16th European Signal Processing Conference (EUSIPCO 2008), Lausanne, Switzerland, August 25-29, 2008, copyright by EURASIP

GENERAL REGION MERGING BASED ON FIRST ORDER MARKOV INFORMATION THEORY STATISTICAL MEASURES

Felipe Calderero, Ferran Marques

Department of Signal Theory and Communications
Technical University of Catalonia (UPC), Barcelona, Spain
{felipe.calderero, ferran.marques}@upc.edu

ABSTRACT

A family of statistical region merging approaches based on a general region model is presented. Each region is modeled as an arbitrary first order finite-state Markov process, characterized by its empirical probability transition matrix. Under this premise, the merging problem is formulated from a statistical point of view, leading to two different merging criteria based on information theory statistical measures: the Kullback-Leibler divergence rate and the Bhattacharyya coefficient. In both cases, a size-independent extension of the previous methods, combined with a modified merging order, is also proposed. Finally, all methods are objectively evaluated and compared with other state-of-the-art region merging techniques.

1. INTRODUCTION

Image segmentation can be considered as a first and key step into image analysis and pattern recognition. However, in a large number of cases, a unique solution for the image segmentation problem does not exist. To overcome this situation, a hierarchical segmentation approach can be used where, instead of a single partition, a hierarchy of partitions is provided.

An important type of hierarchical segmentation approaches are *region merging techniques*. Starting from an initial partition or from the collection of pixels, regions are iteratively merged until a stopping criterion is reached. Thus, region merging algorithms are specified by: a *merging criterion*, defining the cost of merging two regions; a *merging order*, determining the sequence in which regions are merged based on the merging criterion; and a *region model* that determines how to represent the union of regions.

Most attention has been focused on improving the merging criteria, e.g. using criteria combination [1, 2], while applying a simpler region model and merging order. Additionally, most approaches assume either color homogeneous [1, 2] or textured [3] regions, leading to separate approaches to image segmentation.

Our motivation is to investigate region merging techniques using more general region models, and providing a general strategy with no assumptions about the nature of the regions. For that purpose, a statistical framework is proposed, where we can use theoretical results in probability theory and information theory, to formally develop merging criteria leading to the most statistically meaningful partitions. The approach proposed here extends our previous work [4], where the image segmentation problem was tackled from a general statistical point of view but under the assumption that a region was a set of independent and identically distributed (i.i.d.) pixels.

Nevertheless, the independence assumption may not be true in general. Maintaining the general statistical framework, here we propose a more realistic approach, restricting statistical independency only to pixels in different regions and considering some dependency for those pixels in the same region.

In the literature, statistical region merging approaches have been previously proposed. However, they lack this generality principle. For instance, some of them consider different parameter-based families of probability distributions as region model [5]. In turn, although [6] considers a non-parametric region model, regions are assumed to hold a homogeneity property: inside any statistical region, pixels have the same expectation value for each color channel.

The paper is structured as follows. Section 2 proposes to statistically characterize a region as a first order Markov process, using as region model the empirical pairwise pixel distribution (or cooccurrence matrix) of its pixels. Under this premise, two region merging criteria are developed in Section 3 from two perspectives, leading to the size-weighted statistical similarity measures based on the Kullback-Leibler divergence rate and to the Bhattacharyya coefficient directly defined for the first order Markov empirical transition matrices. An alternative approach, combining a size-independent extension of the previous methods and a scale-based merging order, is presented in Section 4. Section 5 presents an objective evaluation and comparison with other techniques, using two image data sets. Finally, conclusions are outlined in Section 6.

2. FIRST ORDER FINITE-STATE MARKOV PROCESS AS GENERAL STATISTICAL REGION MODEL

From a statistical viewpoint, a single channel image can be considered as a realization of a 2D stochastic process. Thus, each pixel is a sample of one of the discrete random variables composing the image process. For simplicity, all results in this work are obtained for single channel images, whose extension is straightforward for the multichannel case under channel independence assumption.

Under this framework, the image segmentation problem can be formulated as the partitioning of the image into (statistical) regions, i.e. connected and disjoint subsets of pixels that share similar statistical properties. To formally tackle this problem we consider that statistical dependency is restricted to pixels belonging to the same region. To simplify the statistical analysis, we will further assume that, for each pixel, the statistical dependency is reduced to the values of its neighboring pixels inside the same region.

Constrained to the low complexity required by a region merging approach, we propose a compromise between the difficulty introduce by a 2D dependency and the simplicity of the i.i.d assumption in [4], and hence, to model each region using a 1D first order Markov model. The reduction of the dimensionality for the Markov model is based on considering only the average pairwise dependency of a pixel on its four closest neighbors, which can be seen as the (empirical) probability transition matrix of a first order finite-state Markov process characterizing the region. This empirical pairwise pixel distribution leads to a second order statistic extensively used in texture analysis, known as *cooccurrence matrix* [7].

