

# Markovian Image Models and their Applications in Unsupervised Image Segmentation

Zoltan Kato

**Abstract**—In this report, we present the main results of our work supported by the OTKA K-46805 grant during 2004–2006:

- 1) We have proposed a monogrid MRF model which is able to combine color and texture features in order to improve the quality of segmentation results. We have also solved the estimation of model parameters [1].
- 2) We have proposed a novel RJMCMC sampling method which is able to identify multi-dimensional Gaussian mixtures. Using this technique, we have developed a fully automatic color image segmentation algorithm [2], [3].
- 3) A new multilayer MRF model has been proposed which is able to segment an image based on multiple cues (such as color, texture, or motion) [4].
- 4) A new shape prior, called 'gas of circles' has been introduced and applied to tree crown segmentation using active contour models [5], [6].

## I. UNSUPERVISED SEGMENTATION: A PROBABILISTIC APPROACH

The simplest statistical model for an image consists of the probabilities of pixel classes. The knowledge of the dependencies between nearby pixels can be modeled by a Markov random Field (MRF). Such models are quite powerful even if it is not easy to determine the values of the parameters which specify a MRF. If each pixel class is represented by a different model then the observed image may be viewed as a sample from a realization of an underlying label field. Unsupervised segmentation can therefore be treated as an *incomplete data problem* where the pixel values are observed, the label field is missing and the associated class model parameters, including the number of classes, need to be estimated. Due to the difficulty of estimating the number of pixel classes (or clusters), unsupervised algorithms often assume that this parameter is *known a priori* [1], [4]. When the number of pixel classes is also being estimated, the unsupervised segmentation problem may be treated as a *model selection problem* over a combined model space.

Our approach [1]–[3] consists of building a Bayesian image model using a first order MRF. The observed image is represented by a mixture of multivariate Gaussian distributions while inter-pixel interaction favors similar labels at neighboring sites. In a Bayesian framework, we are interested in the *posterior distribution* of the unknowns given the observed image. The model assumes that the real world scene consists of a set of regions whose observed features  $\mathcal{F}$  (such as color, texture, or motion) changes slowly, but across the boundary between them, they change abruptly. What we want to infer is a *labeling*  $\omega$  consisting of a simplified, abstract version of the input image: regions has a constant value (called a *label* in our context) and the discontinuities between them form a curve - the contour. Such a labeling  $\omega$  specifies a *segmentation*. Taking the probabilistic approach, one usually wants to come up with a *probability measure* on the set  $\Omega$  of all possible segmentations of the input image and then select the one with the highest probability. Note that  $\Omega$  is finite, although huge. A widely accepted standard, also motivated by the human visual system, is to construct this probability measure in a Bayesian framework. First, we have to quantify how well any occurrence of  $\omega$  fits  $\mathcal{F}$ . This is expressed by the probability distribution  $P(\mathcal{F}|\omega)$  - the *imaging model*. Second, we define a set of properties that any segmentation  $\omega$  must possess regardless the image

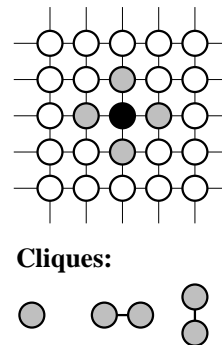


Fig. 1. First order neighborhood system with corresponding cliques [1]–[3].

data. These are described by  $P(\omega)$ , the *prior*, which tells us how well any occurrence  $\omega$  satisfies these properties. For that purpose,  $\omega_s$  is modeled as a discrete random variable taking values in the set of labels  $\Lambda = \{1, 2, \dots, L\}$ . The set of these labels  $\omega = \{\omega_s, s \in \mathcal{S}\}$  is a random field, called the *label process*. Furthermore, the observed color features are supposed to be a realization  $\mathcal{F}$  from another random field, which is a function of the label process  $\omega$ . Basically, the *image process*  $\mathcal{F}$  represents the manifestation of the underlying label process while the prior  $P(\omega)$  represents the simple fact that segmentations should be locally homogeneous. Factoring the above distributions and applying the Bayes theorem gives us the *posterior* distribution  $P(\omega|\mathcal{F}) \propto P(\mathcal{F}|\omega)P(\omega)$ . Note that the constant factor  $1/P(\mathcal{F})$  has been dropped as we are only interested in  $\hat{\omega}$  which *maximizes* the posterior, i.e. the Maximum A Posteriori (MAP) estimate of the hidden field  $\omega$ :

$$\hat{\omega} = \arg \max_{\omega \in \Omega} P(\mathcal{F} | \omega)P(\omega)$$

### A. Unsupervised Segmentation of Color Textured Images

The models of the above distributions depend also on certain parameters. Since neither these parameters nor  $\omega$  is known, both has to be inferred from the only observable entity  $\mathcal{F}$ . This is known in statistics as the *incomplete data* problem.

The proposed segmentation model [1] consists of an MRF defined over a nearest neighborhood system (see Fig. 1) and pixel classes are represented by multivariate Gaussian distributions. This kind of modelization corresponds well to our features: Texture feature images (extracted by Gabor filters) are constructed in such a way that similar textures map to similar intensities. Hence pixels with a given texture will be assigned a well determined value with some variance. Furthermore, pixels with similar color map to their average color. Putting these feature distributions into one multivariate Normal mixture, the modes will correspond to clusters of pixels which are homogeneous in both color and texture properties. Therefore regions will be formed where both features are homogeneous while boundaries will be present where there is a discontinuity in either color or texture. Applying these ideas, the *image process*  $\mathcal{F}$  can be formalized as follows:  $P(\mathcal{F}_s | \omega_s)$  follows a Normal distribution  $N(\vec{\mu}, \Sigma)$ , each pixel class  $\lambda \in \Lambda = \{1, 2, \dots, L\}$  is represented

by its mean vector  $\vec{\mu}_\lambda$  and covariance matrix  $\Sigma_\lambda$ . The whole posterior can now be expressed as a first order MRF by including the contribution of the likelihood term via the singletons (i.e. pixel sites  $s \in \mathcal{S}$ ). Indeed, the singleton energies directly reflect the probabilistic modeling of labels without context, while doubleton clique potentials express relationship between neighboring pixel labels. Thus the energy function of the so defined MRF image segmentation model has the following form:

$$\sum_{s \in \mathcal{S}} \left( \ln(\sqrt{(2\pi)^n |\Sigma_{\omega_s}|}) + \frac{1}{2}(\vec{f}_s - \vec{\mu}_{\omega_s})\Sigma_{\omega_s}^{-1}(\vec{f}_s - \vec{\mu}_{\omega_s})^T \right) + \beta \sum_{\{s,r\} \in \mathcal{C}} \delta(\omega_s, \omega_r) \quad (1)$$

where  $\beta > 0$  is a weighting parameter controlling the importance of the prior. As  $\beta$  increases, the resulting regions become more homogeneous.

