Region Merging Techniques Using Information Theory Statistical Measures

Felipe Calderero, *Student Member, IEEE*, and Ferran Marques, *Senior Member, IEEE*

Abstract—The purpose of the current work is to propose, under a statistical framework, a family of unsupervised region merging techniques providing a set of the most relevant region-based explanations of an image at different levels of analysis. These techniques are characterized by general and nonparametric region models, with neither color nor texture homogeneity assumptions, and a set of innovative merging criteria, based on information theory statistical measures. The scale consistency of the partitions is assured through i) a size regularization term into the merging criteria and a classical merging order, or ii) using a novel scale-based merging order to avoid the region size homogeneity imposed by the use of a size regularization term. Moreover, a partition significance index is defined to automatically determine the subset of most representative partitions from the created hierarchy. Most significant automatically extracted partitions show the ability to represent the semantic content of the image from a human point of view. Finally, a complete and exhaustive evaluation of the proposed techniques is performed, using not only different databases for the two main addressed problems (object-oriented segmentation of generic images and texture image segmentation), but also specific evaluation features in each case: under- and oversegmentation error, and a large set of region-based, pixel-based and error consistency indicators, respectively. Results are promising, outperforming in most indicators both object-oriented and texture state-of-the-art segmentation techniques.

Index Terms—Bhattacharyya coefficient, image region analysis, image segmentation, information theory, Kullback–Leibler divergence, region merging.

I. INTRODUCTION

IMAGE segmentation is a first and key step for image analysis and pattern recognition [1]. Its goal is twofold: from a semantic point of view, image segmentation is a first level of abstraction providing an image representation closer to the object representation than the set of pixels; and from a practical point of view, a region-based representation of the image reduces the number of elementary primitives and allows a more robust estimation of parameters and descriptors. In other words, segmentation simplifies the image providing a representation that is more semantically meaningful and easier to analyze [2].

However, in a large number of cases, a unique solution for the image segmentation problem does not exist (for that reason, sometimes image segmentation is referred as an ill-posed problem[1]), i.e., instead of a single optimal partition, it is possible to find different region-based explanations of an image, at different levels of analysis or detail [3]. To overcome this situation a hierarchical segmentation approach is needed, where instead of a single partition, a hierarchy of partitions is provided [4].

An important type of hierarchical bottom-up segmentation approaches are *region merging techniques* [5]. These techniques are region-based, in the sense that they consider regions not only as the goal of the segmentation process but also as the mean to obtain a partition of the image. Hence, local decisions are directly based on the region properties and features. Starting from an initial partition or from the collection of pixels, regions are iteratively merged until a termination criterion is fulfilled (for instance, a unique region is reached).

Region merging algorithms can be specified by [6]: a *merging criterion* that defines 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. They can be efficiently implemented using graph-based approaches such as the *recursive shortest spanning tree* (RSST) algorithm [7]. Fast implementations of this algorithm can significantly decrease its computational load [8] or even bound its time complexity in the worst case to $\mathcal{O}(n)$ for not significantly complex region models and merging criteria, with n being the number of regions in the initial partition [9].

In the literature, there is an explicit division between two types of region models. For the first type, where the color of the pixels belonging to the region is assumed to be approximatively constant, first-order statistics such as mean [10] or median [11] color values are used as region model. For instance, this assumption is common in object-oriented image segmentation. For the second type, where region merging is applied to texture segmentation, region models are based on second or higher order statistics [12] or in transformations, such as wavelets [13], [14] or Gabor filters [15], [16].

Specially for object-oriented or content-based applications, most researchers have focused their attention on the merging criteria. The basic criteria have relied on color homogeneity, for instance, MSE, Euclidean distance between region color mean or median (or a weighted version), single linkage or complete linkage [17]. These approaches may suffer from small

---

[1]Another reason to refer to the image segmentation problem as ill-posed is that conflicting constraints may lead to a cost function that does not get one closer to a optimal solution by making an improvement in cost.

and meaningless regions into the generated partitions. Other methods overcome this problem introducing a regularization term based on the size of the regions [11]. These merging criteria penalize the merging of large regions, giving priority to the fusion between small regions or regions with similar number of pixels. Although this approach helps to eliminate small meaningless regions, it introduces a bias into the merging process that tends to create partitions with regions of similar sizes. Other techniques incorporate, apart from color and size, contour complexity of the region into the merging criterion [10], for instance, based on the perimeter length or area-perimeter ratio [18]. In turn, some researchers have tried to emulate human perceptual grouping using more complex criteria, for instance defining syntactic features [19], [20].

The merging order has received little attention in the region merging literature. Typically, the merging order is based on iteratively merging the pair of regions with lowest merging cost, given by the chosen merging criterion.

The purpose of the current work is to propose, under a statistical framework, a family of unsupervised region merging techniques with the following characteristics:

- a *general statistical region model*, providing a unified segmentation strategy where arbitrary and non parametric discrete distributions are directly estimated from data and neither color nor texture homogeneity are assumed inside the region;

- a set of *innovative merging criteria*, based on information theory statistical measures between the region models;

- assuring the *scale consistency* of the partitions through two strategies: i) a size regularization term into the merging criteria (in the rest of the paper, we also refer to these methods as area-weighted) and a classical merging order; or ii) a new *scale-based merging order* combined with a size independent criterion (referred also as area-unweighted criterion) to avoid the region size homogeneity imposed by the use of a size regularization term;

- moreover, the definition of a *partition significance index* to automatically determine the subset of most representative partitions from the created hierarchy;

- and the ability of these most significant automatically extracted partitions to *represent the semantic content* of the image from a human point of view.

As just commented, this analysis is performed under a statistical framework, where we can take advantage of well-known results in probability theory and information theory. The set of information theoretical region merging techniques proposed in this work is outlined in Fig. 1.

Nevertheless, the statistical framework is not new in image segmentation. For instance, Markov random fields segmentation techniques have been extensively used [21]; approaches based on the mean-shift algorithm [22] have also been proposed; or even classical techniques such as the well-known Mumford–Shah functional [10] have been interpreted as a parametric maximum *a posteriori* (MAP) estimation [23].

Particularly, other statistical region merging techniques are also found in the literature. Nevertheless, most of these approaches are based on parametric probability region models under color homogeneous [24], [25] or texture [14] assump-
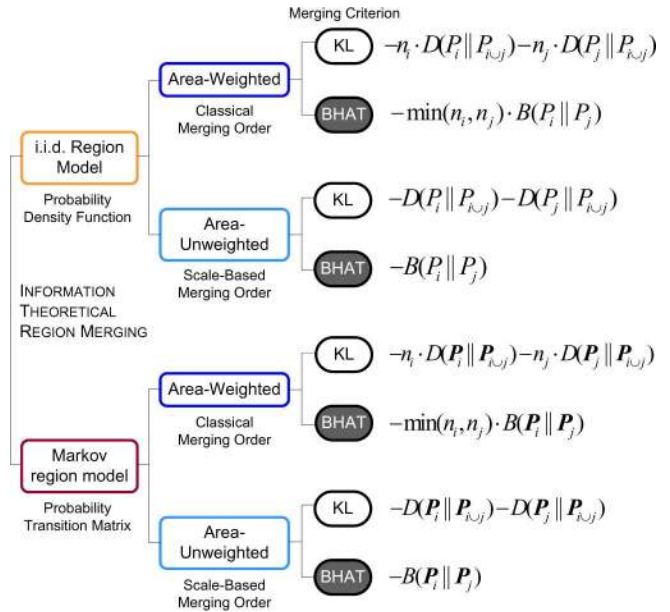

Fig. 1. Information theoretical region merging techniques. The combination of the proposed region models, merging criteria and merging orders leads to eight different techniques, where n, P, and $\mathbf{P}$ refer to the area, the probability density function, and the probability transition matrix of a region, respectively. The Kullback–Leibler divergence and the Bhattacharyya coefficient are represented as $D(\cdot \| \cdot)$ and $B(\cdot \| \cdot)$.

tions. Hence, these techniques cannot be used in a general scenario but are restricted to a particular type of images.

