gaussian Markov random field model (GMRF). This model is obtained if the local conditional density of the MRF model is Gaussian:

$$p(Y_r \mid Y_{r-s} \; \forall s \in I_r) = (2\pi\sigma^2)^{-\frac{1}{2}} e^{-\frac{1}{2}\sigma^{-2}(Y_r - \bar{\mu}_r)^2} \, ,$$

where the mean value is

$$E\{Y_r \mid Y_{r-s} \; \forall s \in I_r\} = \bar{\mu}_r = \mu_r + \sum_{s \in I_r} a_s (Y_{r-s} - \mu_{r-s}) \tag{3}$$

and $\sigma$, $a_s \, \forall s \in I_r$ are unknown parameters.

The 2D GMRF model can be expressed as a stationary non-causal correlated noise driven 2D autoregressive process:

$$Y_r = \sum_{s \in I_r} a_s Y_{r-s} + e_r \tag{4}$$

where the noise $e_r$ is random variable with zero mean $E\{e_r\} = 0$ . The $e_r$ noise variables are mutually correlated

$$R_e = E\{e_r e_{r-s}\} = \begin{cases} \sigma^2 & \text{if } s = (0,0), \\ -\sigma^2 a_s & \text{if } s \in I_r, \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Correlation functions have the symmetry property $E\{e_r e_{r+s}\} = E\{e_r e_{r-s}\}$ hence the neighbourhood support set $I_r$ and its associated coefficients have to be symmetric, i.e., $s \in I_r \Rightarrow -s \in I_r$ and $a_s = a_{-s}$ .

The selection of an appropriate GMRF model support is important to obtain good results in modeling of a given random field. If the contextual neighbourhood is too small it can not capture all details of the random field. Inclusion of the unnecessary neighbours on the other hand add to the computational burden and can potentially degrade the performance of the

model as an additional source of noise. We use hierarchical neighbourhood system $I_r$, e.g., the first-order neighbourhood is $I_r = \{-(0, 1),+(0, 1),-(1, 0),+(1, 0)\}$, etc. An optimal neighbourhood is detected using the correlation method (Haindl et al., 1997) favoring neighbours locations corresponding to large correlations over those with small correlations. Parameter estimation of a MRF model is complicated by the difficulty associated with computing the normalization constant. Fortunately the GMRF model is an exception where the normalization constant is easy obtainable however either Bayesian or ML estimate requires iterative minimization of a nonlinear function. Therefore we use the pseudo-likelihood estimator which is computationally simple although not efficient. The pseudo-likelihood estimate for $a_s$ parameters evaluated for a sublattice $J_r \subset I$ and $J_r = \{s : | r_1 - s_1| \le m \wedge | r_2 - s_2| \le n\}$ centered on the $r$ index. The pseudo-likelihood estimate for $a_s$ parameters has the form

$$\gamma_r = [a_s \quad \forall s \in I_r] = \left[ \sum_{\forall s \in J_r} X_s^T X_s \right]^{-1} \sum_{\forall s \in J_r} X_s^T Y_s \, , \qquad (6)$$

where $X_s = [Y_{s+t} \quad \forall t \in I_s]$.

### 3.3 Mixture model based segmentation

Multi-spectral texture segmentation is done by clustering in the GMRF parameter space $\Theta \in \mathbb{R}^n$ defined on the lattice $I$ where

$$\Theta_r = [\gamma_{r,1}, \zeta_{r,1}, \gamma_{r,2}, \zeta_{r,2}, \cdots \gamma_{r,d}, \zeta_{r,d}]^T \; . \qquad (7)$$

$\gamma_{r,i}$ is the parameter vector (6) computed for the i-th transformed spectral band for the lattice location r and $\zeta_{r,i}$ is the average local spectral value. We assume that this parametric space can be represented using the Gaussian mixture model with diagonal covariance matrices. Hence the GMRF parametric space is first decorrelated using the Karhunen-Loeve transformation (analogously to (1)-(2)). The Gaussian mixture model for GMRF parametric representation is as follows:

$$p(\Theta_r) = \sum_{i=1}^{K} p_i \, p(\Theta_r \,|\, \nu_i, \Sigma_i) \qquad (8)$$

$$p(\Theta_r \,|\, \nu_i, \Sigma_i) = \frac{|\Sigma_i|^{-\frac{1}{2}}}{(2\pi)^{\frac{n}{2}}} \, e^{- \frac{(\Theta_r - \nu_i)^T \Sigma_i^{-1} (\Theta_r - \nu_i)}{2}} . \qquad (9)$$

The mixture equations (8),(9) are solved using the modified EM algorithm. The algorithm is initialized using $\nu_i$, $\Sigma_i$ statistics estimated from the corresponding rectangular subimages obtained by regular division of the input texture mosaic. An alternative initialization can be random choice of these statistics. For each possible couple of rectangles the Kullback Leibler divergence

$$D\left(p(\Theta_r \,|\, \nu_i, \Sigma_i) \,||\, p(\Theta_r \,|\, \nu_j, \Sigma_j)\right) = \int_{\Omega} p(\Theta_r \,|\, \nu_i, \Sigma_i) \, \log\left( \frac{p(\Theta_r \,|\, \nu_i, \Sigma_i)}{p(\Theta_r \,|\, \nu_j, \Sigma_j)} \right) \, d\Theta_r \quad (10)$$

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

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

are merged together in each step. This initialization results in $K_{ini}$ subimages and recomputed statistics $v_i$, $\Sigma_i$. $K_{ini} > K$ where $K$ is the optimal number of textured segments to be found by the algorithm. After initialization two steps of the EM algorithm are repeating:

$$E: \quad p^{(t)}(\omega_i \,|\, \Theta_r) = \frac{p_i \, p(\Theta_r \,|\, v_i, \Sigma_i)}{\sum_{j=1}^{K} p_j \, p(\Theta_r \,|\, v_j, \Sigma_j)}$$

$$M: \qquad\qquad\qquad \forall j = 1, \ldots, K$$

$$p_j^{(t+1)} \quad = \quad \frac{1}{|I|} \sum_{\forall \Theta_r} p^{(t)}(\omega_j \,|\, \Theta_r)$$

$$v_j^{(t+1)} \quad = \quad \frac{\sum_{\forall \Theta_r} \Theta_r \, p^{(t)}(\omega_j \,|\, \Theta_r)}{\sum_{\forall \Theta_r} p^{(t)}(\omega_j \,|\, \Theta_r)}$$

$$\Sigma_j^{(t+1)} = \frac{\sum_{\forall \Theta_r} p(\omega_j \,|\, \Theta_r) \left( \Theta_r - v_j^{(t+1)} \right) \left( \Theta_r - v_j^{(t+1)} \right)^T}{\sum_{\forall \Theta_r} p^{(t)}(\omega_j \,|\, \Theta_r)} \quad . \tag{11}$$

The components with smaller weights $p_j < \xi$ than a given threshold are eliminated. For every pair of components we estimate their Kullback Leibler divergence (10). From the most similar couple, the component with the weight smaller than the threshold is merged to its stronger partner and all statistics are actualized 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.