The proposed segmentation model has the following parameters:

- 1) The weight  $\beta$  of the prior term,
- 2) the number of pixel classes  $L$ ,
- 3) the mean vector  $\vec{\mu}_\lambda$  and covariance matrix  $\Sigma_\lambda$  of each class  $\lambda \in \Lambda$ .

The automatic determination of  $L$  will be addressed in Section I-B. While  $L$  strongly depends on the input image data,  $\beta$  is largely independent of it. Experimental evidence suggests that the model is not sensitive to a particular setting of  $\beta$  [1]. We found that setting  $\beta \geq 2.0$  gives satisfactory and stable segmentations. Unlike the first two parameters, the mean and covariance of the Gaussians must be computed directly from the input image. Our solution to this problem [1] adopts a general iterative algorithm, known as the *EM algorithm*, to compute the maximum likelihood estimates of the parameters of a mixture density. Basically, we will fit a Gaussian mixture of  $L$  components to the histogram of the image features. The observations consist of the histogram data  $\vec{d}_i (i = 1, \dots, D)$  of the feature images.  $D$  denotes the number of histogram points and the dimension of a data point equals to the dimension of the combined color-texture feature space. Assuming there are  $L$  classes, we want to estimate the mean values  $\vec{\mu}_\lambda$  and covariance matrices  $\Sigma_\lambda$  for each pixel class  $\lambda \in \Lambda$ .

The *EM algorithm* aims at finding parameter values which maximize the normalized log-likelihood function:

$$\mathcal{L} = \frac{1}{D} \sum_{i=1}^D \log \left( \sum_{\lambda \in \Lambda} P(\lambda | \vec{d}_i) \right) \quad (2)$$

The underlying model is that the *complete data* includes not only the observable  $\vec{d}_i$  but also the *hidden data* labels  $\vec{\ell}_i$  specifying which Gaussian process generated the data  $\vec{d}_i$ . Actually,  $\vec{\ell}_i$  is also a vector of dimension  $L$  and  $\vec{\ell}_i^\lambda = 1$  if  $\vec{d}_i$  belongs to class  $\lambda$  and 0 otherwise. The idea is that if labels were known, the estimation of model parameters would be equivalent to the supervised case. Hence the following algorithm is alternating two steps: The estimation of a tentative labeling of the data followed by updating the parameter values based on the tentatively labeled data.

*Algorithm 1 (EM for Gaussian mixture identification):*

- ① **[Estimation]** Replace  $\vec{\ell}_i$  with its conditional expectation based on the current parameter estimates. Since the labels may only take values 0 or 1, the expectation is basically equivalent to the posterior probability:

$$P(\lambda | \vec{d}_i) = \frac{P(\vec{d}_i | \lambda)P(\lambda)}{\sum_{\lambda \in \Lambda} P(\vec{d}_i | \lambda)P(\lambda)}, \quad (3)$$

where  $P(\lambda)$  denotes the component weight.

- ② **[Maximization]** Then, using the current expectation of the labels  $\vec{\ell}_i$  as the current labeling of the data, the estimation of the

parameters is simple:

$$P(\lambda) = \frac{K_\lambda}{D} \quad (4)$$

$$\vec{\mu}_\lambda = \frac{1}{K_\lambda} \sum_{i=1}^D P(\lambda | \vec{d}_i) \vec{d}_i \quad (5)$$

$$\Sigma_\lambda = \frac{1}{K_\lambda} \sum_{i=1}^D P(\lambda | \vec{d}_i) (\vec{d}_i - \vec{\mu}_\lambda)^T (\vec{d}_i - \vec{\mu}_\lambda) \quad (6)$$

where  $K_\lambda = \sum_{i=1}^D P(\lambda | \vec{d}_i)$ . Basically the posteriors  $P(\lambda | \vec{d}_i)$  are used as a weight of the data vectors. They express the contribution of a particular data point  $\vec{d}_i$  to the class  $\lambda$ .

③ Go to Step ① until convergence. Each iteration is guaranteed to increase the likelihood of the estimates. The algorithm is stopped when the change of the log-likelihood  $\mathcal{L}$  is less than a predetermined threshold (our test cases used  $10^{-7}$ ).

The proposed algorithm has been tested on a variety of color images. We compared segmentation results using color-only, texture-only and combined (color+texture) features [1] and found in all test-cases that segmentation based purely on texture gives fuzzy boundaries but usually homogeneous regions, whereas segmentation based on color is more sensitive to local variations but provides sharp boundaries. As for the combined features, the advantages of both color and texture based segmentation have been preserved: we obtained sharp boundaries and homogeneous regions. Results has also been compared to those obtained by the JSEG algorithm [7], a recent unsupervised method for color textured image segmentation. Our method clearly outperforms JSEG (see Fig. 2) but JSEG's advantage is that we do not have to specify the image dependent parameter  $L$ .

## B. Segmentation of Color Images via Reversible Jump MCMC Sampling

Our problem becomes much harder when the number of labels  $L$  is also unknown. We have addressed this problem in the context of color-based image segmentation [2], [3]. When this parameter is also being estimated, the unsupervised segmentation problem may be treated as a *model selection* problem over a combined model space. From this point of view,  $L$  becomes a *model indicator* and the observation  $\mathcal{F}$  is regarded as a three-variate Normal *mixture* with  $L$  components corresponding to clusters of pixels which are homogeneous in color.

The goal of our analysis is inference about the number  $L$  of Gaussian mixture components (each one corresponds to a label), the component parameters  $\Theta = \{(\vec{\mu}_\lambda, \Sigma_\lambda) | \lambda \in \Lambda\}$ , the component weights  $p_\lambda$  summing to 1, the inter-pixel interaction strength  $\beta$ , and the segmentation  $\omega$ . A broadly used tool to sample from the posterior distribution is the Metropolis-Hastings method. Classical methods, however, can not be used due to the changing dimensionality of the parameter space. To overcome this limitation, a promising approach, called Reversible Jump MCMC (RJ-MCMC), has been adopted [2], [3]. When we have multiple parameter subspaces of different dimensionality, it is necessary to devise different *move types* between the subspaces. These will be combined in a so called *hybrid sampler*. For the color image segmentation model, the following move types are needed [2], [3]:

- 1) sampling the labels  $\omega$  (i.e. re-segment the image);
- 2) sampling Gaussian parameters  $\Theta = \{(\vec{\mu}_\lambda, \Sigma_\lambda)\}$ ;
- 3) sampling the mixture weights  $p_\lambda (\lambda \in \Lambda)$ ;
- 4) sampling the MRF hyperparameter  $\beta$ ;
- 5) sampling the number of classes  $L$  (splitting one mixture component into two, or combining two into one).

