14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

IDENTIFICATION OF IMAGE STRUCTURE BY THE MEAN SHIFT PROCEDURE FOR HIERARCHICAL MRF-BASED IMAGE SEGMENTATION

R.Gaetano, G.Poggi, G.Scarpa

Dipartimento di Ingegneria Elettronica e delle Telecomunicazioni
Università Federico II di Napoli, via Claudio, 21 – 80125 Napoli, Italy
{gaetano,poggi,giscarpa}@unina.it

ABSTRACT

Tree-structured Markov random fields have been recently proposed in order to model complex images and to allow for their fast and accurate segmentation. By modeling the image as a tree of regions and subregions, the original K-ary segmentation problem can be recast as a sequence of reduced-dimensionality steps, thus reducing computational complexity and allowing for higher spatial adaptivity. Up to now, only binary tree structures have been considered, which simplifies matters but also introduces an unnecessary constraint. Here we use a more flexible structure, where each node of the tree is allowed to have a different number of children, and also propose a simple technique to estimate such a structure based on the mean shift procedure. Experiments on synthetic images prove the structure estimation procedure to be quite effective, and the ensuing segmentation to be more accurate than in the binary case.

1. INTRODUCTION

Segmentation is a low level task that aims at partitioning an image into different homogeneous regions, according to various properties like color, texture, shape, motion, *etc.* It is useful for many high-level applications in such diverse fields as remote-sensing, medical imaging, video coding, image restoration, and so on. The task is often complicated (*e.g.,* in remote sensing applications) by the presence of strong noise components combined to the signal, which calls for the use of sophisticated statistical techniques. Bayesian segmentation based on Markov random field (MRF) image modeling [1, 2] has emerged in recent years as one of the most promising approaches to this problem, guaranteeing a good accuracy at the cost of a significant complexity.

To reduce such a complexity, the class of Tree-Structured MRF (TS-MRF) models has been recently proposed [3, 4]. The main advantage derived from the use of such models is given by their recursive definition, which leads to a recursive formulation and solution of the segmentation problem. The original K-class segmentation is decomposed into a sequence of reduced-dimensionality steps, whose aim is to subdivide a large region in two or more component regions, leading to a hierarchical tree-structured representation of the image. Each step turns out to be relatively simple and manageable, and the overall complexity is significantly reduced with respect to a single K-dimensional step.

In previous works, the proposed techniques used only fixed-dimensionality tree structures (typically binary trees), in order to avoid difficult modeling problems. Such a rigid constraint, however, can sometimes lead to the inappropriate modeling of arbitrarily structured data, as already recognized in [4], where a split-and-merge procedure was proposed to make up for such problems. For this reason, we consider here a more general model, where each node of the tree is allowed to have an arbitrary number of children, that is, each region can be split in two or more subregions. Of course, with such a general model, the problem arises of estimating the tree-structure that better fit the inherent structure of the data (we are considering a strictly unsupervised problem). Therefore, we propose a variation to the unsupervised segmentation algorithm of [4] by providing the dynamic dimensionality selection for MRFs located at each node of the tree. This is obtained by means of the Mean-Shift analysis [5] applied to each region in the spectral domain, which allow us to discover the most relevant modes of the underlying probability distribution and estimate the number of local sub-regions.

The next Section will provide the necessary background on TS-MRF segmentation and the mean-shift procedure, Section 3 will outline the proposed segmentation algorithm, and Section 4 will report some experimental results on synthetic images and draw conclusions.

2. BACKGROUND

2.1 Tree-Structured MRF Segmentation

Image segmentation can be easily formulated as a MAP estimation problem. Suppose each pixel of the image $\mathcal{S}$ belongs to one of K different classes, and let $x_s \in \{1, \ldots, K\}$ indicate the class of pixel s. Then $x = \{x_s, s \in \mathcal{S}\}$ is the segmentation of the image $\mathcal{S}$ in K classes. Of course, x is unknown, and must be recovered from the data $y = \{y_s, s \in \mathcal{S}\}$. Modeling all quantities as random variable/fields (capital letters), we accept as our segmentation $\widehat{x}$ the most likely realization of X given the observed data, namely,

$$\widehat{x} = \arg\max_x p(x|y) = \arg\max_x p(y|x)p(x) \qquad (1)$$

It can be convenient to model the classes as a Markov random field[1], that is a reasonably simple and general model which keeps into account the spatial dependencies in the image through the conditional probability that