The parametric vectors representing texture mosaic pixels are assigned to the clusters according to the highest component probabilities, i.e., $Y_r$ is assigned to the cluster $\omega_j$ if

$$\arg \max_j \sum_{s \in Neigh} weight(s) \, p(\Theta_{r-s} \,|\, v_j, \Sigma_j) \tag{12}$$

$$p(\Theta_r \,|\, v_j, \Sigma_j) > p(\Theta_r \,|\, v_i, \Sigma_i) \ \forall i \neq j \ .$$

The area of single cluster blobs is evaluated in the post-processing thematic map filtration step. Thematic map blobs with area smaller than a given threshold are attached to its neighbour with the highest similarity value. If there is no similar neighbour the blob is eliminated. After all blobs are processed remaining blobs are expanded.

## 4. AR3D-GM segmenter

If we do not like to lose spectral information due to the spectral decorrelation step, we have to use three dimensional models for adequate representation. One of few 3D models which

does not require any approximation and can be treated analytically is the 3D simultaneous causal autoregressive random field model (CAR) used in the AR3D-GM segmenter (Haindl et al., 2006a).

We assume that single multispectral textures can be locally modeled using the CAR model. This model can be expressed as a stationary causal uncorrelated noise driven 3D autoregressive process (Haindl et al., 1992):

$$Y_r = \gamma X_r + e_r \ , \tag{13}$$

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 neighbourhood index set with $\eta = \text{card}(I_r^c)$ 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 (Haindl et al., 1992). The recursive Bayesian parameter estimation of the CAR model is (Haindl et al., 1992):

$$\hat{\gamma}_{r-1}^T = \hat{\gamma}_{r-2}^T + \frac{V_{x(r-2)}^{-1} X_{r-1} (Y_{r-1} - \hat{\gamma}_{r-2} X_{r-1})^T}{(1 + X_{r-1}^T V_{x(r-2)}^{-1} X_{r-1})} \ , \tag{14}$$

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.,

$$\tilde{\gamma}_r^T = \{\hat{\gamma}_r^t, \hat{\gamma}_r^b, \hat{\gamma}_r^r, \hat{\gamma}_r^l\}^T \ . \tag{15}$$

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 $\zeta_r$ to the resulting feature vector ($\Theta_r$).

## 4.1 AR2D-GM segmenter

The AR2D-GM segmenter (Haindl et al., 2005) uses the 2D simultaneous causal autoregressive random field model and thus it requires the spectral decorrelation described in section 3.1. If we stack single decorrelated mono spectral pixel components into $d \times 1$ vectors $Y_r$, the model can be formalized using the same equations as the AR3D model, i.e. (13)-(15). The AR2D models differ in having diagonal parameter matrices $A_s$ and a diagonal white noise covariance matrix.

## 4.2 Mixture based segmentation

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

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

is the modified parameter vector (15) 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 is again (8),(9) and can be solved similarly as (10)-(12).

## 5. Evaluation methodology

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 (Zhang, 1997), this problem is still far from being solved. 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 optimal alternative to check several variants of a developed method and to carefully compare results with state-of-theart in this area is practically impossible because most methods are too complicated and insufficiently described to be implemented in the acceptable time. Because there is no available benchmark fully supporting segmentation method development, we implemented a solution in the form of web based data generator and benchmark software. Proper testing and robust learning of performance characteristics require large test sets and objective ground truth which is unfeasible for natural images. Thus, inevitably all such image sets such as the Berkeley benchmark (Martin et al., 2001) share the same drawbacks - subjectively generated ground truth regions and limited extent which is very difficult and expensive to enlarge. These problems motivated our preference for random mosaics with randomly filled textures even if they only approximate natural image scenes. The profitable feature of this compromise is the unlimited number of different test images with corresponding objective and free ground truth map available for each of them.

The segmentation results can be judged (Zhang, 1997) either by using manually segmented images as reference (Lee et al., 1990), or visually by comparing to the original images (Pal et al., 1993), or just by applying quality measures corresponding to human intuition (Sahoo et al., 1988; Lee et al., 1990; Pal et al., 1993). However it is difficult to avoid subjective ranking conclusions by using either of above approaches on limited test databases.

A prior work on the segmentation benchmark is the Berkeley benchmark presented by Martin et al. (Martin et al., 2001). This benchmark contains more than 1000 various natural images (300 in its public version) from the Corel database, each of which is manually processed by a group of people to get the ground-truth segmentation in the form of partitioning of the image into a set of disjoint segments. Without any special guidance, such

manual segmentations reflect the subjective human perception and therefore, different people usually construct different ground truths on the same image. The Berkeley benchmark suffers from several drawbacks. Apart from subjective ground truth, also its performance criteria a global consistency error (GCE) and a local consistency error (LCE) tolerate unreasonable refinement of the ground truth. Over-segmented machine segmentations have always zero consistency error, i.e., they wrongly suggest an ideal segmentation. The benchmark comparison is based on region borders hence different border localization from the human based drawing can handicap otherwise correct scene segmentation.

Another segmentation benchmark Minerva (Sharma et al., 2001) contains 448 colour and grey scale images of natural scenes which are segmented using four different segmenters, segmented regions are manually labelled and different textural features can be learned from these regions and subsequently used by the kNN supervised classifier. This approach suffers from erroneous ground truth resulting from an imperfect segmenter, manual labelling and inadequate textural feature learning from small regions.