The only randomness in scanning these move types is the random choice between splitting and merging in move (5). One iteration of the hybrid sampler, also called a *sweep*, consists of a complete pass over these moves. The first four move types are conventional in the sense that they do not alter the dimension of the parameter space.

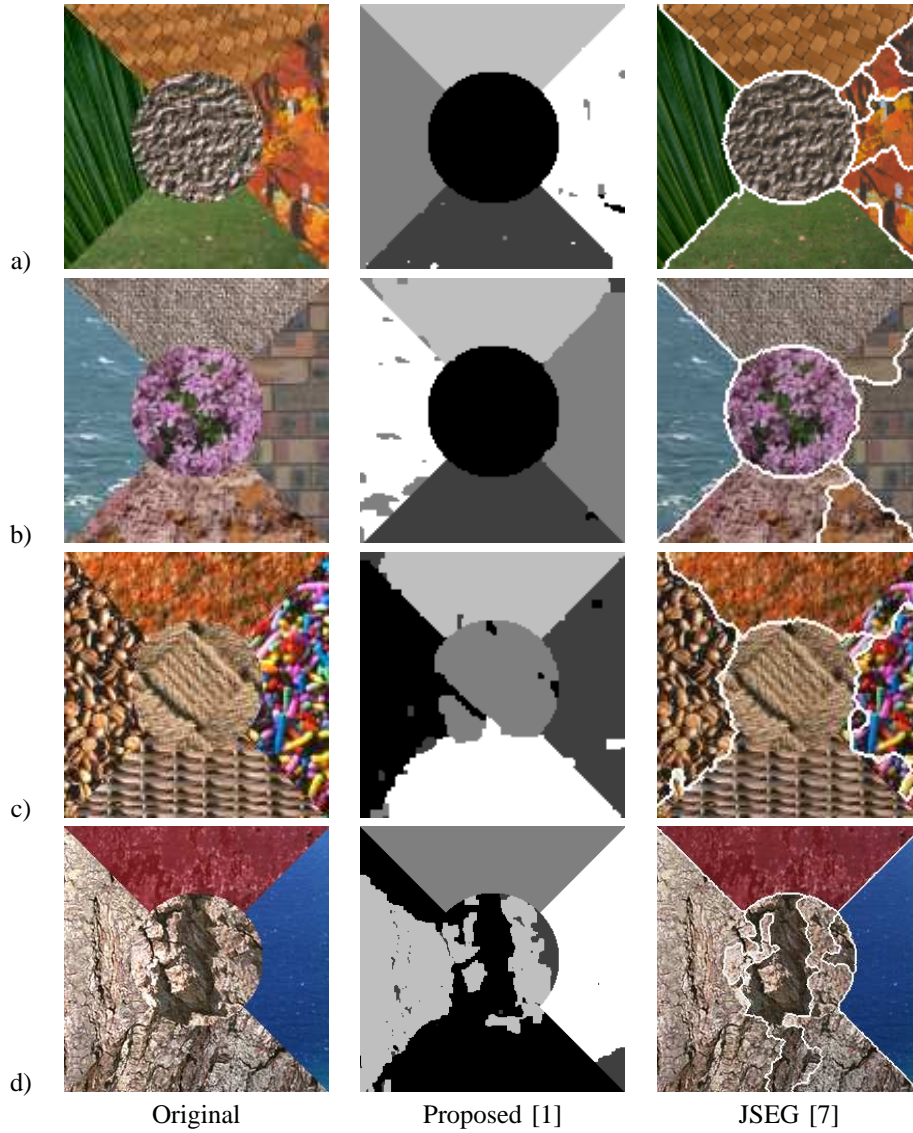


Fig. 2. Segmentation results on synthetic color textured images, each with 5 classes [1].

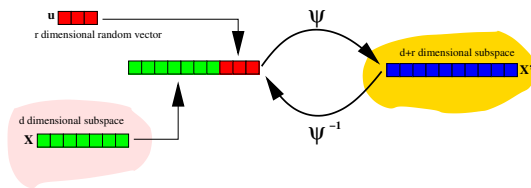


Fig. 3.  $\psi$  is a *diffeomorphism* which transforms back and forth between parameter subspaces of different dimensionality [2], [3]. *Dimension matching* can be implemented by generating a random vector  $u$  such that the dimensions of  $(X, u)$  and  $X'$  are equal.

Hereafter, we will focus on move (5), which requires the use of the reversible jump mechanism. This move type involves changing  $L$  by 1 and making necessary corresponding changes to  $\omega$ ,  $\Theta$  and  $p$ .

The *split proposal* begins by randomly choosing a class  $\lambda$  with a uniform probability  $P_{select}^{split}(\lambda) = 1/L$ . Then  $L$  is increased by 1 and  $\lambda$  is split into  $\lambda_1$  and  $\lambda_2$ . In doing so, a new set of parameters need to be generated. Altering  $L$  changes the dimensionality of the variables  $\Theta$  and  $p$ . Thus we shall define a deterministic function  $\psi$

as a function of these Gaussian mixture parameters:

$$(\Theta^+, p^+) = \psi(\Theta, p, u) \quad (7)$$

where the superscript  $+$  denotes parameter vectors after incrementing  $L$ .  $u$  is a set of random variables having as many elements as the degree of freedom of joint variation of the current parameters  $(\Theta, p)$  and the proposal  $(\Theta^+, p^+)$ . Note that this definition satisfies the *dimension matching* constraint (see Fig. 3), which guarantees that one can jump back and forth between different parameter sub-spaces [2], [3]. This is needed to ensure the convergence of simulated annealing towards a global optimum. The new parameters of  $\lambda_1$  and  $\lambda_2$  are assigned by matching the  $0^{th}$ ,  $1^{th}$ ,  $2^{th}$  moments of the component being split to those of a combination of the two new components [2], [3]:

$$p_\lambda = p_{\lambda_1}^+ + p_{\lambda_2}^+ \quad (8)$$

$$p_\lambda \bar{\mu}_\lambda = p_{\lambda_1}^+ \bar{\mu}_{\lambda_1}^+ + p_{\lambda_2}^+ \bar{\mu}_{\lambda_2}^+ \quad (9)$$

$$p_\lambda (\bar{\mu}_\lambda \bar{\mu}_\lambda^T + \Sigma_\lambda) = p_{\lambda_1}^+ (\bar{\mu}_{\lambda_1}^+ \bar{\mu}_{\lambda_1}^{+T} + \Sigma_{\lambda_1}^+) + p_{\lambda_2}^+ (\bar{\mu}_{\lambda_2}^+ \bar{\mu}_{\lambda_2}^{+T} + \Sigma_{\lambda_2}^+) \quad (10)$$

There are 10 degrees of freedom in splitting  $\lambda$  since covariance matrices are symmetric. Therefore, we need to generate a random

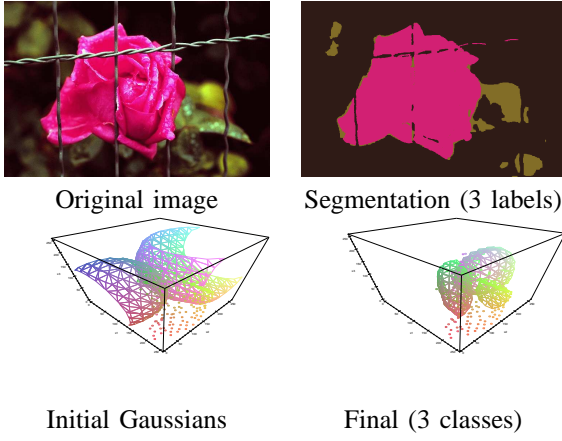
Fig. 4. Segmentation of image *rose41* [2], [3].