Formally, given the set of region pixels $\mathbf{x}$ from an alphabet $\mathscr{X} = \{a_1, a_2, \ldots, a_{|\mathscr{X}|}\}$, their cooccurrence matrix $\mathbf{P_x} \in \mathbb{R}^{|\mathscr{X}|\times|\mathscr{X}|}$ is defined as the relative proportion of occurrences of each pair of pixel values of $\mathscr{X}$ separated by a given displacement Δ, i.e.,

$$\mathbf{P_x}(a,b) = \frac{N(x_i = a,\ x_{i+\Delta} = b \mid \mathbf{x})}{N_p} \quad (1)$$

where $N(x_i = a,\ x_{i+\Delta} = b \mid \mathbf{x})$ is the number of times the pixel value a occurs at a given location, while the pixel value b occurs at a displacement Δ from that location; and N_p is the to-

This work has been partly supported by the projects CENIT-2006 i3media and TEC2007-66858/TCM PROVEC of the Spanish Government.

tal number of pairwise pixel occurrences at displacement Δ in $\mathbf{x}$. Under the previous assumption, the considered displacements are $\Delta = \{(0,1),(0,-1),(1,0),(-1,0)\}$. Averaging on these values, a rotation-invariant cooccurrence matrix is obtained.

The proposed cooccurrence matrix provides a general and unified framework for image segmentation. Arbitrary discrete distributions are directly estimated from data, incorporating spatial information not only about the region itself but also about its interactions with adjacent regions (existence of an edge), with no specific assumptions about the nature of the regions (neither homogeneity nor texture). Moreover, this model can be easily computed and, after the union of a pair of regions, updated:

$$\mathbf{P}_{1\cup 2} = \frac{n_1}{n_1+n_2}\mathbf{P}_1 + \frac{n_2}{n_1+n_2}\mathbf{P}_2 \qquad (2)$$

with n_1, n_2 the number of pixels in regions R_1, R_2, respectively.

The quantization of the alphabet $\mathscr{X}$ can be set to optimize the performance of the algorithm. In this work, we only consider a uniform quantization and directly refer to the number of bins in each dimension of the cooccurrence matrix.

The statistical formulation of the merging problem presented in the next sections is based on considering the cooccurrence matrix as the empirical probability transition matrix of the first order finite-state Markov process characterizing a region. We will assume that this Markov process holds the ergodicity property, and hence, it is completely characterized by its initial state and a probability transition matrix. However, note that in this particular case there is no sense to consider an initial state distribution for the sequence of pixels, because the pixels of a 2D region are not ordered. Consequently, we will assume that all initial states are equally likely, i.e, the probability of the state i is set to $\pi_i = \frac{1}{|\mathscr{X}|}$. Under these considerations, a region is completely characterized by the probability transition matrix of the first order Markov process generating it, estimated by its cooccurrence matrix.

3. AREA-WEIGHTED MERGING CRITERIA FOR FIRST ORDER MARKOV REGION MODEL

3.1 Kullback-Leibler Merging Criterion

The first criterion is based on merging at each step the pair of adjacent regions maximizing the probability of being generated by the same statistical distribution. We tackle this problem as a pairwise hypothesis test. Assume R_1 and R_2 are two adjacent regions with empirical transition matrices $\mathbf{P}_1$, $\mathbf{P}_2$, respectively, whose union would generate a new region with empirical transition matrix $\mathbf{P}_{1\cup 2}$. Then, the two hypotheses considered are:

- $\mathbf{H_0}$: Pixels in the first region, $\mathbf{x}_1 \in R_1$, and pixels in the second region, $\mathbf{x}_2 \in R_2$, are both distributed by the same first order Markov process, with probability transition matrix $\mathbf{P}_{1\cup 2}$;
- $\mathbf{H_1}$: Pixels in the first region, $\mathbf{x}_1 \in R_1$, are distributed by the first order Markov transition matrix $\mathbf{P}_1$; and pixels in the second region, $\mathbf{x}_2 \in R_2$, are distributed by the first order Markov process, with transition matrix $\mathbf{P}_2$.