Approaches without strong probability model assumptions for the regions are less common. For instance, a semiparametric statistical approach is presented in [27] and [29]. In these approaches, image segmentation is formulated as an inference problem. The authors assume that optimal statistical regions have a homogeneity property; i.e., inside any statistical region, the pixels have the same expectation value for each color channel. Hence, pixels are considered statistically independent and having this homogeneity condition, although they are not assumed to be identically distributed. We refer to these techniques as semiparametric because the pixel distribution is modeled by a set of Q independent random variables, where Q is a parameter that has to be set, controlling the statistical complexity of the optimally segmented image. The merging criterion relies on the definition of a merging threshold, based on an upper bound on the difference of the expectation of the pixel value of each region. In [30]–[32], a supervised version of the previous methods is presented, i.e., a segmentation algorithm requiring some user guidance.

Although being more general than parametric approaches, in the previous algorithm the number of random variables depends on the image complexity and, consequently, it is difficult to estimate when no prior knowledge on the image is available. Additionally, some of the resulting partitions suffer from scale inconsistency, i.e., small meaningless regions do not necessarily merge as the scale gets coarser. In this sense, it is difficult that the color homogeneous parts of a texture region can merge to form a coarser level statistical distribution without using a scale consistency mechanism, specially in early stages of the merging process where assumptions such as the homogeneity property
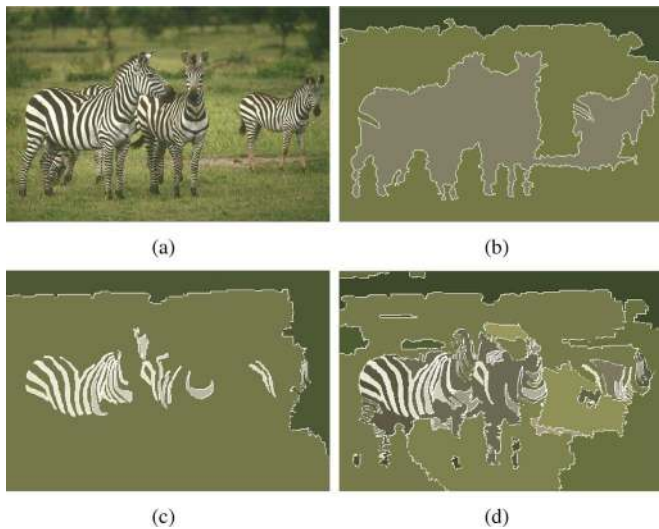
Fig. 2. Illustrative example of a coarse level texture. (a) Original image extracted from the Berkeley Segmentation Dataset [26]. (b) Segmentation obtained by one of our techniques (concretely, using the empirical distribution of the region quantized to ten bins as statistical model and a merging criterion based on the Bhattacharyya coefficient, see Section II-C). The partition shown was automatically determined by the proposed significance index as the most relevant partition in the whole hierarchy (see Section V). (c) Partition provided by the method in [27] with $Q = 32$. Note that this method is not able to capture the zebra black and white texture. (d) Partition provided by the method in [27] with $Q = 128$. Partitions (c) and (d) were obtained using the online available application created by the authors of [27] (http://www.sonycsl.co.jp/person/nielsen/SRM/). This Java implementation is the simplest code provided to the internet community (see also [28] for the C++ code and details).

do not hold. This fact is illustrated in Fig. 2. A region containing the white stripes of a zebra and another region containing the black stripes do not hold the homogeneity property when they are considered independently. Our approach can correctly deal with low scale textures thanks to the preservation of the size consistency of the partitions and the use of accurate information theoretical merging criteria. Moreover, note the ability of the most relevant partition automatically proposed by the significance index [see Fig. 2(b)] to approximately represent the most human-representative semantic content of the image.

Finally, the proposed merging techniques have been evaluated in two different contexts. First, we have performed a complete evaluation in terms of object-oriented segmentation and semantic analysis of generic images. Second, an exhaustive evaluation in terms of natural texture segmentation is provided. In each case we use not only a specific database with available ground truth partitions, but also concrete evaluation features.

For the object-oriented evaluation, the most important types of errors, namely, undersegmentation (merging regions belonging to different objects) and oversegmentation (not merging regions belonging to the same object), are measured using the metrics proposed in [33] and an extension of the methodology in [34].

Our methods are directly compared with the region merging technique proposed in [35]. In this case, the proposed techniques using area-weighted merging criteria obtain similar or better results in terms of undersegmentation (1.5% mean decrease) while clearly outperforming in terms of oversegmentation (10% mean

decrease). The techniques using a scale-based merging order lead to a compromise between under- and oversegmentation error (for instance, a sacrifice of a 9% undersegmentation increase can result into a 30% oversegmentation decrease).

For the texture evaluation, the selected database has associated a system that automatically evaluates and compares the results with nine state-of-the-art texture segmentation algorithms and presents the results on-line. Concretely, a total of 21 indicators are evaluated. In this case, the evaluation is performed on a supervised and on an unsupervised manner. In both cases, the proposed techniques based on a size dependent merging criteria show a good performance in most of the indicators, clearly outperforming eight out of the nine algorithms into the benchmark and being comparable or superior to the best technique (http://mosaic.utia.cas.cz/).

The work presented here completes our preliminary contributions in [36] and [37], first by including a more detailed analytical development of the information theoretical merging criteria, and specially in terms of a deeper and more exhaustive objective evaluation. Concretely, the criteria are proved to be optimal in terms of maximizing the likelihood of the merged regions (criteria in Sections II-B and III-B) or derived from an upper bound on the classification error between a pair of regions (see Sections II-C and III-C and Appendix, this last one including a proof for an extension of the Chernoff bound for first-order Markov processes). The possibility of developing these mathematical proofs has made us chosen these two merging criteria among the myriads of existing information theory statistical distances (for instance, Bregman divergences [38] or Csiszár f-divergences [39]). For the object-oriented evaluation, a study of the performance in terms of the parameter selection for the scale-based merging order and the objective evaluation of the ability of the most significant partition to represent the semantic content are originally included here. For the texture evaluation, a deeper interpretation of the supervised evaluation results and the whole unsupervised evaluation are exclusively presented here. Additional examples from previously used databases and new examples from the Berkeley Segmentation Database [26] are also included.

The rest of the paper is structured as follows. In Section II, a first set of information theoretical region merging techniques is presented. Concretely, a nonparametric region model based on the empirical distribution is proposed in Section II-A and two different size-dependent merging criteria based on the Kullback-Leibler divergence (Section II-B) and the Bhattacharyya coefficient (Section II-C) are formally developed. Following a similar structure, Section III presents a region model based on a first-order Markov process (Section III-A), leading to similar information theory statistical measures (Sections III-B and III-C). An alternative approach, combining a size-independent extension of all previous methods and a scale-based merging order is presented in Section IV. The automatic partition selection criterion is detailed in Section V. Section VI presents an objective evaluation and comparison with other state-of-the-art region merging and image segmentation techniques, using two different data sets. Conclusions are outlined in Section VII. Finally, a novel proof for an extension of the Chernoff bound for first-order Markov processes is included in the Appendix.

II. AREA-WEIGHTED STATISTICAL MERGING CRITERIA FOR I.I.D. PIXEL REGION MODELS

A. I.I.D. General Statistical Region Model

From a statistical point of view, a single channel image can be considered as a realization of a 2-D stochastic process. Therefore, each pixel is a sample of one of the discrete random variables composing the image process. For simplicity, all mathematical developments in this work are obtained for single channel images; their extension to the multichannel case under channel independence assumption is straightforward.