Outex Texture Database (Ojala et al., 2002) provides a public repository for three types of empirical texture evaluation test suites. It contains 14 classification test suites, while one unsupervised segmentation test set is formed by 100 texture mosaics all using the same regular ground truth template and finally one texture retrieval test set. The test suites are publicly available on the website (http://www.outex.oulu.fi), which allows searching, browsing and downloading of the test image databases. Outex currently provides limited test repository but does not allow results evaluation or algorithms ranking.

A psycho-visual evaluation of segmentation algorithms using human observers was proposed in (Shaffrey et al., 2002). The test was designed to visually compare two segmentations in each step and to answer if any consensus of the best segmentation exists. While such human judgement certainly allows meaningful evaluation, this approach is too demanding to be applicable in image segmentation research.

## 5.1 Prague texture segmentation benchmark

The Prague texture segmentation data-generator and benchmark Fig.1 is web based service (http://mosaic.utia.cas.cz) developed as a part of EU NoE no. 507752 MUSCLE project. The goal of the benchmark is to produce score, performance and quality measures for an algorithm's performance for two main reasons: So that different algorithms can be compared to each other, and so that progress toward human-level segmentation performance can be tracked and measured over time. A good experimental evaluation should allow comparison of the current algorithm to several leading alternative algorithms, using as many test images as possible and employing several evaluation measures for comparison (in the absence of one clearly optimal measure). Our benchmark possesses all these features.

Single textures as well as the mosaics generation approach were chosen on purpose to produce unusually difficult tests to allow an improvement space for future better segmentation algorithms.

The benchmark operates either in full mode for registered users (unrestricted mode - U) or in a restricted mode. The major differences between both working modes are that the restricted operational mode does not permanently store visitor's data (results, algorithm details, etc.) into its online database and does not allow custom mosaics creation. To be able to use full-unrestricted benchmark functionalities the user is required to be registered (registration page).
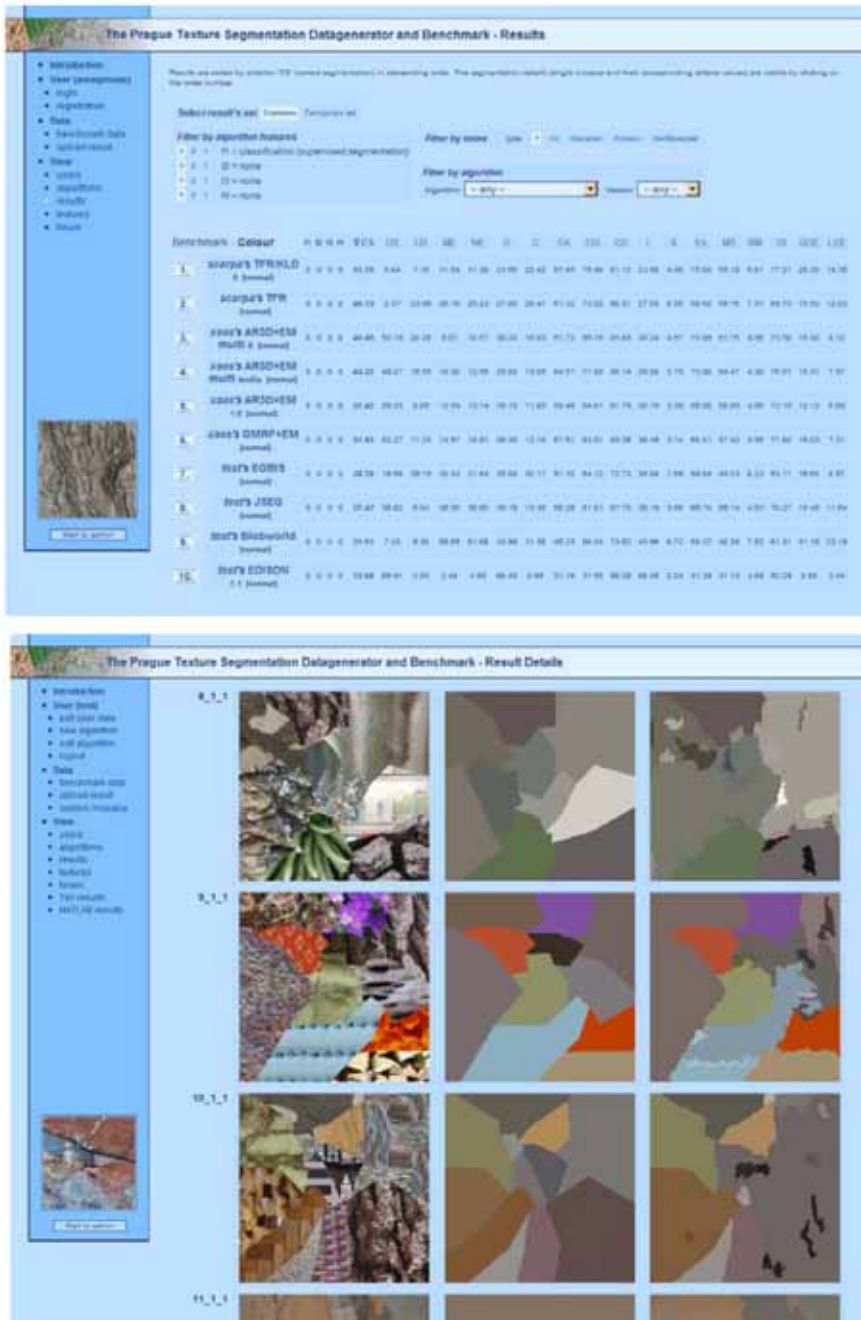
Fig. 1. Benchmark interfaces, the comparison table (top) and detailed method's evaluation on benchmark test mosaics.