[1] We will not go into detail on MRF's, assuming the reader is already familiar with this topic, and referring to the literature (e.g., [2]).

a pixel belong to a given class given the classes of its neighbors. As a result, X has Gibbs distribution

$$p(x) = \frac{1}{Z} \exp[\sum_{c \in \mathcal{C}} V_c(x, \theta)] \tag{2}$$

where Z is a normalizing constant, and the $V_c(\cdot, \theta)$'s are potential functions, defined on suitable cliques c of the image, and depending on some hyperparameters θ.

Observed data Y, instead, are usually modeled as multivariate Gaussian, spatially independent given the class, and characterized for each class k by their mean μ_k and covariance matrix Σ_k.

Given this model, the segmentation problem amounts to maximizing the function $p(y|x)p(x)$ over x, where all the quantities K, μ_k, Σ_k and θ are in general unknown and must be estimated themselves from the data. Due to the inherent complexity of this problem, in practical applications one must resort to heuristics that reduce the search complexity, and accept suboptimal solutions.

To drastically reduce the search complexity, in [3, 4] we have introduced a tree-structured MRF model, where the full segmentation is obtained through a sequence of binary segmentations. More precisely, the whole image is associated to the root node $t = 1$ of a tree T, and is segmented in two regions using a binary MRF model. The two new regions, associated with the children of the root, $t = 2$ and $t = 3$, can be likewise segmented by means of newly defined local binary MRF, and the growth of the tree continues until a suitable stopping condition is met. Therefore, each node t of the tree is associated with a region of the image $\mathcal{S}^t$, a field of observed data Y^t with realization y^t, a binary MRF X^t with realization x^t, and a set of parameters $\{\mu^t, \Sigma^t, \theta^t\}$. The leaves of the tree partition the image in K disjoint regions, namely provide the desired segmentation.

In [3] it is shown that the growth of the tree can be based exclusively on local decisions, taken on the basis of the node split gain defined as the likelihood ratio between the two hypotheses of splitting the region in two or leaving it unaltered.

The use of *binary* fields only, together with the locality of the splitting (the segmentation of a region does not depend on other regions) leads to a significant reduction of the computational complexity with respect to the case where a flat K-class MRF is used. However, it also introduces some additional constraint which tend to impair the segmentation performance, which is why we turn here to more general models, in which the dimensionality of each node's MRF must be estimated in advance.

2.2 Mean-Shift Procedure

The Mean Shift procedure [5] is typically used to detect the local maxima (*modes*) of a probability density function in a given feature space, and applies to many vision tasks, such as discontinuity preserving smoothing and image segmentation. It is based on the *Parzen Window Kernel Density Estimator*: such a technique states that, given a d-dimensional feature space and a set of n data points $(s_1, \cdots, s_n)$, the p.d.f. $p(s)$ can be estimated as:

$$\hat{p}_{h,K}(s) = \frac{c_{k,d}}{nh^d} \sum_{i=1}^{n} k\left(\left\| \frac{s - s_i}{h} \right\|^2 \right), \tag{3}$$

where $k(s)$, called *kernel profile function*, is related to a *kernel function* $K(s)$ by the relation $K(s) = c_{k,d} k(\|s\|^2)$, $c_{k,d}$ is a normalization constant, and h is the "bandwidth" parameter. Applying the gradient operator to both members of (3) yields:

$$\hat{\nabla} p_{h,K}(s) = \frac{2c_{k,d}}{h^2 c_{g,d}} \hat{p}_{h,G}(s) \mathbf{m}_{h,G}(s), \tag{4}$$

where

$$\hat{p}_{h,G}(s) = \frac{c_{g,d}}{nh^d} \sum_{i=1}^{n} g\left(\left\| \frac{s - s_i}{h} \right\|^2 \right), \tag{5}$$

$$\mathbf{m}_{h,G}(s) = \frac{\sum_{i=1}^{n} s_i g\left(\left\| \frac{s-s_i}{h} \right\|^2 \right)}{\sum_{i=1}^{n} g\left(\left\| \frac{s-s_i}{h} \right\|^2 \right)} - s, \tag{6}$$

and $g(s) = -k'(s)$ is the profile of a new kernel function $G(s)$.