To formally tackle the image segmentation problem, we consider a region as a set of independent and identically distributed (i.i.d) pixels which is completely characterized by the probability distribution common to all pixels. We propose a region model based on the estimation of this probability distribution from the empirical distribution of the region.

The *empirical distribution* or *type* $P_\mathbf{x}$ of a sequence $\mathbf{x}$ of n samples from an alphabet $\mathcal{X} = \{a_1, a_2, \ldots, a_{|\mathcal{X}|}\}$ is defined as the relative proportion of occurrences of each value of $\mathcal{X}$, i.e., $P_\mathbf{x}(a) = N(a|\mathbf{x})/n$ for all $a \in \mathcal{X}$, where $N(a|\mathbf{x})$ is the number of times the symbol a occurs in the sequence $\mathbf{x} \in \mathcal{X}^n$. Using a main result of the theory of types [40], the probability of the type of a sequence of i.i.d. observations $\mathbf{x}$ with probability distribution Q, is given by

$$Q^n(\mathbf{x}) = 2^{-n(H(P_\mathbf{x}) + D(P_\mathbf{x}\|Q))} \tag{1}$$

where

$$H(P_\mathbf{x}) = -\sum_{a \in \mathcal{X}} P(a) \log P(a) \tag{2}$$

is the Shannon entropy of the type and

$$D(P_\mathbf{x}\|Q) = \sum_{a \in \mathcal{X}} P_\mathbf{x}(a) \log \frac{P_\mathbf{x}(a)}{Q(a)} \tag{3}$$

is the Kullback–Leibler divergence between the statistical distributions. It can be seen [40] that the empirical distribution $P_\mathbf{x}$ converges to Q, concretely, $D(P_\mathbf{x}\|Q) \to 0$ with probability 1 for $n \to \infty$. Hence, for n sufficiently large, the probability for a particular sequence can be approximated by

$$Q^n(\mathbf{x}) \approx 2^{-nH(P_\mathbf{x})} \tag{4}$$

and the unknown distribution of the data Q can be directly approximated by the empirical distribution of the samples. In practice, to ensure that the value of n is large enough to have a reliable approximation, the statistical model is not directly applied at the pixel level but at the level of an initial partition with a reduced number of regions (see Section VI).

Using the empirical distribution provides a unified and general framework for image segmentation, as arbitrary discrete distributions are directly estimated from data. Apart from pixel independence, no further assumptions are made. Moreover, this model can be easily computed and, after the union of a pair of regions, updated

$$P_{1\cup 2}(a) = \frac{n_1}{n_1 + n_2} \cdot P_1(a) + \frac{n_2}{n_1 + n_2} \cdot P_2(a) \tag{5}$$

where n_1, n_2 are the number of pixels in R_1, R_2, respectively.

The quantization of the alphabet $\mathcal{X}$ 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 considered in the empirical distribution. More sophisticated quantization strategies, such as data-dependent partitions [41], are out of the scope of this paper.

B. 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 distributions P_1, P_2, respectively, whose union would generate a new region with empirical distribution $P_{1\cup 2}$. Then, the two hypotheses considered are as follows:

- **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 $P_{1\cup 2}$;
- **H_1**: pixels $\mathbf{x}_1 \in R_1$ are distributed by P_1; and pixels $\mathbf{x}_2 \in R_2$ are distributed by P_2.

In general, we wish to minimize both probabilities of error. The *Neyman–Pearson lemma* [40] proves that the optimal test for two hypotheses, in that sense, is the so-called *likelihood ratio test*:

$$\frac{P_{H_0}(x_1, x_2, \ldots, x_n)}{P_{H_1}(x_1, x_2, \ldots, x_n)} \gtrless T. \tag{6}$$

Using the result in (4) for the probability of each sequence of pixels, we can write the log-likelihood ratio (in base 2) as

$$\log \frac{P_{H_0}(\mathbf{x}_1, \mathbf{x}_2)}{P_{H_1}(\mathbf{x}_1, \mathbf{x}_2)} = -nH(P_{1\cup 2}) + n_1 H(P_1) + n_2 H(P_2) \tag{7}$$

which can be interpreted as the size-weighted decrease in entropy when the regions are merged. Considering (5) and the Kullback–Leibler divergence between statistical distributions, (7) can be rewritten as

$$\log \frac{P_{H_0}(\mathbf{x}_1, \mathbf{x}_2)}{P_{H_1}(\mathbf{x}_1, \mathbf{x}_2)} = -n_1 \cdot D(P_1\|P_{1\cup 2}) - n_2 \cdot D(P_2\|P_{1\cup 2}). \tag{8}$$

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

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -n_i \cdot D(P_i\|P_{i\cup j}) - n_j \cdot D(P_j\|P_{i\cup j}).$$

This criterion is based on measuring the similarity between the empirical distributions of the regions and the empirical distribution of their merging, weighted by the size of the regions.

C. Bhattacharyya Merging Criterion

In this section we present a new criterion based on a direct statistical comparison between the types of the regions, that is, without using an estimate of the probability distribution of the union of the two regions. Nevertheless, in this case, the Kullback–Leibler divergence becomes impractical, as its convergence cannot be assured anymore. For instance, $D(P_1 \|$

$P_2) \to \infty$ if, for some $a \in \mathcal{X}$, $P_1(a) \neq 0$ and $P_2(a) = 0$. A possible solution to this problem may be to use data-dependent partitions for the divergence estimation, as proposed in [41]. Another possibility may be to use the symmetric version of the Kullback–Leibler divergence, known as Jensen–Shannon divergence, which is always bounded. Nevertheless, it is known to behave differently from the Kullback–Leibler divergence [42].

We tackle the problem from a different perspective. Let us consider the probability simplex in $\mathbb{R}^n$, i.e., the $(n-1)$-dimensional manifold defined by all possible empirical distributions for a sequence of n samples. Each region can be seen as a class in this space, 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 [40], defined as

$$C(P_i, P_j) \triangleq - \min_{0 \leq \lambda \leq 1} \log \left(\sum_x P_i^{\lambda}(x) P_j^{1-\lambda}(x) \right). \quad (9)$$

In other words, the performance of a classifier is determined by the pair of closest classes in the probability simplex, in terms of the Kullback–Leibler divergence.

In our case, 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. Hence, the bound on the error probability becomes a bound on the probability of merging. This way, the bound on the probability of merging for two adjacent regions, with type P_i, P_j, and number of pixels n_i, n_j, respectively, can be written as

$$P_{\text{merging}} \leq e^{-\min(n_i, n_j) \cdot C(P_i, P_j)}. \quad (10)$$

Nevertheless, computing the Chernoff information implies an optimization over λ. To reduce this computational load, in practice, we propose to approximate the Chernoff information by the upper bound corresponding to the case $\lambda = 1/2$, known as the *Bhattacharyya coefficient* [43]

$$B(P_i, P_j) \triangleq - \log \left(\sum_x P_i^{1/2}(x) P_j^{1/2}(x) \right). \quad (11)$$

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

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -\min(n_i, n_j) \cdot B(P_i, P_j). \quad (12)$$

This method is based on a size-weighted direct statistical measure of the empirical region distributions, and we will refer to it as the Bhattacharyya merging criterion (BHAT).

III. AREA-WEIGHTED MERGING CRITERIA FOR FIRST-ORDER MARKOV REGION MODELS

A. First-Order Finite-State Markov Process as General Statistical Region Model

Under the same statistical framework, the region merging problem can be formally tackled considering 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 only with respect to neighboring pixels inside the same region.