The benchmark allows:

- To obtain customized experimental texture mosaics and their corresponding ground truth (U),
- To obtain the benchmark texture mosaic sets with their corresponding ground truth,
- To evaluate visitor's working segmentation results and compare them (Fig.1- top) with state-of-the-art algorithms,
- To update the benchmark database (U) with an algorithm (reference, abstract, benchmark results) and use it for subsequent other algorithms benchmarking,
- To grade noise endurance of an algorithm,
- To check single mosaics evaluation details (criteria values and resulted thematic maps),
- To rank segmentation algorithms according to the most common benchmark criteria,
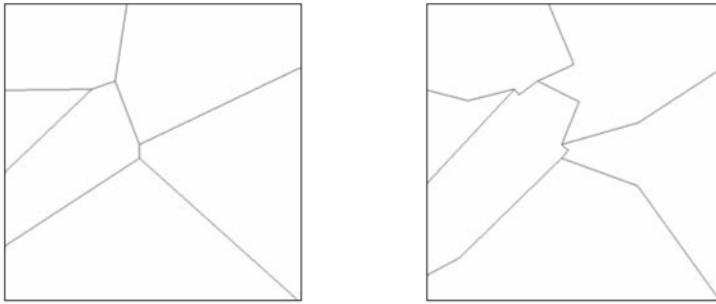- To obtain LaTeX or MATLAB coded resulting criteria tables (U).

Fig. 2. Voronoi (left) and modified (right) mosaic polygons.

### 5.2 Benchmark generation

Benchmark datasets are computer generated $512 \times 512$ random mosaics filled with randomly selected textures. The random mosaics are generated by using the Voronoi polygon random generator (Shewchuk, 1996). It creates firstly a Delaunay triangulation, secondly determines the circumcircle centres of its triangles, and thirdly connects these points according to the neighbourhood relations between the triangles. Resulting Voronoi polygons can further be modified, (see Fig.2), if required by inserting additional border points into each polygon line. We exploit the fact that segmenting smaller and irregular objects is more difficult than segmenting bigger and regular objects such as squares or circles.

### 5.3 Performance criteria

The submitted benchmark results are evaluated and stored (U) in the server database and used for the algorithm ranking according to a chosen criterion. We have implemented the twenty seven most frequented evaluation criteria categorized into four groups: region-based (5+5), pixel-wise (12), consistency measures (2) and clustering comparison criteria (3). The performance criteria mutually compare ground truth image regions with the corresponding machine segmented regions. Symbols ↑ / ↓ further denote required increase or decrease of the corresponding criterion. The basic region-based criteria available are correct, over-segmentation, undersegmentation, missed and noise. All these criteria are available either for a single threshold parameter setting or as the performance curves (e.g. Fig.3) and their

integrals. Our pixel-wise criteria group contains the most frequented classification criteria such as the omission and commission errors, class accuracy, recall, precision, mapping score, etc. The consistency criteria group incorporates the global and local consistency errors. Finally, the last criterion set contains three clustering comparison measures. By clicking on a required criterion the evaluation table is reordered, according to this chosen criterion.
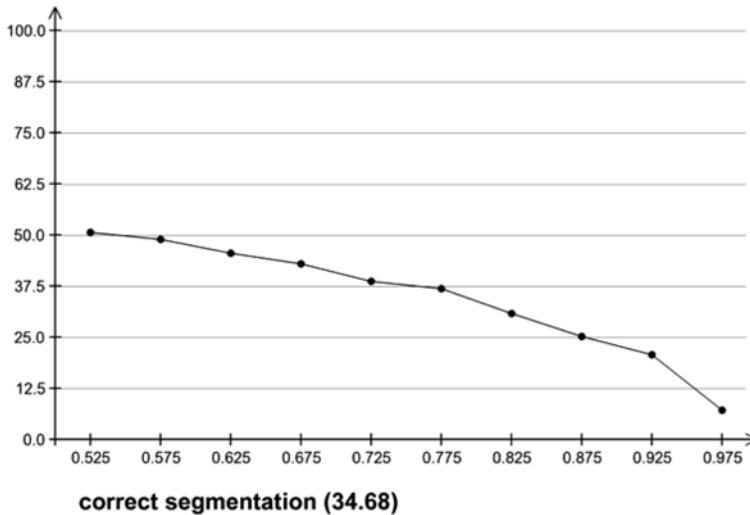


Fig. 3. The correct segmentation sensitivity (performance) curve and its integral for the AR3D-GM segmenter.

## 6. Results

Our algorithms were tested on natural colour textures mosaics from the Prague Texture Segmentation Data-Generator and Benchmark (Mikeš et al., 2006; Haindl et al., 2008). 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.

Tab.1 compares the overall benchmark performance of our algorithms (segmentation time 7 min/img on the Athlon 2GHz processor) with the Blobworld (Carson et al., 1999) (30 min/img), JSEG (Deng et al., 2001) (30 s/img), Edison (Christoudias et al., 2002) (10 s/img), respectively. These results demonstrate very good pixel-wise, correct region segmentation and low undersegmentation properties of both our methods while the oversegmentation results are only average. For all the pixel-wise criteria or the consistency measures our methods are either the best ones or the next best with marginal difference from the best one. Fig.4 shows four selected 512×512 experimental benchmark mosaics created from five to eleven natural colour textures. The last four columns demonstrate comparative results from the four alternative algorithms. Hard natural textures were chosen rather than synthesized (for example using Markov random field models) ones because they are expected to be more difficult for the underlying segmentation model. The third column demonstrates robust