Similarly to the Neyman-Pearson lemma for i.i.d. observations, in [12] it is proved that the best achievable error exponent for testing between two stationary and irreducible Markov sources (thus, ergodic Markov processes) is given by the *likelihood ratio test*:

$$\frac{P_{H_0}(x_1,x_2,\ldots,x_n)}{P_{H_1}(x_1,x_2,\ldots,x_n)} \gtrless T \qquad (3)$$

Considering that the probability of a first order Markov sequence can be written as $P(\mathbf{x}) = P(x_1)\prod_{i=2}^{n} P(x_i|x_{i-1})$ and referring to $\mathbf{x}$ as the concatenation of the pixels of both regions, i.e., $\mathbf{x} = (\mathbf{x}_1,\mathbf{x}_2)$, the log-likelihood ratio can be formulated as:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} = \log\frac{P_{1\cup 2}(x_1)}{P_1(x_1)P_2(x_1)} + \sum_{i=2}^{n}\log P_{1\cup 2}(x_{i-1}|x_i) - \sum_{i=2}^{n_1}\log P_1(x_{i-1}|x_i) - \sum_{i=n_1+2}^{n}\log P_2(x_{i-1}|x_i)$$

where n_1, n_2 are the number of pixels in R_1 and R_2, respectively; and $n = n_1 + n_2$.

As the Markov process modeling each region holds the ergodicity property (as stated in Section 2), by the ergodic theorem, each one of the terms $\sum_i \log P(x_{i-1}|x_i)$ approaches the statistical average with probability 1 under the probability distribution P. For instance,

$$\sum_{i=2}^{n}\log P_{1\cup 2}(x_{i-1}|x_i) = (n-1)\cdot E\{\log P_{1\cup 2}(x_{i-1}|x_i)\} \qquad (4)$$

where $E\{.\}$ corresponds to the statistical mean under the distribution of $P_{1\cup 2}$. Considering $P_{1\cup 2}(x_1 = a_i) = \pi_i^{1\cup 2}$ as the initial states distribution of the process, and $\mathbf{P}_{1\cup 2} = (p_{ij}^{1\cup 2}) = P_{1\cup 2}(x_{i-1}|x_i)$ as the transition matrix,

$$E\{\log P_{1\cup 2}(x_{i-1}|x_i)\} = \sum_{i\in\mathscr{X}}\pi_i^{1\cup 2}\sum_{j\in\mathscr{X}} p_{ij}^{1\cup 2}\log p_{ij}^{1\cup 2} = -H(P_{1\cup 2})$$

where $H(P) = -\sum_{i,j\in\mathscr{X}}\pi_i p_{ij}\log p_{ij}$ is the Shannon entropy rate of the first order Markov process [11].

Thus, we can rewrite the log-likelihood ratio test in terms of the Shannon entropy rate of the processes as:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} = \log\frac{P_{1\cup 2}(x_1)}{P_1(x_1)P_2(x_1)} - (n-1)\cdot H(P_{1\cup 2}) + (n_1-1)\cdot H(P_1) + (n_2-1)\cdot H(P_2) \qquad (5)$$

In general, for n_1 and n_2 sufficiently large, the first term can be dismissed. Under this condition, the log-likelihood ratio can be approximated as:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx -n\cdot H(P_{1\cup 2}) + n_1\cdot H(P_1) + n_2\cdot H(P_2) \qquad (6)$$

that can be interpreted as the size-weighted decrement on the entropy rate when the regions are merged. Considering (2), the equiprobable initial state assumption, and using the Kullback-Leibler divergence rate between a first order Markov sequence of n samples p^n with stationary distribution π and transition matrix $\mathbf{P} = (p_{ij})$, and another first order Markov sequence of n observations q^n, with transition matrix $\mathbf{Q} = (q_{ij})$:

$$\lim_{n\to\infty}\frac{1}{n}D(p^n||q^n) = \sum_{i,j\in\mathscr{X}}\pi_i\cdot p_{ij}\cdot\log\frac{p_{ij}}{q_{ij}} \qquad (7)$$

then, (6) can be rewritten as:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx -n_1 D(P_1||P_{1\cup 2}) - n_2 D(P_2||P_{1\cup 2}) \qquad (8)$$

or equivalently,

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx \frac{1}{|\mathscr{X}|}\left(n_1\sum_{i,j}p_{ij}^1\log\frac{p_{ij}^{1\cup 2}}{p_{ij}^1} + n_2\sum_{i,j}p_{ij}^2\log\frac{p_{ij}^{1\cup 2}}{p_{ij}^2}\right) \qquad (9)$$

Defining the Kullback-Leibler divergence between transition matrices as:

$$D(\mathbf{P}||\mathbf{Q}) \triangleq \sum_{i,j}p_{ij}\log\frac{p_{ij}}{q_{ij}} \qquad (10)$$

we can rewrite the previous expression depending only on the transition matrices of the candidate regions:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \propto -n_1 D(\mathbf{P}_1||\mathbf{P}_{1\cup 2}) - n_2 D(\mathbf{P}_2||\mathbf{P}_{1\cup 2}) \qquad (11)$$

Consequently, at each merging stage, the two adjacent regions (written as $R_i \sim R_j$) with maximum log-likelihood should be merged. We will refer to this statistical criterion as the *Markov Kullback-Leibler merging criterion* (M-KL), formally stated as:

$$\{R_1,R_2\} = \arg\max_{R_i\sim R_j} -n_i\cdot D(\mathbf{P}_i||\mathbf{P}_{i\cup j}) - n_j\cdot D(\mathbf{P}_j||\mathbf{P}_{i\cup j}) \qquad (12)$$

This criterion measures the similarity between the empirical probability transition matrices of the regions and the empirical transition matrix of their merging, weighted by the size of the regions.