TABLE I

F-MEASURE AND CPU TIME COMPARISON [3]

Method	F-measure	CPU time
Human segmentation	0.79	—
RJMCMC	0.57	15 min
JSEG	0.56	2 min

variable  $u1$ , a random vector  $\vec{u}2$  and a symmetric random matrix  $\mathbf{u}3$ . We can now define the diffeomorphism  $\psi$  which transforms the old parameters  $(\Theta, p)$  to the new  $(\Theta^+, p^+)$  using the above moment equations and the random numbers  $u1$ ,  $\mathbf{u}2$ , and  $\mathbf{u}3$  [2], [3]:

$$p_{\lambda_1}^+ = p_{\lambda} u1 \quad (11)$$

$$p_{\lambda_2}^+ = p_{\lambda} (1 - u1) \quad (12)$$

$$\mu_{\lambda_1, i}^+ = \mu_{\lambda, i} + u2_i \sqrt{\Sigma_{\lambda, i, i} \frac{1 - u1}{u1}} \quad (13)$$

$$\mu_{\lambda_2, i}^+ = \mu_{\lambda, i} - u2_i \sqrt{\Sigma_{\lambda, i, i} \frac{u1}{1 - u1}} \quad (14)$$

$$\Sigma_{\lambda_1, i, j}^+ = \begin{cases} u3_{i, i} (1 - u2_i^2) \Sigma_{\lambda, i, i} \frac{1}{u1} & \text{if } i = j \\ u3_{i, j} \Sigma_{\lambda, i, j} \sqrt{(1 - u2_i^2)} & \text{if } i \neq j \\ \times \sqrt{(1 - u2_j^2)} u3_{i, i} u3_{j, j} & \end{cases} \quad (15)$$

$$\Sigma_{\lambda_2, i, j}^+ = \begin{cases} (1 - u3_{i, i}) (1 - u2_i^2) & \text{if } i = j \\ \times \Sigma_{\lambda, i, i} \frac{1}{u1} & \text{if } i = j \\ (1 - u3_{i, j}) \Sigma_{\lambda, i, j} & \text{if } i = j \\ \times \sqrt{(1 - u2_i^2) (1 - u2_j^2)} & \text{if } i \neq j \\ \times \sqrt{(1 - u3_{i, i}) (1 - u3_{j, j})} & \text{if } i \neq j \end{cases} \quad (16)$$

The random variables  $u$  are chosen from the interval  $(0, 1]$ . In order to favor splitting a class into roughly equal portions, beta(1.1, 1.1) distributions are used. The next step is the reallocation of those sites  $s$  where  $\omega_s = \lambda$ . This reallocation is based on the new parameters and has to be completed in such a way as to ensure the resulting labeling  $\omega^+$  is drawn from the posterior distribution with  $\Theta = \Theta^+$ ,  $p = p^+$  and  $L = L + 1$ .

Merging of a pair  $(\lambda_1, \lambda_2)$  is basically the inverse of the split operation [2], [3].

Finally, the split or merge proposal is accepted with a probability relative to the probability ratio of the current and the proposed states. The segmentation and parameter estimation is then obtained as a MAP estimation implemented via simulated annealing:

*Algorithm 2 (RJMCMC Segmentation):*

- ① Set  $k = 0$ . Initialize  $\hat{\beta}^0, \hat{L}^0, \hat{p}^0, \hat{\Theta}^0$ , and the initial temperature  $T_0$ .
- ② A sample  $(\hat{\omega}^k, \hat{L}^k, \hat{p}^k, \hat{\beta}^k, \hat{\Theta}^k)$  is drawn from the posterior distribution using the *hybrid sampler* outlined earlier. Each sub-chain

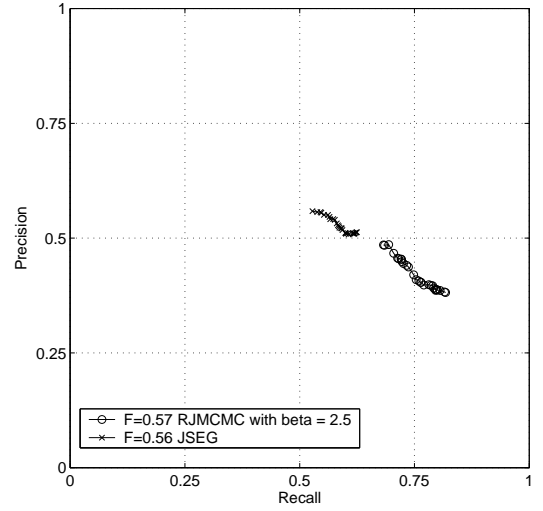


Fig. 6. Precision-recall curve for JSEG and RJMCMC [3].

is sampled via the corresponding move-type while all the other parameter values are set to their current estimate.

- ③ Goto Step ② with  $k = k + 1$  and  $T_{k+1}$  until  $k < \mathcal{K}$ .

As usual, an exponential annealing schedule ( $T_{k+1} = 0.98T_k$ ,  $T_0 = 6.0$ ) was chosen so that the algorithm would converge after a reasonable number of iterations. In our experiments, the algorithm was stopped after 200 iterations ( $T_{200} \approx 0.1$ ).

The proposed algorithm has been tested [2], [3] on a variety of real color images and results have also been compared to those produced by JSEG [7]. In Fig. 5, we show a couple of results obtained on the Berkeley Segmentation Dataset, and in Fig. 6, we plot the corresponding precision-recall curves. Note that RJMCMC has a slightly higher *F-measure* (see Table I) which ranks it over JSEG. However, it is fair to say that both methods perform equally well but behave differently: while JSEG tends to smooth out fine details (hence it has a higher precision but lower recall value), RJMCMC prefers to keep fine details at the price of producing more edges (i.e. its recall values are higher at a lower precision value).