In order to achieve low complexity region merging, we propose a compromise between the difficulty introduced by a 2-D dependency and the simplicity of the i.i.d assumption in Section II-A, and hence, to model each region using a 1-D first-order Markov model. The reduction of the dimensionality is based on the scanning of the region pixels in four different directions (left–right, right–left, up–down, down–up), estimating the directional empirical transition matrices of the 1-D Markov process associated to each scanning. The simplified 1-D Markov model of the region is obtained averaging the four directional transition matrices. In other words, the 1-D 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 *co-occurrence matrix* [44].

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

$$\mathbf{P_x}(a, b) = \frac{N(x_i = a, \ x_{i+\Delta} = b \mid \mathbf{x})}{N_p} \quad (13)$$

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 total 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 co-occurrence matrix is obtained.

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 (in terms of homogeneity or texture). Moreover, this model can be easily computed and, after the union of a pair of regions, updated

$$\mathbf{P}_{1 \cup 2} = \frac{n_1}{n_1 + n_2} \mathbf{P}_1 + \frac{n_2}{n_1 + n_2} \mathbf{P}_2 \quad (14)$$

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

The quantization of the alphabet $\mathcal{X}$ can be set to optimize the performance of the algorithm. As in Section II-A, we only consider a uniform quantization and directly refer to the number of bins in each dimension of the co-occurrence matrix.

The statistical formulation of the merging problem presented in Sections IV–VI is based on considering the co-occurrence matrix as the empirical probability transition matrix of the first-order finite-state Markov process characterizing a region [45]. We will assume that this Markov process is ergodic, and hence, it is completely characterized by its initial state and a probability transition matrix. However, note that in this particular case it is not meaningful to consider an initial state distribution for the sequence of pixels, because the pixels of a 2-D 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 $\pi_i = (1/|\mathcal{X}|)$. Under these considerations, a region is completely characterized by the probability transition matrix of the first-order Markov process generating it, estimated by its co-occurrence matrix.

B. Kullback-Leibler Merging Criterion

Similarly to the i.i.d. case in Section II-B, we can define a criterion based on merging at each step the pair of adjacent regions maximizing the probability of being generated by the same first-order Markov process. Assuming that R_1 and R_2 are two adjacent regions with n_1, n_2 pixels, and with empirical transition matrices $\mathbf{P}_1$, $\mathbf{P}_2$, respectively, whose union would generate a new region with empirical transition matrix $\mathbf{P}_{1\cup2}$, the two hypotheses considered are as follows:

- $\mathbf{H_0}$: 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 $\mathbf{P}_{1\cup2}$;
- $\mathbf{H_1}$: 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 [46] 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*:

$$\frac{P_{H_0}(x_1, x_2, \ldots, x_n)}{P_{H_1}(x_1, x_2, \ldots, x_n)} \gtrless T. \quad (15)$$

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

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} = \log\frac{P_{1\cup2}(x_1)}{P_1(x_1)P_2(x_1)} + \sum_{i=2}^{n}\log P_{1\cup2}(x_{i-1}|x_i)$$
$$- \sum_{i=2}^{n_1}\log P_1(x_{i-1}|x_i) - \sum_{i=n_1+2}^{n}\log P_2(x_{i-1}|x_i)$$

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 is assumed to be ergodic (as stated in Section III-A), 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,

$$\sum_{i=2}^{n}\log P_{1\cup2}(x_{i-1}|x_i) = (n-1)\cdot E\left\{\log P_{1\cup2}(x_{i-1}|x_i)\right\}$$

where $E\{\cdot\}$ corresponds to the statistical mean under the distribution of $P_{1\cup2}$. Considering $P_{1\cup2}(x_1 = a_i) = \pi_i^{1\cup2}$ as the initial states distribution of the process, and $\mathbf{P}_{1\cup2} = (p_{ij}^{1\cup2}) = P_{1\cup2}(x_{i-1}|x_i)$ as the transition matrix

$$E\{\log P_{1\cup2}(x_{i-1}|x_i)\} = \sum_{i\in\mathcal{X}}\pi_i^{1\cup2}\sum_{j\in\mathcal{X}}p_{ij}^{1\cup2}\log p_{ij}^{1\cup2} = -H(P_{1\cup2})$$

$$(16)$$

where

$$H(P) = -\sum_{i,j\in\mathcal{X}}\pi_i p_{ij}\log p_{ij} \quad (17)$$

is the Shannon entropy rate of the first-order Markov process [47].

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

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} = \log\frac{P_{1\cup2}(x_1)}{P_1(x_1)P_2(x_1)} - (n-1)\cdot H(P_{1\cup2})$$
$$+ (n_1-1)\cdot H(P_1) + (n_2-1)\cdot H(P_2). \quad (18)$$

In general, for n_1 and n_2 sufficiently large, the first term can be dismissed [46]. Particularly in our case, the asymptotic conditions for n_1 and n_2 are not required as we consider an equiprobable initial state distribution. Thus, the first term is constant for any $\mathbf{x}$, not affecting the maximization of the log-likelihood ratio test. Under this condition, the log-likelihood ratio can be written (apart from a constant) as

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} = -(n-1)\cdot H(P_{1\cup2})$$
$$+ (n_1-1)\cdot H(P_1) + (n_2-1)\cdot H(P_2) \quad (19)$$

and, for $n_1, n_2 \gg 1$, it can be simply approximated as

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx -n\cdot H(P_{1\cup2}) + n_1\cdot H(P_1) + n_2\cdot H(P_2) \quad (20)$$

that can be interpreted as the size-weighted decrement on the entropy rate when the regions are merged. Considering (14), 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 $\mathbf{Q} = (q_{ij})$:

$$D'(P\|Q) = \lim_{n\to\infty}\frac{1}{n}D(p^n\|q^n) = \sum_{i,j\in\mathcal{X}}\pi_i\cdot p_{ij}\cdot\log\frac{p_{ij}}{q_{ij}} \quad (21)$$

then, (20) can be rewritten as:

$$\log\frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx -n_1 D'(P_1\|P_{1\cup2}) - n_2 D'(P_2\|P_{1\cup2}) \quad (22)$$

or equivalently,

$$\log \frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \approx \frac{1}{|\mathcal{X}|} \left(n_1 \sum_{i,j} p_{ij}^1 \log \frac{p_{ij}^{1\cup2}}{p_{ij}^1} \right.$$

$$\left. + n_2 \sum_{i,j} p_{ij}^2 \log \frac{p_{ij}^{1\cup2}}{p_{ij}^2} \right).$$

Defining the Kullback–Leibler divergence between transition matrices as

$$D(\mathbf{P}\|\mathbf{Q}) \triangleq \sum_{i,j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \qquad (23)$$

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

$$\log \frac{P_{H_0}(\mathbf{x})}{P_{H_1}(\mathbf{x})} \propto -n_1 D(\mathbf{P}_1\|\mathbf{P}_{1\cup2}) - n_2 D(\mathbf{P}_2\|\mathbf{P}_{1\cup2}). \quad (24)$$

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

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

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.

C. Bhattacharyya Merging Criterion

Identically to Section II-C, the idea behind this approach is to use the Chernoff information bound on the exponent of the probability of error of a classifier, based on the maximum intersection between two distributions, as a measure of similarity, and consequently, to perform a clustering procedure based on the maximization of this bound.

Proceeding analogously to the classical derivation of the Chernoff bound for the i.i.d. case, in the Appendix we develop an extension of this bound for the case of first-order Markov sequences. Hence, the Chernoff information between the transition matrices of two first-order Markov processes is defined as

$$C(\mathbf{P}_1, \mathbf{P}_2) \triangleq -\min_{0 \le \lambda \le 1} \log \left(\sum_{i,j} \left(p_{ij}^1 \right)^\lambda \left(p_{ij}^2 \right)^{1-\lambda} \right). \quad (25)$$

Following the reasoning in Section II-C, 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

$$P_{\text{merging}}(R_i, R_j) \le e^{-\min(n_i, n_j) \cdot C(\mathbf{P}_i, \mathbf{P}_j)}. \quad (26)$$

As in Section II-C, in order to avoid the optimization over λ, we propose to approximate the Chernoff information by the *Bhattacharyya coefficient* between the transition matrices ($\lambda = 1/2$)

$$B(\mathbf{P}_i, \mathbf{P}_j) \triangleq -\log \left(\sum_x \mathbf{P}_i^{1/2}(x) \mathbf{P}_j^{1/2}(x) \right). \quad (27)$$

In conclusion, this approach leads to an analogous expression to the criteria in (12) for the i.i.d. case:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -\min(n_i, n_j) \cdot B(\mathbf{P}_i, \mathbf{P}_j). \quad (28)$$

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).