3.2 Bhattacharyya Merging Criterion

In this section we present a new criterion based on a direct statistical comparison between the cooccurrence matrices of the regions. Nevertheless, in this case, the Kullback-Leibler divergence becomes impractical, as its convergence cannot be warranted anymore.

We tackle the problem from a different perspective. As introduced in [4], each region can be seen as a class centered at the point generated by its empirical distribution on the probability simplex. The exponent of the probability of error of such a classifier is bounded by the minimum *Chernoff information* between the statistical distribution of any pair of classes [8].

Proceeding analogously to the classical derivation of the Chernoff bound for the i.i.d. case, we can obtain a similar result for first order Markov sequences. As shown in [12], the posteriori probability decision rule minimizes the Bayesian probability of error for testing between hypotheses H_1 and H_2. Being A the decision region for H_1, the probability of error for this rule is

$$P_e = \sum_{A^C} \pi_1 P_1 + \sum_{A} \pi_2 P_2 \tag{13}$$

$$= \sum_{x \in \mathscr{X}} \min\{\pi_1 P_1, \pi_2 P_2\} \tag{14}$$

where A^C refers to the complementary region of A, i.e. the decision region for H_2. Now for any two positive numbers a and b, we have

$$\min\{a,b\} \leq a^\lambda b^{1-\lambda} \quad \text{for all } 0 \leq \lambda \leq 1 \tag{15}$$

Using this to continue the chain, we have

$$P_e \leq \sum_{x \in \mathscr{X}} (\pi_1 P_1)^\lambda (\pi_2 P_2)^{1-\lambda} \tag{16}$$

$$\leq \sum_{x \in \mathscr{X}} P_1^\lambda P_2^{1-\lambda} \tag{17}$$

For a sequence of observations, we have

$$P_e^{(n)} \leq \sum_{x^n \in \mathscr{X}^n} (P_1^{(n)})^\lambda (P_2^{(n)})^{1-\lambda} \tag{18}$$

Particularizing for first order Markov sequences,

$$P_e^{(n)} \leq \sum_{x^n \in \mathscr{X}^n} P_1^\lambda(x_1) P_2^{1-\lambda}(x_2) \prod_{k=2}^{n} P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1})$$

As in Section 2, we can assume all initial states equally likely, which removes the dependency on the state probabilities. Thus:

$$\begin{aligned} P_e^{(n)} &\leq \sum_{x^n \in \mathscr{X}} \prod_{k=2}^{n} P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1}) \\ &\leq \prod_{k=2}^{n} \sum_{x_k, x_{k-1}} P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1}) \\ &\leq \prod_{k=2}^{n} \sum_{i,j} P_1^\lambda(i,j) P_2^{1-\lambda}(i,j) \\ &= \left(\sum_{i,j} P_1^\lambda(i,j) P_2^{1-\lambda}(i,j) \right)^{n-1} \end{aligned} \tag{19}$$

where $P_1(i,j)$, $P_2(i,j)$ correspond to the transition matrices of the processes, respectively (referred as $\mathbf{P}_1 = (p_{ij}^1)$ and $\mathbf{P}_2 = (p_{ij}^2)$ in the previous sections, but changed here to simplify the notation).

Hence, the exponent of the error probability is bounded by:

$$\log P_e^{(n)} \leq (n-1) \log \left(\sum_{i,j} P_1^\lambda(i,j) P_2^{1-\lambda}(i,j) \right) \tag{20}$$

When sequences have a different number of samples, an identical upper bound on the probability of error can be obtained considering only the smaller number of samples in both sequences.