## II. MULTILAYER MRF MODELIZATION

The human visual system is not treating different features sequentially. Instead, multiple cues are perceived simultaneously and then they are integrated by our visual system in order to explain the observations. Therefore different image features have to be handled in a parallel fashion. In this project, we attempt to develop such a model in a Markovian framework based on our earlier work on color-texture segmentation [8]. We propose a new MRF image segmentation model which aims at combining color and motion features for video object segmentation [4], [9]. The model has a multi-layer structure (see Fig. 7): Each feature has its own layer, called *feature layer*, where an MRF model is defined using only the corresponding feature. A special layer is assigned to the combined MRF model. This layer interacts with each feature layer and provides the segmentation based on the combination of different features. Unlike previous methods, our approach doesn't assume motion boundaries being part of spatial ones. The uniqueness of the proposed method is the ability to detect boundaries that are visible only in the motion feature as well as those visible only in the color one.

Perceptually uniform color values and precomputed optical flow data is used as features for the segmentation. The proposed model consists of 3 layers. At each layer, we use a first order neighborhood system and extra inter-layer cliques (Fig. 7). The image features are represented by multivariate Gaussian distributions. For example, on the color layer, the observed image  $\mathcal{F}^c = \{\vec{f}_s^c | s \in \mathcal{S}^c\}$  consists of three spectral component values ( $L^*u^*v^*$ ) at each pixel  $s$  denoted by

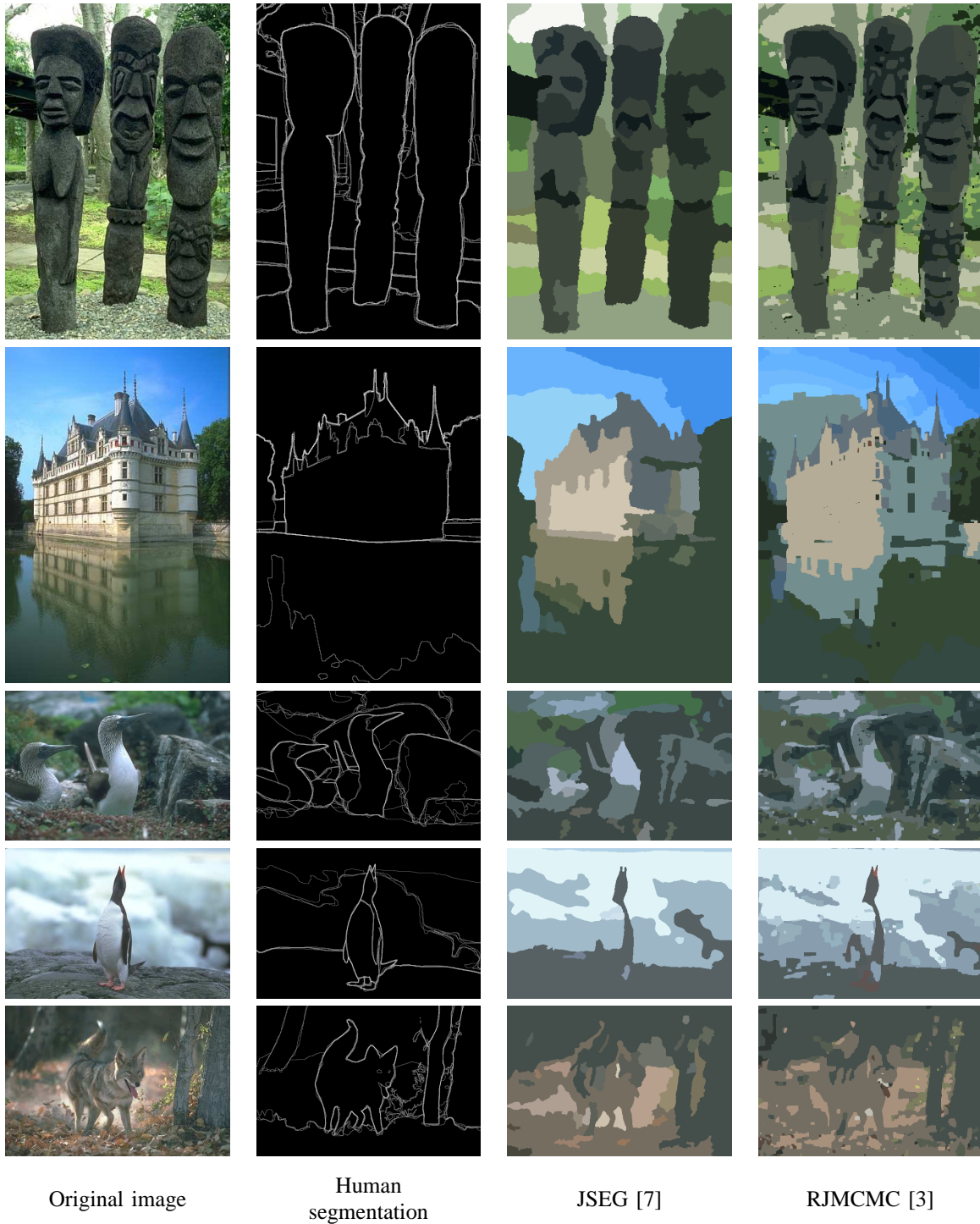


Fig. 5. Benchmark results on images from the Berkeley Segmentation Dataset [3]

the vector  $\vec{f}_s^c$ . The class label assigned to a site  $s$  on the color layer is denoted by  $\psi_s$ . The energy function  $U(\psi, \mathcal{F}^c)$  of the so defined MRF layer has the following form:

$$\sum_{s \in \mathcal{S}^c} \mathcal{G}^c(\vec{f}_s^c, \psi_s) + \beta \sum_{\{s,r\} \in \mathcal{C}} \delta(\psi_s, \psi_r) + \sum_{s \in \mathcal{S}^c} V^c(\psi_s, \eta_s^c)$$

where  $\mathcal{G}^c(\vec{f}_s^c, \psi_s)$  denotes the Gaussian energy term. The last term ( $V^c(\psi_s, \eta_s^c)$ ) is the inter-layer clique potential. The motion layer adopts a similar energy function with some obvious substitutions.

The combined layer only uses the motion and color features indirectly, through inter-layer cliques. A label consists of a pair of

color and motion labels such that  $\eta = \langle \eta^c, \eta^m \rangle$ , where  $\eta^c \in \Lambda^c$  and  $\eta^m \in \Lambda^m$ . The set of labels is denoted by  $\Lambda^x = \Lambda^c \times \Lambda^m$  and the number of classes  $L^x = L^c L^m$ . Obviously, not all of these labels are valid for a given image. Therefore the combined layer model also estimates the number of classes and chose those pairs of motion and color labels which are actually present in a given image. The energy function  $U(\eta)$  is of the following form:

$$\sum_{s \in \mathcal{S}^x} (V_s(\eta_s) + V^c(\psi_s, \eta_s^c) + V^m(\phi_s, \eta_s^m)) + \alpha \sum_{\{s,r\} \in \mathcal{C}} \delta(\eta_s, \eta_r)$$

where  $V_s(\eta_s)$  denotes singleton energies,  $V^c(\psi_s, \eta_s^c)$  (resp.