IV. EXTENSION TO AREA-UNWEIGHTED STATISTICAL MERGING CRITERIA

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 areas of the regions tend to increase as the number of regions into the partition decreases.

The size term favors the fusion of smaller regions, slowing the merging of larger regions, even when they are similarly distributed. On one hand, as it may be possible to make a mistake during the merging process, merging small regions causes less significant errors in terms of number of pixels, minimizing undersegmentation. On the other hand, as the fusion of larger regions is slowed even when they are similarly distributed, area-weighted methods suffer generally from oversegmentation (see Section VI).

Therefore, the goal of this section is to propose an extension of the previous methods providing a tradeoff between under- and oversegmentation, while increasing the size resolution of the partitions (i.e., the region with minimum size). 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 distribution (see Section VI for further details on how to assure this condition in practice), the area dependency can be removed from the previous merging criteria:

- area-unweighted Kullback–Leibler merging criterion for the i.i.d. region model:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -D(P_i\|P_{i\cup j}) - D(P_j\|P_{i\cup j}); \quad (29)$$

- Area-unweighted Bhattacharyya merging criterion for the i.i.d. region model:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -B(P_i, P_j); \quad (30)$$

- area-unweighted Kullback–Leibler merging criterion for the first-order Markov region model:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -D(\mathbf{P}_i \| \mathbf{P}_{i \cup j}) - D(\mathbf{P}_j \| \mathbf{P}_{i \cup j}); \quad (31)$$

- area-unweighted Bhattacharyya merging criterion for the first-order Markov region model:

$$\{R_1, R_2\} = \arg \max_{R_i \sim R_j} -B(\mathbf{P}_i, \mathbf{P}_j). \quad (32)$$

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 (29)–(32) with a scale-based merging order, incorporating the size consistency constraints. The idea is to define a *scale threshold* for each level of resolution. Regions beyond this threshold are considered as *out-of-scale* and are merged with the highest priority, fusing them with their 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 until new out-of-scale regions appear.

The scale threshold is defined as

$$T_{\text{scale}} = \alpha \cdot \frac{\text{Image Area}}{\text{Number of Regions}} \quad (33)$$

i.e., regions that are smaller than a given percentage of the mean region area at the current scale are considered out-of-scale. 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 performance curves for different values of the scale parameter α are evaluated in Section VI (see Experiment 4 and Fig. 11).

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 their size, because the merging cost only measures the statistical similarity of the empirical distributions, without being size biased.

V. PARTITION SELECTION CRITERION

A. Stopping Criterion Versus Selection Criterion

In Section I, the importance of a hierarchical approach to provide different region-based explanations of the image at different levels of detail was discussed. For instance, the level of resolution may be application dependent, considering the type of analysis to be performed. The nature of the region merging techniques succeeds to provide a hierarchy of partitions. Nevertheless, these processes usually lead to a huge number of partitions with no clues about which of them is more representative or contains the most meaningful information.

The goal of this section is to propose a partition selection criterion, i.e., an automatic technique to extract from the hierarchy of partitions the most statistical significant partitions at different resolutions. Although being at an early level of analysis, and

without any *a priori* knowledge on the image, we expect statistical meaningful partitions to contain most human-representative regions, for different levels of analysis.

In the region merging literature, other researchers have proposed the use of stopping criteria for the merging process. These approaches determine a single step in the merging process (that is, a single partition). This proposed partition is usually related with the last merging step where the merging criteria has a certain degree of validity or where a certain compromise is fulfilled. Some simple criteria are based on the maximum merging cost (or minimum similarity), on the definition of a certain threshold on the number of regions [11], [48], or on the minimum value of peak-to-noise ratio (PSNR) between the original image and an image partition reconstructed using the mean region color [11]. When heterogenous image data sets are considered, it is not possible to generalize the value of these parameters for a given purpose (for instance: extraction of partitions with high level semantic content, or with minimum undersegmentation error and providing high accuracy).

In [49], a stopping criterion based on the evolution of the accumulated merging cost during the overall process is proposed. Its goal is to extract the most salient semantic regions or objects present in the scene. It determines the number of regions that divides the accumulated merging curve into two segments: the highest variation part of the accumulated cost and the lowest variation part (geometrically, it corresponds with the first-order polygonal approximation of the accumulated curve using a standard recursive subdivision method). Although this criterion is dependent on the image content itself, the determined partition depends on the number of regions of the initial partition used by the region merging process.

Another stopping criterion based on the accumulated cost is proposed in [34]. In this case, its purpose is to determine partitions with correct characterization of the objects in terms of low undersegmentation error and a moderate number of regions. The proposed partition is given by the step of the merging process leading to a relative increase of the accumulated cost over a certain percentage. This method succeeds in its particular goal, but no studies are presented about its validity in other cases.

We would like to remark the significant difference between a stopping criterion (as those proposed in the previous approaches) and the partition selection proposed here. While a stopping criterion determines a single partition or iteration, usually referred to the merging step where the used merging criteria cannot assure a certain degree of reliability or where a given tradeoff is fulfilled, the selection criteria proposes a reduced set of meaningful partitions at different levels of detail. Thus, the selection criteria is application independent, and it is not specified together with a particular purpose. It can be seen as a tool to simplify and facilitate a higher level image analysis, but independently of the type of analysis itself. In that sense, the proposed partition selection criterion is related with the simplification or edition techniques for hierarchical region-based image representations in [50] and [51]. In those approaches, the merging sequence is stored into a hierarchical tree structure that, in turn, is collapsed to a reduced subset of relevant partitions. Nevertheless, the simplified version of the hierarchy is based on the analysis of the tree structure

Fig. 3. Analysis of the partition selection criterion: Berkeley Segmentation Database [26] image subset. Columns, from left to right: original image, first and second selected partitions using the significance index in (34). Partitions are computed using Markov based region model criteria. Concretely, the method used for the examples are: M-BHAT area-weighted in the first row, and M-KL area-weighted in the second row, both with empirical transition matrices quantized to 10 bins per dimension.

Fig. 4. Analysis of the partition selection criterion: Corel image subset. Columns, from left to right: Original image, first and second selected partitions using the significance index in (34). The left columns show i.i.d. KL area-unweighted criterion (types quantized to five bins); and the right columns show Markov M-BHAT area-unweighted criterion (empirical transition matrices quantized to five bins per dimension).

(parent-children relationships) and not directly on the sequence of merging similarities as the criterion presented here.

B. Statistical Partition Selection Criterion

Experimentally, we have observed that the merging-similarity sequences for area-weighted and area-unweighted methods (in this case, without considering the costs of out-of-scale regions) present a similar behavior. For that reason, the proposed selection strategy can be applied to all merging techniques presented in this work.

The proposed strategy relies on selecting the partitions associated to a significant decrease into the sequence of merging similarities. Therefore, we consider a nondecreasing version of the sequence of merging similarities $s(n)$, where n is the number of remaining regions, defined as $\hat{s}_{\min}(n) = \min_{n \leq k < \infty} \{s(k)\}$.