Analogously to the i.i.d. case, we can define the Chernoff information between the probability transition matrices of two first order Markov processes, as the minimum exponent of error given by (20):

$$C(\mathbf{P}_1, \mathbf{P}_2) \triangleq - \min_{0 \leq \lambda \leq 1} \log \left(\sum_{i,j} P_1^\lambda(i,j) P_2^{1-\lambda}(i,j) \right) \tag{21}$$

We propose to merge the pair of regions with maximum Chernoff information, redefining the probability of error of a classifier as the probability of fusion in a clustering method. This way, the bound on the probability of merging for two adjacent regions, with empirical probability transition matrices $\mathbf{P}_i$, $\mathbf{P}_j$, and number of pixels n_i, n_j, respectively, can be written as:

$$P_{\text{merging}}(R_i, R_j) \leq e^{-\min(n_i, n_j) \cdot C(\mathbf{P}_i, \mathbf{P}_j)} \tag{22}$$

Nevertheless, computing the Chernoff information implies an optimization over λ. To reduce this computational load, we propose to approximate the Chernoff information by the *Bhattacharyya coefficient* between the transition matrices ($\lambda = 1/2$):

$$B(\mathbf{P}_i, \mathbf{P}_j) \triangleq - \log \left(\sum_x \mathbf{P}_i^{\frac{1}{2}}(x) \mathbf{P}_j^{\frac{1}{2}}(x) \right) \tag{23}$$

In conclusion, a statistical clustering approach leads to the merging of the adjacent pair of regions with maximum (bound of the) probability of fusion, or equivalently, maximizing its exponent:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} - \min(n_i, n_j) \cdot B(\mathbf{P}_i, \mathbf{P}_j) \tag{24}$$

This method is based on a size-weighted direct statistical measure of the empirical probability transition matrices, and we will refer to it as the *Markov Bhattacharyya merging criterion* (M-BHAT).

4. AREA-UNWEIGHTED MERGING CRITERIA FOR FIRST ORDER MARKOV REGION MODEL

The obtained merging costs depend on the size of the involved regions, establishing, in some sense, the confidence of the estimated empirical models. This approach assures that the resulting partitions are size consistent, meaning that the area of the regions tends to increase as the number of regions into the partition decreases.

The size term prioritizes the fusion of the smaller regions, slowing the merging of the larger regions, even when they are similarly distributed. Here, we propose an extension of the previous methods providing a trade-off between under- and oversegmentation, while increasing the size resolution of the partition. This is achieved by removing the size dependency from the merging criteria and incorporating it into the merging order to assure size consistency.

Hence, under the assumption that regions are large enough to have a high confidence on the estimated transition matrices, the area dependency can be removed from the previous merging criteria:

- Area-unweighted Kullback-Leibler merging criterion:
$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -D(\mathbf{P}_i || \mathbf{P}_{i \cup j}) - D(\mathbf{P}_j || \mathbf{P}_{i \cup j}) \tag{25}$$
- Area-unweighted Bhattacharyya merging criterion:
$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -B(\mathbf{P}_i, \mathbf{P}_j) \tag{26}$$

In practice, we cannot always assure that the distribution of all regions is perfectly estimated, specially, in early stages of the merging process. For this reason, and to assure the size consistency of the partitions, an agglomerative force is needed into the merging process. Our proposal is to combine the criteria in (25) and (26) with a *scale-based merging order*, incorporating the size consistency constraints. The idea is to define a scale threshold for each level of resolution. A region below this threshold is considered as *out-of-scale* and merged with the highest priority, fusing it with its most similar region in the partition. Finally, when no out-of-scale regions remain, the algorithm continues merging *in-scale* regions normally. At each merging step the scale threshold is updated, and normal merging continues till new out-of-scale regions appear. The scale threshold is defined as:

$$T_{\text{scale}} = \alpha \cdot \frac{\text{Image Area}}{\text{Number of Regions}} \tag{27}$$

The α parameter controls the minimum resolution at each scale. Heuristically, we have found that values around $\alpha = 0.15$ provide a good compromise between under- and oversegmentation.

The benefit of this approach is that the fusion of large regions is not penalized, once out-of-scale regions have been removed. All regions are equally likely to merge despite its size, because the merging cost only measures the statistical similarity of the empirical transition matrices, without being size biased.

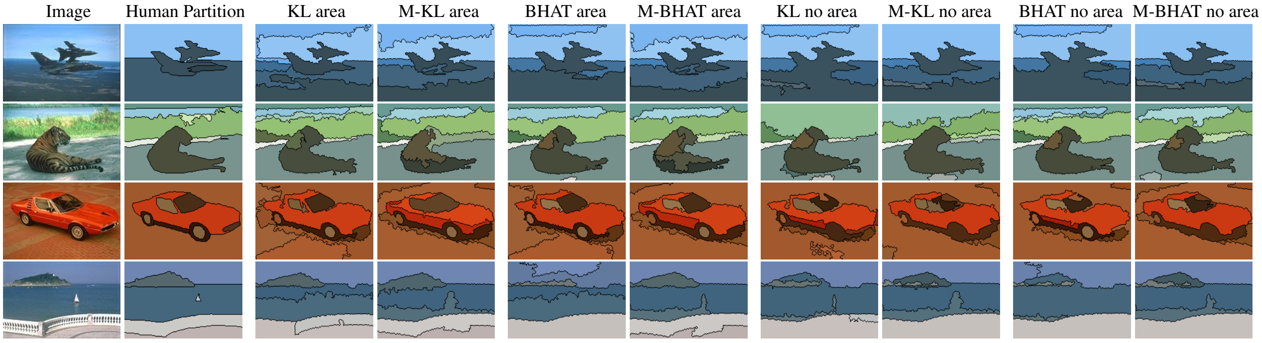

Figure 1: Merging criteria comparison for the subset of the Corel© database. In all methods, region models are quantized to 5 bins.