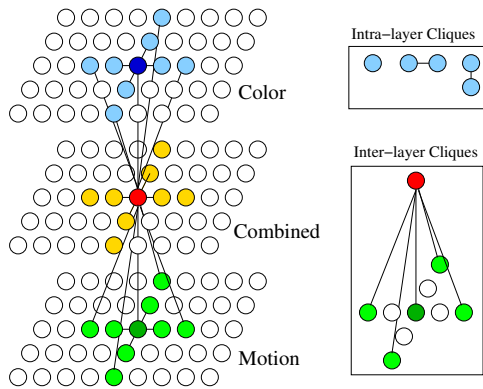


Fig. 7. Multi-layer MRF model [4], [9].

$V^m(\phi_s, \eta_s^m)$  denotes inter-layer clique potentials. The last term corresponds to second order intra-layer cliques which ensures homogeneity of the combined layer.  $\alpha$  has the same role as  $\beta$  in the color layer model and  $\delta(\eta_s, \eta_r) = -1$  if  $\eta_s = \eta_r$ , 0 if  $\eta_s \neq \eta_r$  and 1 if  $\eta_s^c = \eta_r^c$  and  $\eta_s^m \neq \eta_r^m$  or  $\eta_s^c \neq \eta_r^c$  and  $\eta_s^m = \eta_r^m$ . The idea is that region boundaries present at both color and motion layers are preferred over edges that are found only at one of the feature layers. At any site  $s$ , we have 5 inter-layer interactions between two layers: Site  $s$  interacts with the corresponding site on the other layer as well as with the 4 neighboring sites two steps away (see Fig. 7). This potential is based on the difference of the first order potentials at the corresponding feature layers. Clearly, the difference is 0 if and only if both the feature layer and the combined layer has the same label. If the labels are different then it is proportional to the energy difference between the two labels. Finally, the singleton energy controls the number of classes at the combined layer by penalizing small classes.

The proposed algorithm has been tested on real video sequences [4], [9]. We also compare the results to motion only and color only segmentation (basically a monogrid model similar to the one defined for the feature layers but without inter-layer cliques). The mean vectors and covariance matrices were computed over representative regions selected by the user. The number of motion and color classes is known a priori but classes on the combined layer are estimated during the segmentation process. Fig. 8 shows some segmentation results. Note that the head of the men on this image can only be separated from the background using motion features. Clearly, the multi-layer model provides significantly better results compared to color only and motion only segmentations. See Fig. 9 to compare the performance of the proposed method with the one from [10] on the *Mother and Daughter* standard sequence. Note that some of the contours are lost by [10] (the sofa, for example) while our method successfully identifies region boundaries. In particular, our method is able to separate the hand of the mother from the face of the daughter in spite of their similar color. This demonstrates again that the proposed method is quite powerful in combining motion and color features in order to detect boundaries visible only in one of the features.

### III. SHAPE PRIORS FOR SEGMENTATION

The aim of this work is to introduce prior shape knowledge into existing image segmentation models. To accomplish we extended the recently introduced higher-order active contour framework for region and image modeling by introducing a model for a ‘gas of circles’, the ensemble of regions in the image domain consisting of an unknown number of circles, with approximately fixed radius and short range of interactions. We applied the developed models of current interest in remote sensing image processing: the extraction of tree crowns. Forestry services are interested in various quantities associated with forests and plantations, such as the density of trees, the mean crown area and diameter, etc.

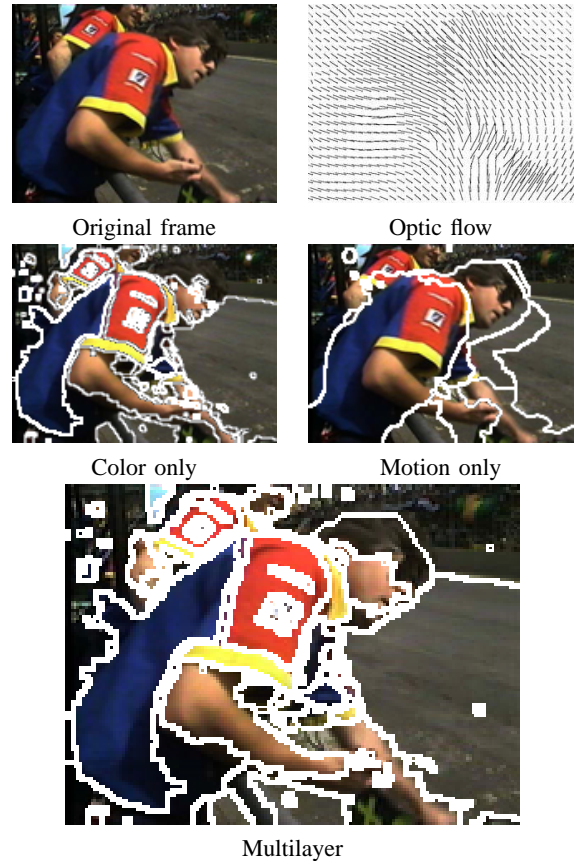


Fig. 8. Segmentation results [4], [9].

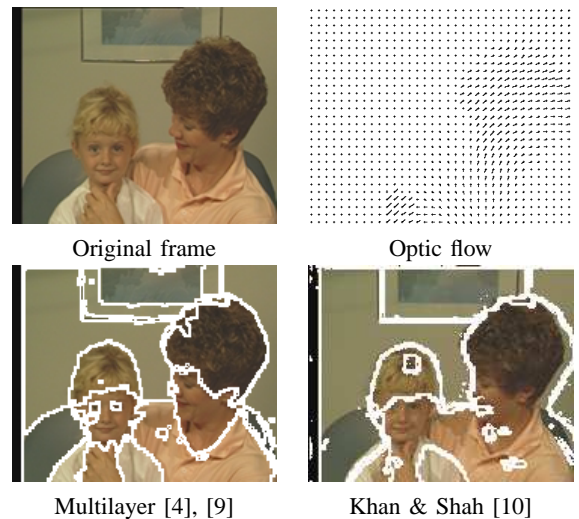


Fig. 9. Comparison of the segmentation results obtained by the proposed method [4], [9] and those produced by the algorithm of Khan &amp; Shah [10].