Determining the most important decrements on $\hat{s}_{\min}(n)$ provides the set of statistically significant partitions. These partitions may be ordered using some significance index. Here, we propose an importance weight based on the relative increase with respect to the current similarity value. Given $\delta(n) = (d/dn)(\hat{s}_{\min}(n))$ and a nonincreasing version of $\delta(n)$, $\hat{\delta}_{\max}(n) = \max_{n \leq k < \infty} \{\delta(k)\}$, the importance weight is defined as

$$\omega_k = \frac{\hat{\delta}_{\max}(k-1) - \hat{\delta}_{\max}(k)}{\hat{\delta}_{\max}(k)}. \tag{34}$$

In order to illustrate the behavior of the proposed partition selection criterion and the fact that it performs similarly with all proposed merging criteria, examples of automatically selected partitions using this significance order are shown in Figs. 3–5. A complete evaluation is given in Section VI. Note that in Figs. 3 and 4, in general, the first selection corresponds to a coarse partition, whose regions are good approximations of the objects. Usually, the second proposal gives a finer partition with most representative regions in the scene. Examples in Fig. 5 (natural texture mosaics from the Prague texture segmentation data generator, see Section VI-B) show that the partition selection criteria can be a valuable solution to provide with correct image explanations at different levels of detail, correctly capturing the increase in complexity of the region pattern as the resolution decreases.

VI. EXPERIMENTAL RESULTS

In this section, we provide a complete and exhaustive evaluation of the proposed techniques, using not only different databases for the two main addressed problems: object-oriented segmentation of generic images (Corel Image Database, Section VI-A) and texture image segmentation (Prague Segmentation Datagenerator, Section VI-B), but also specific evaluation features in each case: under- and oversegmentation error, and a large set of region-based, pixel-based and error consistency indicators, respectively.
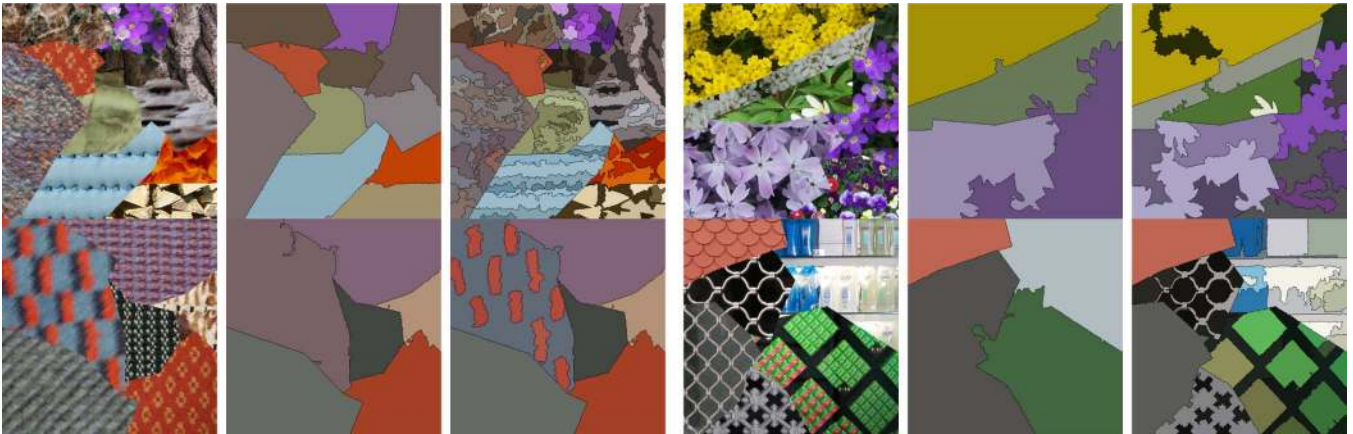
Fig. 5. Analysis of the partition selection criterion: texture mosaic. Columns, from left to right: original image, first and second selected partitions using the significance index in (34). The left columns show i.i.d. region model criteria; concretely, in descending order: KL area-weighted, BHAT area-weighted (types quantized to ten bins). The right columns show Markov region model criteria; concretely, in descending order: M-KL area-weighted, M-BHAT area-weighted (empirical transition matrices quantized to ten bins per dimension).

In the results shown in the paper, we have tried to select images presenting objects (single or not) showing different characteristics. For instance, objects immerse into textured background, objects formed by textured regions or objects formed not only by color homogeneous regions but also by different plain color regions. Additionally, we also illustrate the behavior in more general images, such as landscapes. With this selection of illustrative examples, including also a set of natural texture mosaics, we try to cover a large amount of general image types with different levels of complexity.

In order to ensure that all initial regions are large enough to have a high confidence on the statistical model estimation, the proposed region merging techniques were applied on an initial partition of the original image. Except otherwise stated, the initial partitions were computed using the watershed algorithm [52].

A. Corel Image Database

The first set of experiments was performed over a set of 100 images from the Corel image database [34]. The set contains ten images of ten different complexity classes: tigers, horses, eagles, mountains, fields, cars, jets, beaches, butterflies and roses. 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 as in [34] the distances defined in [33]. Initially, an asymmetric distance is proposed $d_{\mathrm{asym}}(P, Q)$. By definition, it measures the minimum number of pixels whose labels should be changed so that partition P becomes finer than partition Q, normalized by the image size. Note that, in general, $d_{\mathrm{asym}}(P, Q) \neq d_{\mathrm{asym}}(Q, P)$. When P is the partition to evaluate and Q the ground truth partition, the first ordering measures the degree of undersegmentation, and the second, the oversegmentation in P with respect to the ground truth partition.

Nevertheless, some applications may not be uniquely concerned about under- or oversegmentation but they may be interested in establishing a good compromise between both types of error. For that purpose, a symmetric distance $d_{\mathrm{sym}}(P, Q)$ is proposed that provides a measure of the global error between partitions with equal number of regions. 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 was originally defined in the pattern analysis field, in terms of data clustering comparison [53].

1) Experiment 1. Under- and Oversegmentation Evaluation: This experiment evaluates separately, as a function of the number of regions into the partitions: i) the degree of undersegmentation and ii) the degree of oversegmentation of the generated partitions, referred in both cases to the ground truth partitions. The asymmetric partition distance is used for this evaluation, concretely, $d_{\mathrm{asym}}(P, Q)$ is chosen for i), and $d_{\mathrm{asym}}(Q, P)$ for ii), where P and Q are the computed partition and the ground truth partition, respectively.

These results are compared with the region merging technique proposed in [35]. The merging criterion in [35] 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. In order to fix the same test conditions as in [35], the proposed statistical techniques were applied on initial partitions (with 500 regions) computed using the same color-based criterion used in [35] to initialize the merging process (the weighted Euclidean norm of the color mean difference of the regions). The initial partitions were computed using the same color-based criteria as in [35]. The residual undersegmentation mean error associated to the initial partitions (their average asymmetric distance with respect to the ground truth) is 0.0211, with 500 regions per partition.