The function $\mathbf{m}_{h,G}(s)$ is called *mean shift*: it is possible to prove that, for each point s, this vector always points towards the direction of the maximum increase in the density. Such "alignment" allows the mean shift to define a path leading from a starting kernel center s to a stationary point of the p.d.f., that corresponds to a mode of the density. Hence, it is possible to define the *mean shift procedure* as the iterative execution of three steps, once a starting kernel center s is fixed:

1. computation of mean shift vector $\mathbf{m}_{h,G}(s)$,
2. translation of kernel $G(s)$ by $\mathbf{m}_{h,G}(s)$,
3. update of kernel center $s = s + \mathbf{m}_{h,G}(s)$.

This procedure is guaranteed to converge at a nearby local maximum of the density. Clearly, to detect all significant modes, it must be executed multiple times, each time with a different initialization, in order to cover most of the feature space with kernels. A typical choice is to initialize the procedure with a randomly selected subset of the available data points, whose dimension is chosen as a compromise between computational effort and desired precision. Another important issue is the selection of kernel shape: our choice is the use of *multivariate normal kernel*, defined as

$$K_N(s) = (2\pi)^{-d/2} \exp\left(-\frac{1}{2} \|s\|^2 \right). \tag{7}$$

But the most critical issue when applying this mode detection procedure is the selection of the bandwidth parameter h. Such a parameter is related to the resolution of modes: a large value for h typically leads to estimating a smoother density, and hence to detecting less modes; on the other hand, using a small bandwidth, the density estimate will suffer from too much variability, and too many modes can be detected.

3. PROPOSED SEGMENTATION ALGORITHM

We implement a recursive segmentation algorithm. The whole image, associated with the root of the tree, is initially split in an arbitrary number of regions, not necessarily two. Each of these regions is then split in its turn, again in two or more subregions, and this continues recursively until a stopping criterion is satisfied. To each node t, a local Potts MRF with K_t labels is associated, and the local segmentation is obtained by maximizing the posterior probability, by means of any suitable technique.

The main innovation proposed here with respect to the algorithm described in [4] is the use of generic rather than binary splits. This allows one to avoid important segmentation inaccuracies; for example, if a region presents three equally significant classes, and we attempt to subdivide it in just two sub-regions (following a fixed binary approach) the split will very likely generate a false contour.

The key problem becomes the automatic selection of the number of classes, and we estimate it as the number of dominant modes in the p.d.f. of the current region, as provided by the mean shift procedure. Although this approach makes certainly sense, in principle, it is clear that, in the empirical multivariate p.d.f.'s we deal with, the mean shift procedure can find anything from a single mode to a large number of modes, depending on the critical bandwidth parameter h which governs mode resolution. We therefore decided to set $h = \lambda_1^{1/d}$, with λ_1 the largest eigenvalue associated with the data. This way, when there is a large spread in the region's data, indicating two or more clearly distinguishable classes, a large bandwidth is used so as to overlook finer distinction and avoid a proliferation of modes. On the contrary, when the data are more homogeneous, the bandwidth reduces automatically and allows one to distinguish classes that are more similar to one another. As stopping conditions, we accept both the obvious case of a single mode found, and that of a large number of modes because too complex structures are typically unreliable, meaning that the corresponding region is likely non-structured at all. In addition, the tree growth stops when the split gain [4] is smaller than 1.

Another important innovation concerns the split initialization. In MRF-based segmentation, in fact, it is necessary to provide an initial split of each region in order to apply map optimization procedures. In the original algorithm [4] this was obtained by using a simple clustering algorithm, the GLA [6], applied on the whole region, but now the mode estimation procedure provides such pre-classification, for a suitable sample set, as a by-product with no need for further processing and increased reliability. Starting from this pre-classification, we estimate the mean and covariance for the subregions and compute the full initial split using the maximum likelihood technique.