Symmetric Distance		
	1. Method in [1]	0.3809
	2. KL Area-weighted	0.3191
	3. KL Area-unweighted	0.2372
	4. BHAT Area-weighted	0.3081
	5. BHAT Area-unweighted	0.2275
	6. Markov KL Area-weighted	0.3255
	7. Markov KL Area-unweighted	0.2624
	8. Markov BHAT Area-weighted	0.3263
	9. Markov BHAT Area-unweighted	0.2573

Table 1: Symmetric distance for the subset of the Corel© database.

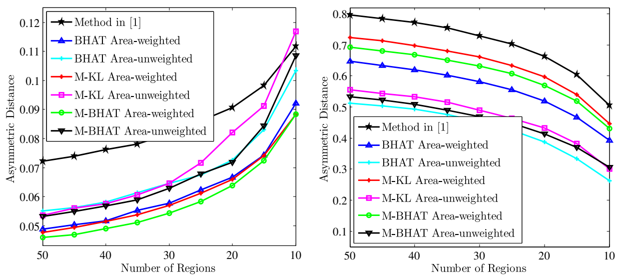

Figure 2: Asymmetric distance for the Corel© subset database. Left: from computed to ground truth partition (oversegmentation); right: viceversa (undersegmentation). Statistical methods were computed using matrices quantized to 5 bins per dimension.

5. EXPERIMENTAL RESULTS

All region merging techniques compared in this section were applied on an initial partition of the original image in HSV color space, computed using the watershed algorithm. The first set of experiments was performed over a set of 100 images from the Corel© image database [1]. Ground truth partitions were manually segmented in the context of the SCHEMA project (http://www.iti.gr/SCHEMA/).

To evaluate the quality of the partitions created by the proposed methods, we use the distances defined in [9]. Initially, a symmetric distance is proposed $d_{\text{sym}}(P,Q)$. This distance is defined in terms of the minimum number of pixels whose labels should be changed between regions in P to achieve a perfect matching with Q (P and Q become identical), normalized by the total number of pixels in the image. This measure evaluates the global quality of a partition, and its compromise between under- and oversegmentation.

The definition is extended to an asymmetric distance: $d_{\text{asym}}(P,Q)$. In this case, it measures the minimum number of pixels whose labels should be changed so that partition Q becomes finer than partition P, normalized by the image size. Note that, in general, $d_{\text{asym}}(P,Q) \neq d_{\text{asym}}(Q,P)$. When P is the partition to evaluate and Q its ground truth partition, the first ordering measures the degree of oversegmentation, and the second, the undersegmentation in P with respect to the ground truth partition.

Table 1 shows the mean symmetric distance between ground truth partitions and partitions generated by the proposed methods, both with the same number of regions. These results are compared with the region merging technique proposed in [1] and the statistical region merging techniques in [4], using the same watershed-based initial partitions. The merging criterion in [1] combines color similarity and contour complexity of the regions, normalized by the component dynamic range, and was shown to outperform most color based merging techniques. As early commented, [4] presents the analogous area-weighted and area-unweighted statistical techniques for a simplified statistical model, where statistical independence is assumed for all image pixels. The merging criteria are based on the Kullback-Leibler divergence and the Bhattacharyya coefficient for the normalized histograms of the regions or empirical distributions. We refer to these methods as KL area-weighted, BHAT area-weighted, KL area-unweighted, BHAT area-unweighted. Note that all statistical criteria, Markov-based or not, outperform [1] and, as expected, area-unweighted methods present the best trade-off between under- and oversegmentation. The symmetric distance is slightly larger for Markov-based methods compared to methods in [4]. Nevertheless, in Fig. 1 (see *http://gps-tsc.upc.es/imatge/_Felipe/eusipco08/* for an extension) it can be seen that the subjective quality of the partitions is similar or slightly superior for the Markov-based strategies.