In Fig. 6, the results for the mean asymmetric distance for different number of regions with i.i.d. based techniques are presented. Fig. 6-left shows $d_{\mathrm{asym}}(P, Q)$, measuring the degree of undersegmentation. In this case, area-weighed methods outperform area-unweighted methods. On the contrary, in Fig. 6-right, for $d_{\mathrm{asym}}(Q, P)$, area-unweighted methods clearly present less oversegmentation. As discussed in Section IV, this fact is expected due to the increase in the partition resolution provided by area-unweighted methods. Note that, in general and for both
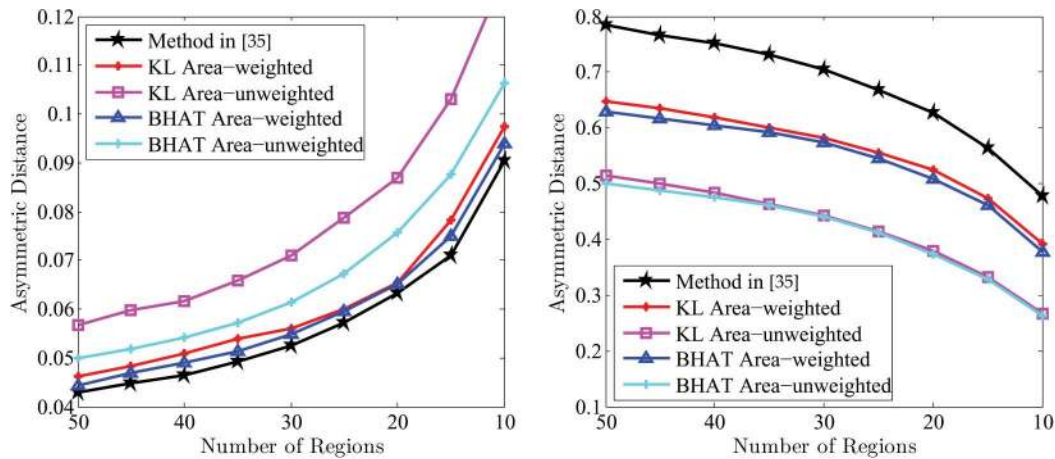
Fig. 6. Asymmetric distance for the subset of the Corel database. Left: From computed to ground truth partition (degree of undersegmentation); right: vice versa (degree of oversegmentation). Note that left and right figures are shown at different vertical scales. Statistical methods were computed using types quantized to five bins.
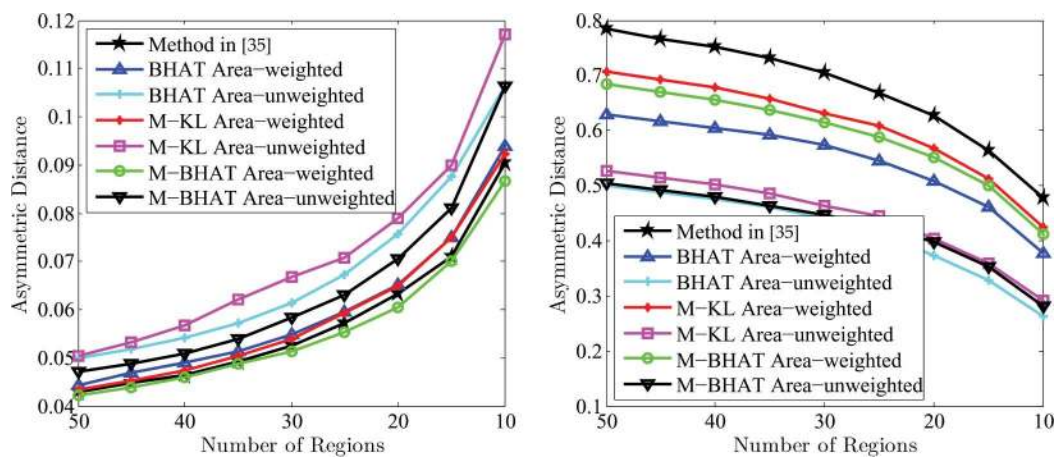

Fig. 7. Asymmetric distance for the Corel subset database. Left: From computed to ground truth partition (undersegmentation); right: vice versa (oversegmentation). Left and right figures are shown at different vertical scales. Statistical methods were computed using matrices quantized to five bins per dimension.

results, the performance of the Bhattacharyya criterion is superior to that of the Kullback–Leibler criterion.

Compared to [35], i.i.d. based techniques suffer from larger undersegmentation error, although the mean error increase for the Bhattacharyya and the Kullback–Leibler area-weighted versions is less than 4% and 6%, respectively. Nevertheless, in terms of oversegmentation, all statistical techniques significantly outperform the method in [35]: a mean 15% decrease of the asymmetric distance for area-weighed methods, and a mean 30% decrease for area-unweighted methods (see Fig. 6-right).

In Fig. 7, the results for the mean asymmetric distance for different numbers of regions for the Markov based techniques are presented. For the sake of clarity, only the best i.i.d. model based area-weighted and area-unweighed methods are shown (corresponding in both cases to the Bhattacharyya merging criteria). Fig. 7-left shows $d_{\mathrm{asym}}(P,Q)$, measuring the degree of undersegmentation. In this case, area-weighted methods outperform area-unweighted methods, and generally, Markov-based techniques are superior to i.i.d. methods. The Markov-based Bhattacharyya area-weighted method presents the most significant improvement with respect to the best technique based on the i.i.d. region model (a 5% distance decrease), outperforming

also [35] with a 1.5% less undersegmentation error. On the contrary, in Fig. 7-right, for $d_{\mathrm{asym}}(Q,P)$, Markov-based techniques suffer from more oversegmentation than i.i.d. model based techniques. Despite this fact, they still outperform [35] in terms of oversegmentation (a mean 10% distance decrease). Hence, from our experiments it can be observed that there is a compromise between under- and oversegmentation. Again, the Bhattacharyya versions show a better performance than the Kullback–Leibler techniques.

2) Experiment 2. Global Partition Quality Evaluation: The goal of this experiment is to perform a general comparison in terms of under- and oversegmentation between the ground truth partitions and the partitions generated by the proposed method with the same number of regions than the corresponding ground truth partitions. To measure the dissimilarity between human-created and computed partitions, the symmetric partition distance is used.

Table I shows the mean symmetric distance between ground truth partitions and partitions with the same number of regions generated by the proposed methods. As in the previous experiment, the initial partitions were computed using the same color-based criteria as in [35].
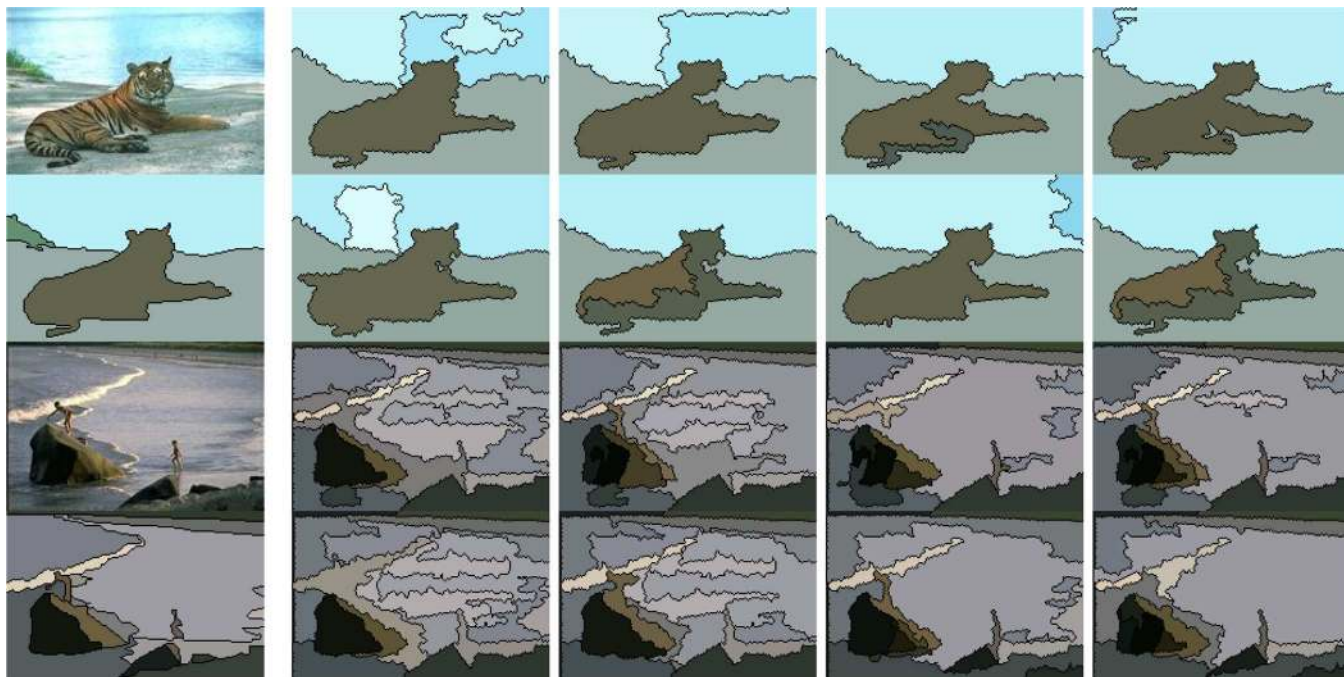
Fig. 8. Merging criteria comparison: Corel image subset. For each example, columns from left to right: Original image (first row), human partition (second row), KL area-weighted (first row), M-KL area-weighted (second row); BHAT area-weighted (first row), M-BHAT area-weighted (second row); KL area-unweighted (first row), M-KL area-unweighted (second row); BHAT area-unweighted (first row), M-BHAT area-unweighted (second row). Types and empirical transition matrices quantized to five bins per dimension.