| | Benchmark – Colour | | | | |
|---|---|---|---|---|---|
| | AR3D-GM | GMRF-GM | JSEG | Blobworld | EDISON |
| ↑ CS | **37.42** | 31.93 | 27.47 | 21.01 | 12.68 |
| ↓ OS | 59.53 | 53.27 | 38.62 | **7.33** | 86.91 |
| ↓ US | 8.86 | 11.24 | 5.04 | 9.30 | **0.00** |
| ↓ ME | 12.55 | 14.97 | 35.00 | 59.55 | **2.48** |
| ↓ NE | 13.14 | 16.91 | 35.50 | 61.68 | **4.68** |
| ↓ O | **34.32** | 33.61 | 37.94 | 41.45 | 73.17 |
| ↓ C | 100.00 | 100.00 | 92.77 | **58.94** | 100.00 |
| ↑ CA | **59.46** | 57.91 | 55.29 | 46.23 | 31.19 |
| ↑ CO | **64.81** | 63.51 | 61.81 | 56.04 | 31.55 |
| ↑ CC | 91.79 | 89.26 | 87.70 | 73.62 | **98.09** |
| ↓ I. | **35.19** | 36.49 | 38.19 | 43.96 | 68.45 |
| ↓ II. | 3.39 | 3.14 | 3.66 | 6.72 | **0.24** |
| ↑ EA | **69.60** | 68.41 | 66.74 | 58.37 | 41.29 |
| ↑ MS | **58.89** | 57.42 | 55.14 | 40.36 | 31.13 |
| ↓ RM | 4.88 | 4.86 | 4.96 | 7.96 | **3.21** |
| ↑ CI | **73.15** | 71.80 | 70.27 | 61.31 | 50.29 |
| ↓ GCE | 12.13 | 16.03 | 18.45 | 31.16 | **3.54** |
| ↓ LCE | 6.69 | 7.31 | 11.64 | 23.19 | **3.44** |
| ↓ dM | 15.43 | 15.27 | **15.19** | 20.03 | 16.84 |
| ↓ dD | **19.76** | 20.63 | 23.38 | 31.11 | 35.37 |
| ↓ dVI | 17.10 | 17.32 | 17.37 | **15.84** | 25.65 |
| ↑ $\overline{CS}$ | **34.68** | 31.04 | 29.13 | 19.10 | 12.95 |
| ↓ $\overline{OS}$ | 53.32 | 49.74 | 37.70 | **10.81** | 76.35 |
| ↓ $\overline{US}$ | 9.24 | 11.33 | 6.38 | 8.35 | **0.00** |
| ↓ $\overline{ME}$ | 19.90 | 21.92 | 34.72 | 58.54 | **13.91** |
| ↓ $\overline{NE}$ | 20.80 | 23.59 | 35.38 | 61.24 | **15.29** |
| ↑ $\overline{F}$ | **72.08** | 70.79 | 69.23 | 60.46 | 47.42 |

Table 1. Benchmark criteria (×100): CS = correct segmentation; OS = oversegmentation; 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;

behaviour of our GMRF-GM 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 JSEG (Deng et al., 2001), Blobworld (Carson et al., 1999) and Edison (Christoudias et al., 2002) algorithms on these data performed steadily worse as can be seen in the last two columns of Fig.4, some areas are undersegmented while other parts of the mosaics are oversegmented. The GMRF-GM (Haindl et al., 2004) method is slower and its results are surprisingly also slightly worse than the AR3D-GM results. Resulting segmentation results are promising however comparison with all state-of-the-art

algorithms is difficult because of lack of sound experimental evaluation results in this area. Our results can be further improved by an appropriate postprocessing.
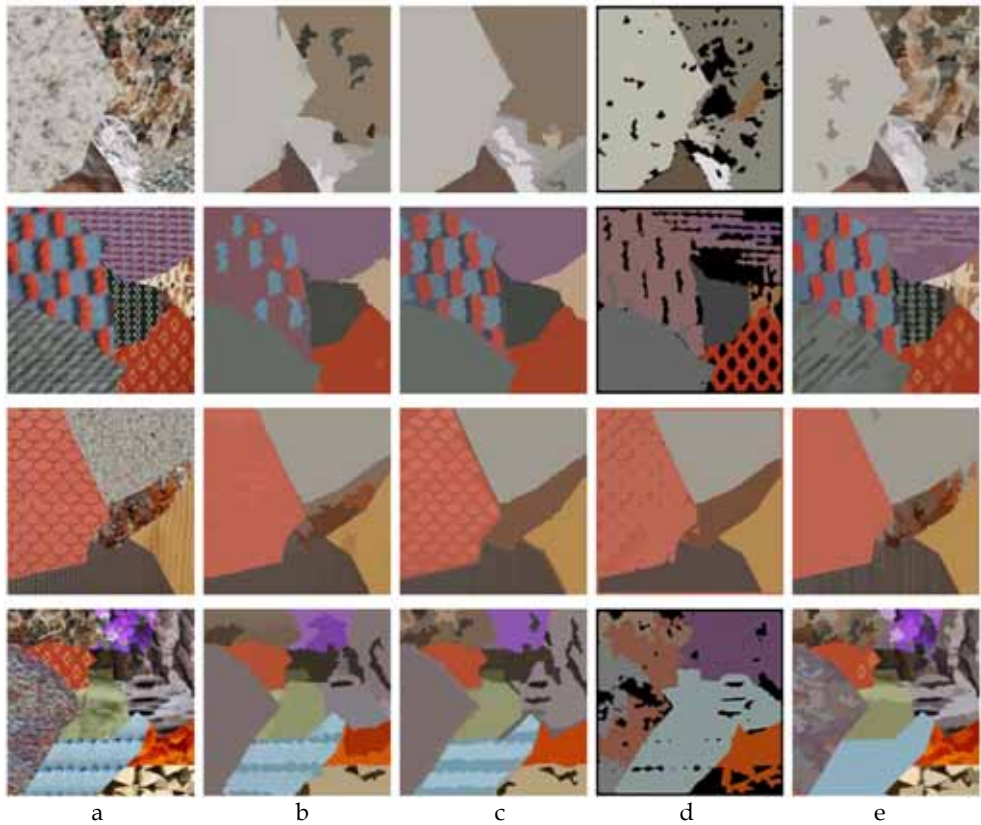


Fig. 4. Selected benchmark texture mosaics (a), AR3D-GM (b), GMRF-GM (c), Blobworld (d), and Edison segmentation results (e), respectively.

### 6.1 Mammography

Breast cancer is the leading cause of death (Tweed et al., 2002; Qi et al., 2003) among all cancers for middle-aged women in most developed countries. Current effort is focused on cancer prevention and early detection which can significantly reduce the mortality rate. X-ray screening mammography is the most frequented method for breast cancer early detection although not without problems (Qi et al., 2003) such as rather large minimum detectable tumor size, higher mammogram sensitivity for older women or radiation exposition. Automatic mammogram analysis is still difficult task due to wide variation of breast anatomy, nevertheless a computer-aided diagnosis system can successfully assist a radiologist, and can be used as a second opinion. The first step in a such system is detection of suspicious potentially cancerous regions of interest . Several approaches to detect these regions of interest (ROI) were published (Tweed et al., 2002), mostly based on supervised learning. We propose an unsupervised segmentation method for fast automatic

mammogram segmentation into the regions of interest (ROI) using a statistical random field based texture representation.