The proposed TS-MRF-based unsupervised algorithm can be summarized as follows:

1. Initialize the tree.
2. For each newly created leaf:
   (a) Compute the largest eigenvalue and hence the bandwidth parameter h_t;
   (b) Execute the mean-shift mode detection procedure on the current region to obtain the dimensionality K_t for the current node, and to initialize the splitting procedure;
   (c) If $K_t = 1$ or $K_t > K_{\max}$ set this node as terminal, otherwise complete the test split on the current region and evaluate the associated split gain.
3. If all leaves are set as terminal, then exit. Else find the non-terminal leaf with the largest split gain. If this value is smaller than 1 exit, otherwise validate the split and go to step 2.

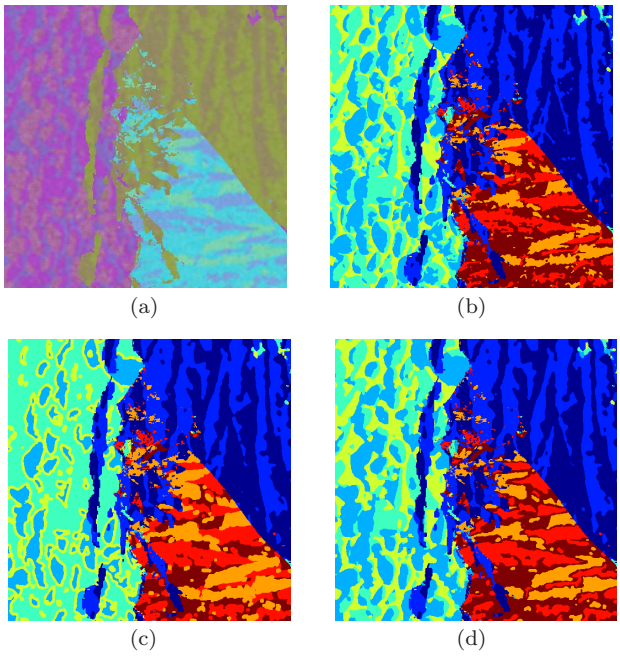

Figure 1: test image (a), ground truth (b), segmentation with TS-MRF/B (c), with TS-MRF/K (d).

4. EXPERIMENTAL RESULTS

Our experiments are carried out, for the time being, on a synthetic test image so as to measure objectively the performance of the new segmentation algorithm as opposed to the reference algorithm based on purely binary splits.

The three-band synthetic image, shown in Fig.1(a), has been obtained by projecting the ground truth of Fig.1(b) on the data space, adding white noise, and finally performing a light spatial filtering. The reference algorithm, referred to as TS-MRF/B from now on, generates the tree structure shown in Fig.2 and the segmentation map of Fig.1(c), while the new algorithm, called TS-MRF/K, generates the tree structure of Fig.3 and the segmentation map of Fig.1(d).

It is clear that TS-MRF/B has a hard time fitting the intrinsic (non-binary) structure of the data. Whenever a ternary split is needed (for example in the root node) the algorithm must simulate it by means of a sequence of two binary splits. Sometimes, this has no detrimental effect, like in the root, where the dark-blue, light-blue and orange macroregions are correctly singled out, but

	DB.	B.	LB.	C.	G.	O.	R.	Br.	Total
D.Blue	**49083**	2749							51832
Blue	2467	**44573**							47040
L.Blue			**20922**		6				20928
Cyan			8	**29944**	20361				50313
Green			14866	11383	**5472**				31721
Orange			6			**13129**	1436	6574	21145
Red						2	**14987**	6863	21852
Brown						5		**17308**	17313
Total	51550	47322	35802	41327	25839	13136	16423	30745	262144

Figure 2: Confusion matrix for TS-MRF/B.

	DB.	B.	LB.	C.	G.	O.	R.	Br.	Total
D.Blue	**48928**	2632							51560
Blue	2622	**44690**							47312
L.Blue			**32512**	9337	77				41926
Cyan			6	**26883**	372				27261
Green			3278	5107	**25390**				33775
Orange			6			**11429**		3032	14461
Red						1	**14844**	3338	18183
Brown						1706	1579	**24375**	27666
Total	51550	47322	35802	41327	25839	13136	16423	30745	262144

Figure 3: Confusion matrix for TS-MRF/K.

in at least one instance, the split of the light-blue regions, this leads to a grossly inaccurate segmentation, as also testified by the confusion matrix[2] reported in Fig.1. From another point of view, this inaccuracy can be seen as the detection of a false contour. A further split of the cyan region using TS-MRF/B succeeds in revealing the correct missing contour, but the overall result will be an obvious oversplitting of the macroregion.