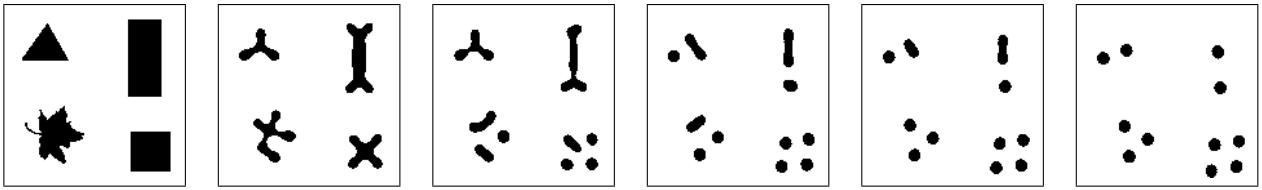


Fig. 10. Sequences of curve evolution using  $E_g$  itself, from left to right: from the initialization to stable state.

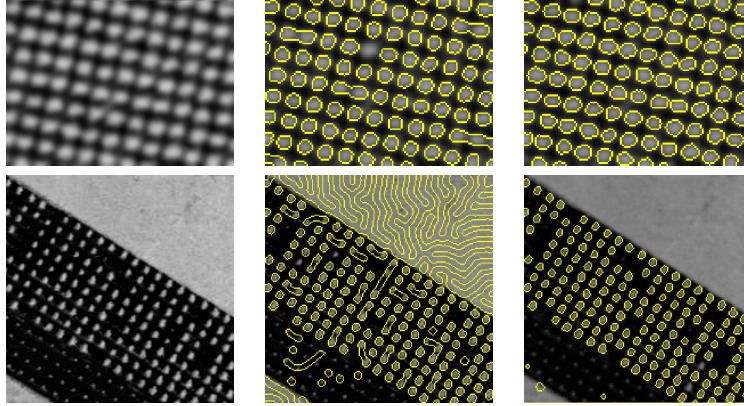


Fig. 11. Results on real aerial images, first column: original, second: results with [5], [11] model, last column: results using [6]. IFN ©

To include more complex prior knowledge, longer-range interactions are needed. There is a large body of work that does this implicitly, via a template region or regions to which the segmented region  $R$  is compared. However, such energies effectively limit  $R$  to a bounded subset of region space close to the template(s), which excludes, *inter alia*, cases like tree crown extraction in which  $R$  has an unknown number of connected components. ‘Higher-order active contours’ (HOACs) provide a complementary approach. HOACs generalize classical active contours to include multiple integrals over the contour. Thus HOAC energies explicitly model long-range interactions between boundary points without using a template. This allows the inclusion of complex prior knowledge while permitting the region to have an arbitrary number of connected components, which furthermore may interact amongst themselves. The approach is very general: classical energies are linear functionals on the space of regions; HOACs include all polynomial functionals.

In [5], [11], a HOAC energy was used for tree crown extraction. In this ‘gas of circles’ model, collections of mutually repelling circles of given radius  $r_0$  are local minima of the geometric energy. The model has many potential applications in varied domains, but it suffers from a drawback: such local minima can trap the gradient descent algorithm used to minimize the energy, thus producing phantom circles even with no supporting data. The model as such is not at fault: an algorithm capable of finding the global minimum would not produce phantom circles. This suggests two approaches to tackling the difficulty. One is to find a better algorithm. The other is to compromise with the existing algorithm by changing the model to avoid the creation of local minima, while keeping intact the prior knowledge contained in the model. We solved the problem of phantom circles in [5], [11]’s model by calculating, via a Taylor expansion of the energy, parameter values that make the circles into inflection points rather than minima. In addition, we find that this constraint halves the number of model parameters, and severely constrains one of the two that remain, while improving the empirical success of the model [6].

#### A. The ‘gas of circles’ model

HOAC energies generalize classical active contour energies by including multiple integrals over the contour. The simplest such

generalizations are quadratic energies, which contains double integrals. There are several forms that such multiple integrals can take, depending on whether or not they take into account contour direction at the interacting points. The Euclidean invariant version of one of these forms is

$$E_g(\gamma) = \lambda L(\gamma) + \alpha A(\gamma) - \frac{\beta}{2} \int \int \tau(p) \cdot \tau(p') \Psi(|p, p'|) dp dp',$$

where  $\gamma$  is the contour, parameterized by  $p$ ;  $L$  is the length of the contour;  $A$  is the area;  $|p, p'| = |\gamma(p) - \gamma(p')|$ ;  $\tau = \dot{\gamma}$  is the (unnormalized) tangent vector to the contour; and  $\Psi$  is an interaction function that determines the geometric content of the model. With an appropriate choice of interaction function  $\Psi$ , the quadratic term creates a repulsion between antiparallel tangent vectors. This has two effects. First, for particular ranges of  $\alpha$ ,  $\beta$ , and  $d_{min}$  ( $\lambda = 1$  wlog), circular structures, with a radius  $r_0$  dependent on the parameter values, are stable to perturbations of their boundary. Second, such circles repel one another if they approach closer than  $2d_{min}$ . Regions consisting of collections of circles of radius  $r_0$  separated by distances greater than  $2d_{min}$  are thus local energy minima. We [5], [11] called this the ‘gas of circles’ model.

Via a stability analysis, we [5], [11] found the ranges of parameter values rendering circles of a given radius stable as functions of the desired radius. Stability, however, created its own problems, as circles sometimes formed in places where there was no supportive data. To overcome this problem, in [6], the criterion that circles of a given radius be local energy minima was replaced by the criterion that they be points of inflexion. As well as curing the problem of ‘phantom’ circles, this revised criterion allowed the fixing of the parameters  $\alpha$ ,  $\beta$ , and  $d_{min}$  as functions of the desired circle radius, leaving only the overall strength of the prior term,  $\lambda$ , unknown. For energy-based models, parameter adjustment is a problem, so this is a welcome advance.

To illustrate the behavior of the prior model, figure 10 shows the result of gradient descent starting from the region on the left. Note that there is no data term. The parameter values in these experiments render the circles involved stable. With the parameter values calculated in [6], they would disappear. Figure 11 illustrates results using the published models.

#### IV. DISSEMINATION AND FUTURE WORK

Our results have been published in

- two top tier peer-reviewed international conference proceedings [2], [5],
- two LNCS book series of Springer [4], [6]
- two peer-reviewed international journals [1], [3],
- two peer-reviewed [9], [12] and four non-refereed national conference proceedings [13]–[16],
- one INRIA Research Report [11].

The project's achievements have also been presented at leading international conferences

- 2004** British Machine Vision Conference.
- 2006** Asian Conference on Computer Vision, International Conference on Pattern Recognition, Indian Conference on Computer Vision, Graphics and Image Processing.

and national conferences:

- 2004,2007** Conference of the Hungarian Association for Image Analysis and Pattern Recognition.
- 2005** Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition.