The presented method detects the fibroglandular tissue regions from either craniocaudal (CC) or mediolateral oblique (MLO) views and thus can help focus a radiologist to this most important breast region.

**Breast Detector** The method starts with automatic breast area detection because it can be cheaply computed and simplifies the subsequent fibroglandular tissue region detection. This is performed using simple histogram thresholding with an automatically selected threshold. In this step the method also recognizes several label areas on a mammogram. We compute their areas and all but the largest one are discarded and merged with the
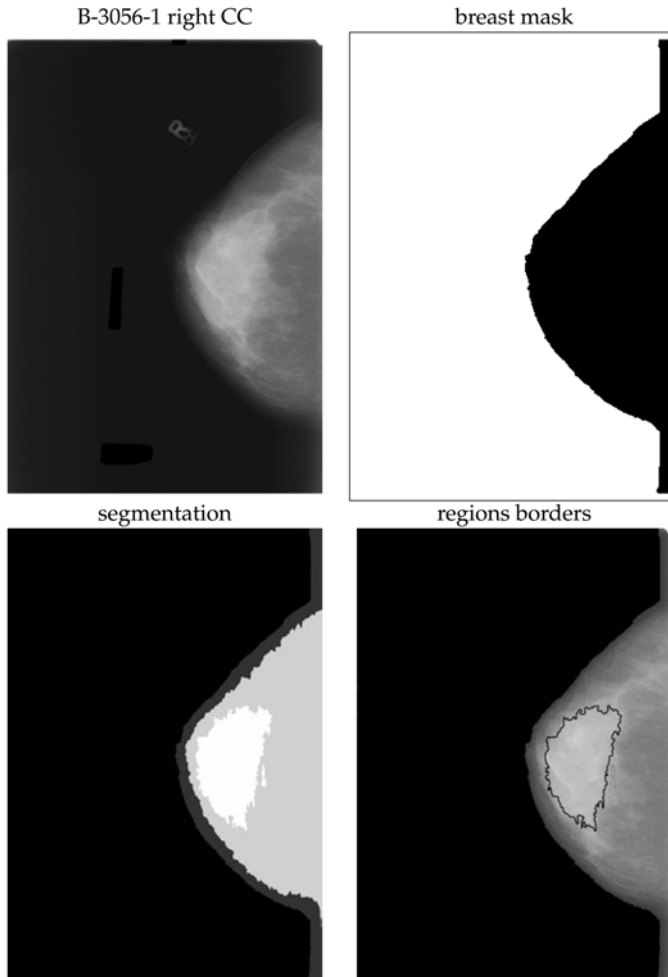


Fig. 5. Normal right breast mammogram (patient age 58, but with a cancerous lesion in the left breast), the detected breast area, segmentation result and detected regions of interest, respectively.

background. In this stage the algorithm also decides the breast orientation on the mammogram (left or right). Fig. 5 - breast mask show resulting detected breast area (in inverted grey levels). The following detection of regions of interest is performed only in the breast region ignoring the background area set in the mask template.

**Breast Tissue Texture Model** Our method segments pseudo-colour multiresolution mammograms each created from the original greyscale mammogram and its two nonlinear gamma transformations. We assume to down-sample input image $Y$ into $M = 3$ different resolutions $Y^{(m)} = \downarrow^{\iota_m} Y$ with sampling factors $\iota_m$ $m = 1, \ldots, M$ identical for both directions and $Y^{(1)} = Y$. Local texture for each pixel $Y_r^{(m)}$ is represented using the 3D CAR model parameter space $\Theta_r^{(m)}$. The concept of decision fusion (Kittler et al., 1997) 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
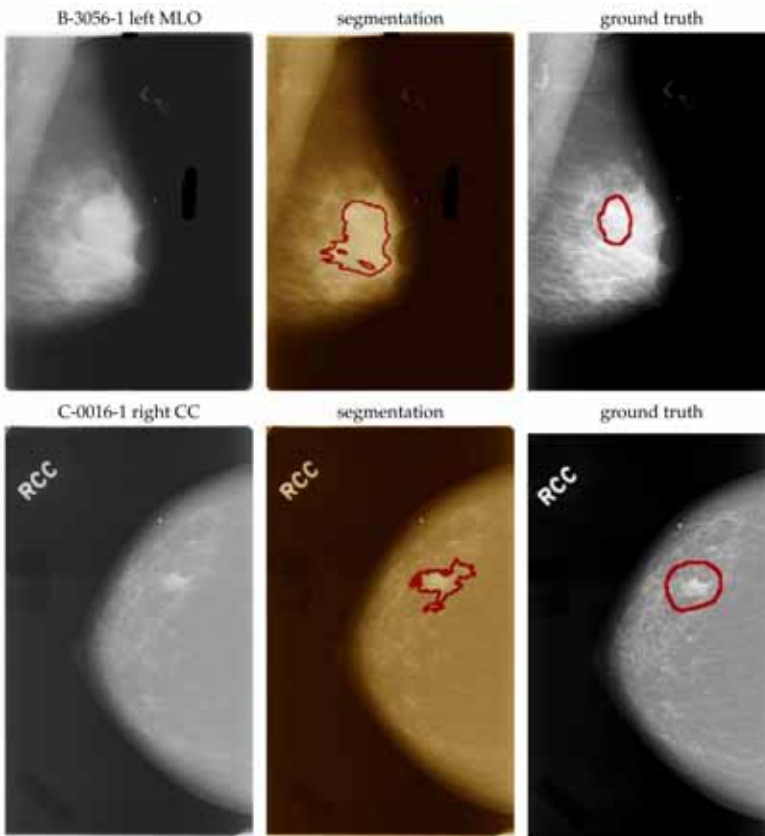


Fig. 6. Cancerous mammograms (patients age 58 (top) and 80 (bottom)), radiologist associated ground truth and detected regions of interest using the multiple segmenter approach, respectively.

classifier applications. The proposed method circumvents the problem of multiple unsupervised segmenters combination (Haindl et al., 2007) by fusing multiple-processed measurements into a single segmenter feature vector.