On the contrary, the proposed TS-MRF/K provides the correct (or *a* correct) tree structure for the test image, with a first ternary split at the root node that singles out the correct macroregions, each of which is then split in two or three regions, following their actual composition. As a consequence, the segmentation map is usually accurate, but for some random sparse errors, as obvious from the analysis of the corresponding confusion matrix of Fig.2. Major improvements have been obtained on the light-blue macroregion, due to the direct ternary split, and on the orange one. For the latter case, we observed in particular that the new split initialization method leads to a more accurate contour detection, thus pointing out a limit of the old GLA-based algorithm.

Some interesting results concern also the total number of classes detected by the two algorithms. Using TS-MRF/B the segmentation process does not stop until the maximum number of classes is reached, while with TS-MRF/K, it stops automatically after 9 classes are detected, with only one elementary region oversplit, thus resulting, for the case, in a drastic reduction of oversegmentation phenomena.

5. CONCLUSIONS

In this work we proposed a modification to the unsupervised TS-MRF segmentation algorithm, aimed at extending hierarchical data modeling from only binary trees to any generic tree structure. The mean shift procedure has been used to select the dimensionality of each node of the tree. Some experimental results have been obtained on synthetic data, showing that the new technique can perform better in the presence of arbitrarily structured data and can also address several drawbacks of the existing TS-MRF algorithm.

[2] A *confusion matrix* is a table that allows comparison between a predicted class membership and an actual one. Each element of the matrix in position (i, j) represents the number of pixels belonging to class i actually classified as belonging to class j.

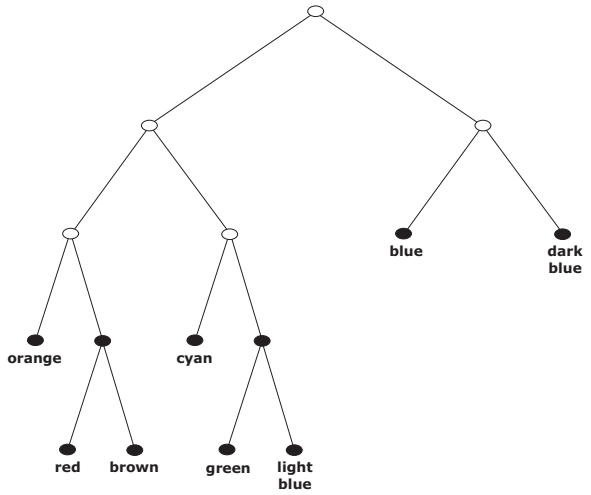

Figure 4: 8-class binary tree structure.

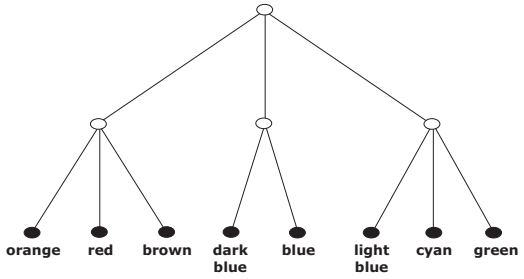

Figure 5: 8-class K-ary tree structure.

REFERENCES

[1] S.Geman, D.Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp.721-741, Nov.1984.

[2] S.Z.Li, *Markov random field modeling in image analysis*, Springer-Verlag, 2001.

[3] G.Poggi, A.R.P.Ragozini, "Image segmentation by tree-structured Markov random field," *IEEE Signal Processing Letters*, pp.155-157, July 1999.

[4] C.D'Elia, G.Poggi, G.Scarpa, "A tree-structured Markov random field model for Bayesian image segmentation", *IEEE Transactions on Image Processing*, pp.1259-1273, Oct. 2003.

[5] D.Comaniciu, P.Meer, "Mean Shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp.603-619, May 2002.

[6] A.Gersho, R.M.Gray, "Vector Quantization and Signal Compression," *Kluwer Academic,* Jan. 1992.