In Figure 2, the results for the mean asymmetric distance for different numbers of regions are presented. For the sake of clarity, only the best area-weighted and area-unweighed methods in [4] are shown (corresponding in both cases to the Bhattacharyya merging criteria). Figure 2-left shows $d_{\text{asym}}(P,Q)$, measuring the degree of oversegmentation. In this case, area-weighed methods outperform area-unweighted methods, and generally, Markov-based techniques are slightly superior. The Markov-based Bhattacharyya area-weighted method presents the most significant improvement with respect to the best technique in [4] (a 7% distance decrement). On the contrary, in Fig. 2-right, for $d_{\text{asym}}(Q,P)$, Markov-based techniques suffer from more undersegmentation than techniques in [4]. Hence, as observed in [4], there is a compromise between under- and oversegmentation. Note that in all cases the Bhattacharyya versions are superior to Kullback-Leibler ones.

Another evaluation was performed on the benchmark system presented in [10]. It contains a set of synthetic mosaics of natural textures and their corresponding ground truth partitions, allowing an online evaluation and comparison with other state-of-the-art techniques with respect to a large set of indicators, divided in three classes: region-based, pixel-wise average and error consistency. A complete description of these methods, as well as the results for all the proposed methods and a comparison with other state-of-the-art texture segmentation techniques are also available online at *http://mosaic.utia.cas.cz/*. In this context, the evaluation was performed in a supervised manner, i.e., the number of regions in the evaluated partitions was manually set to the number of regions in the ground truth partitions.

All area-weighted statistical approaches, Markov or not, outperform the rest of techniques in the benchmark. Nevertheless, due to size homogeneity in the textured regions generating the mosaics,

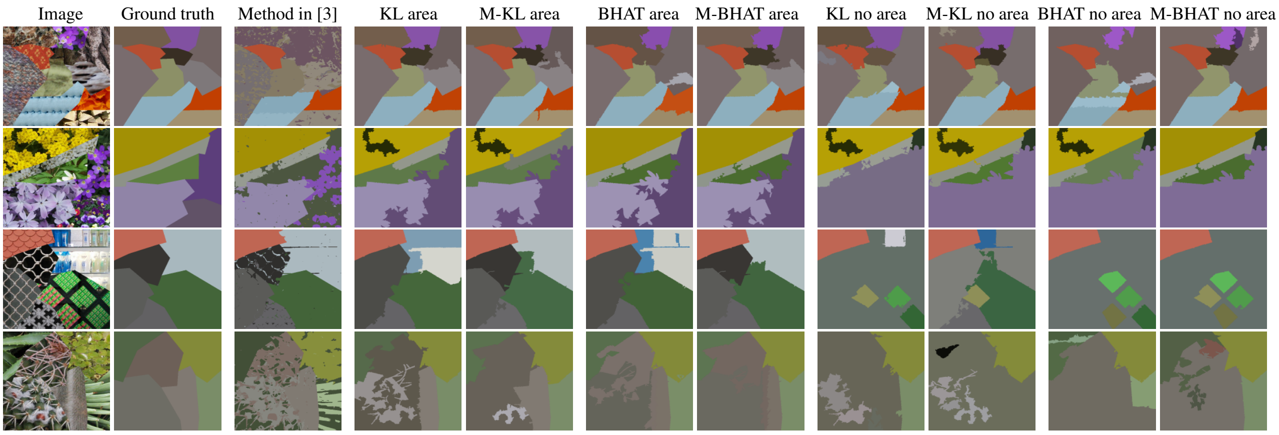

Figure 3: Merging criteria comparison for the texture database in [10]. In all statistical methods, region models are quantized to 10 bins.

	KL Area	M-KL Area	M-BHAT Area	BHAT Area	TFR/KLD	BHAT NoArea	M-BHAT NoArea	KL NoArea	M-KL NoArea
↑*CS*	68.72	67.55	64.07	61.12	51.25	30.14	28.66	26.52	21.25
↓ *OS*	9.00	11.36	8.59	9.32	5.84	8.02	5.38	6.92	6.13
↓ *US*	6.67	9.11	6.49	6.09	7.16	51.23	56.22	46.13	49.95
↓ *ME*	15.09	12.58	18.58	20.61	31.64	14.95	9.41	21.50	24.57
↓ *NE*	15.16	13.54	17.64	19.76	31.38	11.63	7.90	20.99	23.37
↓ *O*	15.26	15.70	15.56	16.44	23.6	38.59	38.15	41.10	46.73
↓ *C*	15.26	15.70	15.56	16.44	22.42	38.59	38.15	41.10	46.73
↑ *CA*	78.90	78.07	77.86	76.73	67.45	47.27	46.74	44.41	37.52
↑ *CO*	84.74	84.30	84.44	83.56	76.40	61.41	61.85	58.90	53.27
↑ *CC*	89.30	87.19	87.76	88.10	81.12	59.94	66.02	64.55	61.10
↓ *I.*	15.26	15.70	15.56	16.44	23.60	38.59	38.15	41.10	46.73
↓ *II.*	2.10	2.91	2.92	2.68	4.09	15.45	15.82	15.74	18.42
↑ *EA*	85.01	84.32	84.44	83.77	75.80	54.30	53.90	52.16	45.16
↑ *MS*	77.12	76.44	76.66	75.33	65.19	42.12	42.78	38.35	29.90
↓ *RM*	4.54	4.26	4.61	4.94	6.87	15.00	15.39	15.95	19.35
↑ *CI*	85.98	85.01	85.25	84.76	77.21	56.98	57.32	56.08	49.58
↓ *GCE*	13.29	13.33	14.13	15.60	20.35	11.25	10.00	13.96	12.46
↓ *LCE*	6.93	7.21	7.36	8.64	14.36	5.06	5.31	6.67	6.53
↓ *dM*	6.84	7.61	8.07	8.32	12.64	32.08	32.80	32.62	41.93
↓ *dD*	10.88	10.85	11.18	12.37	18.01	21.58	21.45	23.06	26.02
↓ *dVI*	14.16	14.13	14.14	14.20	14.06	12.06	11.87	12.31	11.59