**Experimental Results** The algorithm was tested on mammograms from the Digital Database for Screening Mammography (DDSM) from the University of South Florida (Heath et al., 2000). This database contains 2620 four view (left and right craniocaudal (CC) and mediolateral oblique (MLO)) mammograms in different resolutions. Single mammograms cases are divided into normal, benign, benign without callback volumes and cancer. All our experiments are done with three resolutions ($M$=3) using sampling factors $\iota_1 = 2, \iota_2 = 4, \iota_3 = 8$ and the causal neighbourhood with fourteen neighbours ($\eta$= 14). Fig. 6-top show left MLO mammogram of a patient age 58 with detected malignant asymmetric lesion and the right CC mammogram (Fig. 6-bottom) of a patient age 80 with detected irregular, spiculated malignant lesion type. The segmenter correctly found the region of interest with the cancer lesion on both mammograms. The detected region of interest results Figs. 5-6 demonstrate very good region segmentation and low oversegmentation properties of our method.

### 6.2 Remote sensing

Segmentation of remote sensing imagery for various applications (e.g. agriculture, geological survey, military and security, weather forecast, terrain classification, astronomy, the detection of changes and anomalies, etc.) is challenging task due to huge amounts of data measured by satellite or airborne sensors. Large remote sensing images suffer not only with geometric and radiometric distortions problems but also with various challenges due to the high heterogeneity both within and across classes. The within class heterogeneity is due to the difference of acquisition process, orientation, and intrinsic appearance (Fauqueur et al., 2005).

We modified our unsupervised segmentation methods (sections 3,4) to be able to handle large aerial images (up to 8000 × 8000) distributed by the British National Space Centre (BNSC) as a CDROM called "Window On The UK". These aerial images (Fig.7) cover both urban and rural areas of the United Kingdom. The parametric space Θ (7) build over large images from this set requires efficient memory handling and distance based region class merging to avoid expensive memory swapping during the segmentation. Segmentation results illustrated on Fig.7-bottom do not use any prior information except the minimal region area. This parameter can be easily determined from the image resolution and the intended thematic map application.

## 7. Conclusions

We discussed three efficient and robust methods for unsupervised texture segmentation with unknown number of classes based on the underlying Markovian and GM texture models and their modifications for medical mammographics and remote sensing applications, respectively. Although these algorithm use the random field type models they are fast because they use efficient recursive or pseudo-likelihood parameter estimation of the underlying texture models and therefore they are much faster than the usual Markov

Fig. 7. Aerial Lmw 4800 × 4800 image (top left), its detail (top right), the corresponding unsupervised segmentation (bottom left) and parrac 8000 × 8000 image segmentation result, respectively.

chain Monte Carlo estimation approach. Usual handicap of segmentation methods is their lot of application dependent parameters to be experimentally estimated. Our methods require only a contextual neighbourhood selection and two additional thresholds. Their performance is demonstrated on the extensive benchmark tests on natural texture mosaics. They perform favorably compared with four alternative segmentation algorithms. The AR-GM methods are faster than the GMRF-GM method. These segmenters allow to build efficient methods for completely automatic unsupervised detection of mammogram fibroglandular tissue regions of interest which can assist a radiologist, and can be used as a second opinion. Alternatively our segmenters can be used to detect meaningful areas in large remote sensing images and in various other image segmentation applications.

## 8. Acknowledgements

## 9. References

Andrey, P. & Tarroux, P. (1998). Unsupervised segmentation of markov random field modeled textured images using selectionist relaxation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3), 252–262.

Barbu, A. & Zhu, S. C. (2004). Multigrid and multi-level swendsen-wang cuts for hierarchic graph partition. In *CVPR (2)*, volume 2 pp. 731–738.

Belongie, S.; Carson, C.; Greenspan, H., & Malik, J. (1998). Color and texture-based image segmentation using em and its application to content-based image retrieval. In *International Conference on Computer Vision*: IEEE.

Boykov, Y. & Kolmogorov, V. (2003). Computing geodesics and minimal surfaces via graph cuts. In *ICCV03* pp. 26–33. Nice, France.

Brox, T. & Weickert, J. (2006). Level set segmentation with multiple regions. *IEEE Trans. Image Processing*, 15(10), 3213–3218.

Carson, C.; Thomas, M.; Belongie, S.; Hellerstein, J. M., & Malik, J. (1999). Blobworld: A system for region-based image indexing and retrieval. In *Third International Conference on Visual Information Systems*: Springer.

Christoudias, C.; Georgescu, B., & Meer, P. (2002). Synergism in low level vision. In R. Kasturi, D. Laurendeau, & C. Suen (Eds.), *Proceedings of the 16th International Conference on Pattern Recognition*, volume 4 pp. 150–155. Los Alamitos: IEEE Computer Society.

Comaniciu, D. & Meer, P. (2002). Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5), 603–619.

Deng, H. & Clausi, D. A. (2004). Unsupervised image segmentation using A simple MRF model with A new implementation scheme. *Pattern Recognition*, pp. 2323–2335.

Deng, Y. & Manjunath, B. (2001). Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(8), 800–810.

Estrada, F. & Jepson, A. (2005). Quantitative evaluation of a novel image segmentation algorithm. In *CVPR*, volume II pp. 1132–1139.: IEEE.

Fauqueur, J.; Kingsbury, N. G., & Anderson, R. (2005). Semantic discriminant mapping for classification and browsing of remote sensing textures and objects. In *ICIP (2)* pp. 846–849.

Felzenszwalb, P. & Huttenlocher, D. (1998). Image segmentation using local variation. In *CVPR* pp. 98–104.: IEEE.

Galun, M.; Sharon, E.; Basri, R., & Brandt, A. (2003). Texture segmentation by multiscale aggregation of filter responses and shape elements. In *International Conference on Computer Vision* pp. 716–723.

Haindl, M. (1991). Texture synthesis. *CWI Quarterly*, 4(4), 305–331.

Haindl, M. (1999). Texture segmentation using recursive markov random field parameter estimation. In K. Bjarne & J. Peter (Eds.), *Proceedings of the 11th Scandinavian Conference on Image Analysis* pp. 771–776. Lyngby, Denmark: Pattern Recognition Society of Denmark.