I also gave the following invited talks about the project's results at leading international research institutes:

- 2005** *Reversible Jump Markov Chain Monte Carlo for Unsupervised MRF Color Image Segmentation*, 25 April 2005, INRIA Sophia Antipolis, France.
- 2006** *Energy Minimization Methods in Image Segmentation*, 24 January 2006, IIT Bombay, India.
- 2007** *Multilayer Markovian Models*, 17 April 2007, INRIA Sophia Antipolis, France.

Although the project officially finished by the end of 2006, there are some ongoing works as well as submitted and planned publications. A journal paper about the results presented in Section III has been submitted to *Pattern Recognition* [17]. Another application of the multilayer MRF model in Section II has been submitted to *IEEE International Conference on Image Processing* [18].

There is an ongoing bilateral (Hungarian-French) PhD work by Mr. Peter Horvath which is strongly related to Section III. French co-supervisors are Ian Jermyn and Josiane Zerubia from the Ariana Group of INRIA Sophia Antipolis, France. Defense expected in 2007.

A software licence agreement is currently being signed by the *Hungarian Forest Service, University of Szeged, and INRIA Sophia Antipolis, France*. This will allow the *Hungarian Forest Service* to use our program outlined in Section III in exchange for aerial images. The importance of this contract is two-fold: First, these images are needed for further research. Second, the use of our program in a real environment will help to improve it and potentially find other industrial applications.

#### REFERENCES

- [1] Z. Kato and T. C. Pong, "A Markov random field image segmentation model for color textured images," *Image and Vision Computing*, vol. 24, no. 10, pp. 1103–1114, Oct. 2006.
- [2] Z. Kato, "Reversible jump Markov chain Monte Carlo for unsupervised MRF color image segmentation," in *Proceedings of British Machine Vision Conference*, A. Hoppe, S. Barman, and T. Ellis, Eds., vol. 1. Kingston, UK: BMVA, Sept. 2004, pp. 37–46.
- [3] —, "Segmentation of color images via reversible jump MCMC sampling," *Image and Vision Computing*, 2007, in press.
- [4] Z. Kato and T. C. Pong, "A multi-layer MRF model for video object segmentation," in *Proceedings of Asian Conference on Computer Vision*, ser. Lecture Notes in Computer Science, P. J. Narayanan, S. K. Nayar, and H.-Y. Shum, Eds., vol. 3852. Hyderabad, India: Springer, Jan. 2006, pp. 953–962.
- [5] P. Horvath, I. Jermyn, Z. Kato, and J. Zerubia, "A higher-order active contour model for tree detection," in *Proceedings of International Conference on Pattern Recognition*, vol. 2, IAPR. Hong Kong, China: IEEE, Aug. 2006, pp. 130–133.
- [6] —, "An improved 'gas of circles' higher-order active contour model and its application to tree crown extraction," in *Proceedings of Indian Conference on Computer Vision, Graphics and Image Processing*, ser. Lecture Notes in Computer Science, P. Kalra and S. Peleg, Eds., vol. 4338. Madurai, India: Springer, Dec. 2006, pp. 152–161.
- [7] Y. Deng, , and B. S. Manjunath, "Unsupervised segmentation of color-texture regions in images and video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 8, pp. 800–810, Aug. 2001. [Online]. Available: <http://vision.ece.ucsb.edu/segmentation/jseg/>
- [8] Z. Kato, T. C. Pong, and G. Q. Song, "Unsupervised segmentation of color textured images using a multi-layer MRF model," in *Proceedings of International Conference on Image Processing*, vol. I. Barcelona, Spain: IEEE, Sept. 2003, pp. 961–964.
- [9] Z. Kato and T. C. Pong, "Video object segmentation using a multicue Markovian model," in *Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition*, D. Chetverikov, L. Czuni, and M. Vincze, Eds., KEPAF, OAGM/AAPR. Veszprem, Hungary: Austrian Computer Society, May 2005, pp. 111–118.
- [10] S. Khan and M. Shah, "Object based segmentation of video using color, motion and spatial information," in *Proceedings of International Conference on Computer Vision and Pattern Recognition*, vol. II. Kauai, Hawaii: IEEE, Dec. 2001, pp. 746–751.
- [11] P. Horvath, I. Jermyn, Z. Kato, and J. Zerubia, "A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction," INRIA, Sophia Antipolis, France, Research Report 6026, Nov. 2006. [Online]. Available: <http://hal.inria.fr/inria-00115631>
- [12] P. Horvath, A. Bhattacharya, I. Jermyn, J. Zerubia, and Z. Kato, "Shape moments for region based active contours," in *Joint Hungarian-Austrian Conference on Image Processing and Pattern Recognition*, D. Chetverikov, L. Czuni, and M. Vincze, Eds., KEPAF, OAGM/AAPR. Veszprem, Hungary: Austrian Computer Society, May 2005, pp. 187–194.
- [13] P. Horvath and Z. Kato, "Optical flow computation using an energy minimization approach," in *Conference of Hungarian Association for Image Analysis and Pattern Recognition*, Z. Gacsi, P. Barkoczy, and G. Sarkozi, Eds., Miskolc-Tapolca, Hungary, Jan. 2004, pp. 125–130, non-refereed.
- [14] —, "Color, texture and motion segmentation using gradient vector flow," in *Conference of Hungarian Association for Image Analysis and Pattern Recognition*, Z. Gacsi, P. Barkoczy, and G. Sarkozi, Eds., Miskolc-Tapolca, Hungary, Jan. 2004, pp. 131–137, non-refereed.
- [15] Z. Kato, T. C. Pong, and G. Q. Song, "Color textured image segmentation using a multi-layer Markovian model," in *Conference of Hungarian Association for Image Analysis and Pattern Recognition*, Z. Gacsi, P. Barkoczy, and G. Sarkozi, Eds., Miskolc-Tapolca, Hungary, Jan. 2004, pp. 152–158, non-refereed.
- [16] P. Horvath, I. Jermyn, Z. Kato, and J. Zerubia, "Kör alakú objektumok szegmentálása magasabb rendű aktív kontúr modellek segítségével," in *Conference of Hungarian Association for Image Analysis and Pattern Recognition*, A. Fazekas and A. Hajdu, Eds., Debrecen, Hungary, Jan. 2007, pp. 133–140, non-refereed, in Hungarian.
- [17] —, "A higher-order active contour model of a 'gas of circles' and its application to tree crown extraction," *Pattern Recognition*, 2007, submitted.
- [18] C. Benedek, T. Sziranyi, Z. Kato, and J. Zerubia, "A multi-layer MRF model for object-motion detection in unregistered airborne image-pairs," in *Proceedings of International Conference on Image Processing*, 2007, submitted.