Table 2: For indicators with up arrow, larger values are preferred; for down arrows, the opposite. For each parameter, the first and second best values among all methods is shown in blue (dark grey) and orange (light grey), respectively. Statistical methods were quantized to 10 bins. Benchmark criteria are explained in [10].

area-unbiased methods present a lower performance. Table 2 outlines the evaluation results for the statistical methods proposed here and those in [4], using as reference the best texture segmentation technique into the benchmark [3]. Some examples for these techniques are shown in Fig. 3. Again, it can be seen that the subjective quality of the partitions generated by area-weighted statistical methods is superior to partitions obtained by [3]. Markov-based approaches and methods in [4] provide similar segmentation results.

6. CONCLUSIONS

We have experimentally confirmed that the set of region merging techniques using a richer and more general statistical region model provides higher quality partitions. Concretely, the area-weighted methods exhibit an excellent performance in terms of minimizing the merging error or undersegmentation. Nevertheless, an area-unweighted extension of the previous methods has been proposed to obtain a better trade-off between under- and oversegmentation. In general, the quality of the obtained partitions is as good as or slightly better than those generated by the statistical methods using pixel independence assumptions. Markov-based methods can be useful in applications where the segmentation errors may be crucial, while applications being more error tolerant or computationally constrained may find into the simple statistical model a more valuable solution.

REFERENCES

[1] V. Vilaplana, F. Marques, "On Building a Hierarchical Region-Based Representation for Generic Image Analysis", *Proc. ICIP'07*, vol. 4, pp. 325–328, Sept. 2007.

[2] T. Adamek, N.E. O'Connor, "Using Dempster-Shafer Theory to Fuse Multiple Information Sources in Region-Based Segmentation", *Proc. ICIP'07*, vol. 2, pp. 269–272, Sept. 2007.

[3] G. Scarpa, M. Haindl, J. Zerubia, "A Hierarchical Finite-State Model for Texture Segmentation", *Proc. ICASSP'07*, vol. 1, pp. 1209–1212, April 2007.

[4] F. Calderero, F. Marques, "General Region Merging Approaches Based on Information Theory Statistical Measures", *Submitted to ICIP'08* (available online at *http://gps-tsc.upc.es/imatge/_Felipe/submitted/*).

[5] V. Gies, T.M. Bernard, "Statistical solution to watershed oversegmentation", *Proc. ICIP'04*, vol. 3, pp. 1863–1866, Oct. 2004.

[6] R. Nock, F. Nielsen, "Statistical region merging", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp.1452–1458, Nov. 2004.

[7] R.M. Haralick, "Statistical and structural approaches to texture", *Proceedings of IEEE*, vol. 67, no. 5, pp. 786–804, May. 1979.

[8] T. Cover, J. Thomas, *Elements of Information Theory*, New York: John Wiley & Sons, Inc., second edition, 2006

[9] J.S. Cardoso, L. Corte-Real, "Toward a generic evaluation of image segmentation", *IEEE Trans. Image Process.*, vol. 14, no. 11, pp. 1773–1782, Nov. 2005.

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

[11] Z. Rached, F. Alajaji, L.L. Campbell, "The Kullback-Leibler divergence rate between Markov sources", *IEEE Trans. Information Theory*, vol. 50, no. 5, pp. 917–921, May 2004.

[12] Z. Rached, "Information Measures for Sources with Memory and Their Application to Hypothesis Testing and Source Coding", *PhD. Thesis*, Queen's University, Ontario, Canada. August 2002.