Haindl, M. & Havlíček, V. (1997). *Prototype Implementation of the Texture Analysis Objects*. Technical Report 1939, ÚTIA AV Č R, Praha, Czech Republic.

Haindl, M. & Mikeš, S. (2005). Colour texture segmentation using modelling approach. *Lecture Notes in Computer Science*, (3687), 484–491.

Haindl, M. & Mikeš, S. (2006a). Unsupervised texture segmentation using multispectral modelling approach. In Y. Tang, S.Wang, D. Yeung, H. Yan, & G. Lorette (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, volume II pp. 203–206. Los Alamitos: IEEE Computer Society.

Haindl, M. & Mikeš, S. (2007). Unsupervised texture segmentation using multiple segmenters strategy. In M. Haindl, J. Kittler, & F. Roli (Eds.), *MCS 2007*, volume 4472 of *Lecture Notes in Computer Science* pp. 210–219.: Springer.

Haindl, M. & Mikeš, S. (2008). Texture segmentation benchmark. In B. Lovell, D. Laurendeau, & R. Duin (Eds.), *Proceedings of the 19th International Conference on Pattern Recognition, ICPR 2008* pp. accepted. Los Alamitos: IEEE Computer Society.

Haindl, M. & Mikeš, S. (2004). Model-based texture segmentation. *Lecture Notes in Computer Science*, (3212), 306 – 313.

Haindl, M. & Šimberová, S. (1992). *Theory & Applications of Image Analysis*, chapter A Multispectral Image Line Reconstruction Method, pp. 306–315. World Scientific Publishing Co.: Singapore.

Haindl, M. & Vácha, P. (2006b). Illumination invariant texture retrieval. In Y. Tang, S. Wang, D. Yeung, H. Yan, & G. Lorette (Eds.), *Proceedings of the 18th International Conference on Pattern Recognition, ICPR 2006*, volume III pp. 276–279. Los Alamitos: IEEE Computer Society.

Heath, M.; Bowyer, K.; Kopans, D.; Moore, R., & Kegelmeyer, P. (2000). The digital database for screening mammography. In *Proc. of the 5th Int. Workshop on Digital Mammography*: Medical Physics Publishing.

Kashyap, R. (1986). Image models. In K. F. T.Y. Young (Ed.), *Handbook of Pattern Recognition and Image Processing*. New York: Academic Press.

Kittler, J.; Hojjatoleslami, A., & Windeatt, T. (1997). Weighting factors in multiple expert fusion. In *Proc. BMVC* pp. 41–50.: BMVA BMVA.

Lee, S. U.; Chung, S. Y., & Park, R. H. (1990). A comparitive performance study of several global thresholding techniques for segmentation. *Computer Vision, Graphics, and Image Processing*, 52, 171–190.

Manjunath, B. & Chellapa, R. (1991). Unsupervised texture segmentation using markov random field models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13, 478–482.

Martin, D.; Fowlkes, C.; Tal, D., & Malik, J. (2001). 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*, volume 2 pp. 416–423.

Mikeš, S. & Haindl, M. (2006). Prague texture segmentation data generator and benchmark. *ERCIM News*, (64), 67–68.

Ojala, T.; Maenpaa, T.; Pietikainen, M.; Viertola, J.; Kyllonen, J., & Huovinen, S. (2002). Outex: New framework for empirical evaluation of texture analysis algorithms. In *International Conference on Pattern Recognition* pp. I: 701–706.

Ojala, T. & Pietikainen, M. (1999). Unsupervised texture segmentation using feature distributions. *Pattern Recognition*, 32(477-486).

Pal, N. & Pal, S. (1993). A review on image segmentation techniques. *Pattern Recognition*, 26(9), 1277–1294.

Panjwani, D. & Healey, G. (1995). Markov random field models for unsupervised segmentation of textured color images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(10), 939–954.

Qi, H. & Diakides, N. A. (2003). Thermal infrared imaging in early breast cancer detection - a survey of recent research. In *25th Annual Int. Conference of the IEEE EMBS* pp. 448–452.

Reed, T. R. & du Buf, J. M. H. (1993). A review of recent texture segmentation and feature extraction techniques. *CVGIP–Image Understanding*, 57(3), 359–372.

Sahoo, P. K.; Soltani, S., & Wong, A. K. C. (1988). Survey of thresholding techniques. *Computer Vision, Graphics and Image Processing*, 41(2), 233–260.

Scarpa, G. & Haindl, M. (2006). Unsupervised texture segmentation by spectralspatial-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*, volume II pp. 151–154. Los Alamitos: IEEE Computer Society.

Scarpa, G.; Haindl, M., & Zerubia, J. (2007). A hierarchical texture model for unsupervised segmentation of remotely sensed images. *Lecture Notes in Computer Science*, 4522, 303–312.

Shafarenko, L.; Petrou, M., & Kittler, J. (1997). Automatic watershed segmentation of randomly textured color images. *IEEE Trans. on Image Processing*, 6(11), 1530–1544.

Shaffrey, C. W.; Jermyn, I. H., & Kingsbury, N. G. (2002). Psychovisual evaluation of image segmentation algorithms. In *ACIVS 2002*.

Sharma, M. & Singh, S. (2001). Minerva scene analysis benchmark. In *Seventh Australian and New Zealand Intelligent Information Systems Conference* pp. 231–235.: IEEE.

Shewchuk, J. R. (1996). Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. In M. C. Lin & D. Manocha (Eds.), *Applied Computational Geometry: Towards Geometric Engineering*, volume 1148 of *Lecture Notes in Computer Science* pp. 203–222. Springer-Verlag. From the First ACMWorkshop on Applied Computational Geometry.

Shi, J. & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(8), 888–905.

Tweed, T. & Miguet, S. (2002). Automatic detection of regions of interest in mammographies based on a combined analysis of texture and histogram. In *ICPR*, volume 02 pp. 448–452. Los Alamitos, CA, USA: IEEE Computer Society.

Vacha, P. & Haindl, M. (2007). Image retrieval measures based on illumination invariant textural mrf features. In *CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval* pp. 448–454. New York, NY, USA: ACM Press.

Zhang, Y. J. (1997). Evaluation and comparison of different segmentation algorithms. *Pattern Recognition Letters*, 18, 963–974.