TABLE I
MEAN SYMMETRIC DISTANCE FOR THE SUBSET OF THE COREL DATABASE.
THE TWO CRITERIA WITH MINIMUM MEAN SYMMETRIC
DISTANCE ARE HIGHLIGHTED IN GRAY

	Symmetric Distance
1. Method in [35]	0.3369
2. KL Area-weighted	0.3138
3. KL Area-unweighted	0.2480
4. BHAT Area-weighted	0.3044
5. BHAT Area-unweighted	0.2292
6. Markov KL Area-weighted	0.3277
7. Markov KL Area-unweighted	0.2532
8. Markov BHAT Area-weighted	0.3117
9. Markov BHAT Area-unweighted	0.2347

Note that all statistical criteria, i.i.d. or Markov based, outperform [35]. As expected, area-unweighted methods present the best tradeoff between under- and oversegmentation, and particularly, the Bhattacharyya-based criteria. The symmetric distance is slightly larger for Markov-based methods compared to i.i.d. model based methods. Fig. 8 presents several results. In it, it can be observed that, in spite of the differences observed in Table I, the subjective quality of the partitions is similar for all statistical approaches.

3) Experiment 3. Quantization Effect Evaluation: Here, we study the effect of the degree of quantization, in terms of number of bins, in the empirical models used in the proposed statistical methods. Similarly to Experiment 1, we measure independently the degree of under- and oversegmentation as a function of the number of regions into the partition for different number of bins into the model.

Figs. 9 and 10 present examples of the performance variation for different number of bins on the normalized histogram for each region. Concretely, the $d_{\mathrm{asym}}(P, Q)$ for the area-weighted version of the Bhattacharyya method for the i.i.d. model and the Markov model are shown, respectively.

On one hand, it can be seen than the oversegmentation error monotonically decreases as number of bins decreases and presents a large range of variation (see Fig. 9-right). On the other hand, the variation on the degree of undersegmentation when the number of bins is not excessively small (more than four bins) is not so important and only a moderate improvement can be achieved. Note that in this case, the increase of the number not always improves the performance (for instance, in Fig. 9-left the minimum curve of the asymmetric distance is obtained for ten bins). This effect is being currently analyzed as part of our future work (see discussion in Section VII). Taking into account not only the segmentation quality but also the computational time, we can conclude that good performance can be obtained with a reduced number of bins: five or ten bins is a good compromise between the partition quality and the computational load of the algorithm. Remaining methods present a similar behavior.

4) Experiment 4. Scale-Parameter Effect Evaluation: For completeness, the performance curves for different values of the scale parameter α, introduced in Section IV to control the scale-based merging order for area-unweighted merging criteria, are evaluated. Concretely, in Fig. 11, the asymmetric distance measuring the degree of under- (left) and oversegmentation (right) for the i.i.d. version of the Bhattacharyya area-unweighted method are shown to illustrate the general behavior of the presented methods.
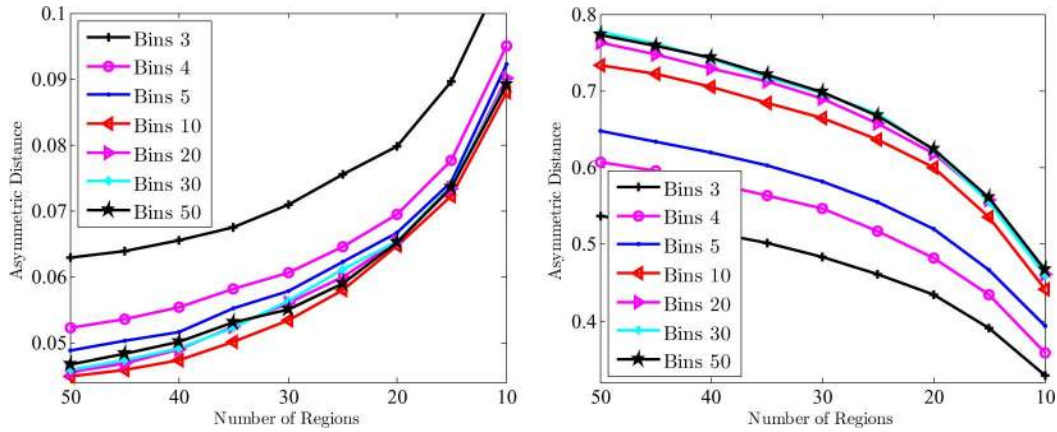
Fig. 9. Asymmetric distance for the BHAT area-weighted method for different number of bins in the empirical distributions computed over the subset of the Corel database. Left: From computed to ground truth partition; right: vice versa. Left and right figures are shown at different vertical scales.
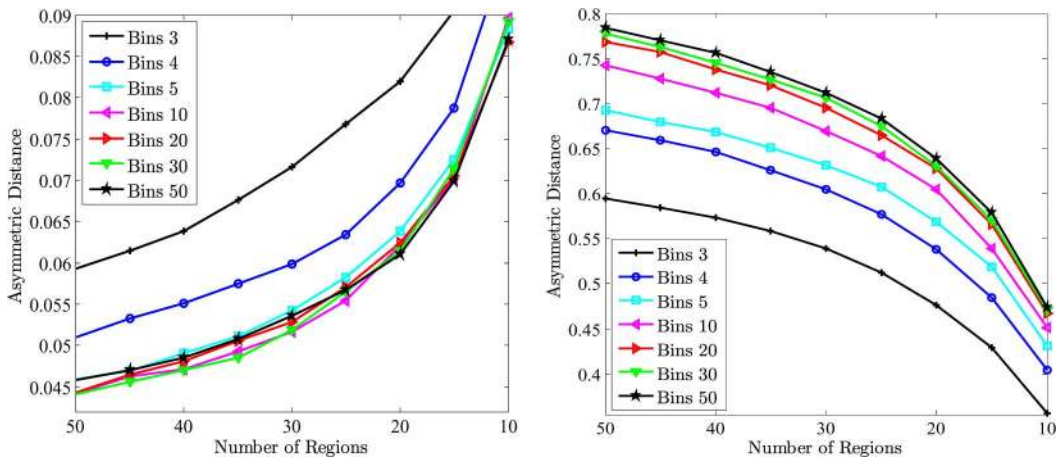

Fig. 10. Asymmetric distance for the M-BHAT area-weighted method for different number of bins in the empirical distributions for the Corel subset database. Left: From computed to ground truth partition; right: vice versa. Left and right figures are shown at different vertical scales.

Fig. 11. Asymmetric distance for the BHAT area-unweighted method for different values of the scale parameter α in the scale-based merging order for the Corel subset database. Left: From computed to ground truth partition; right: vice versa. Left and right figures are shown at different vertical scales.
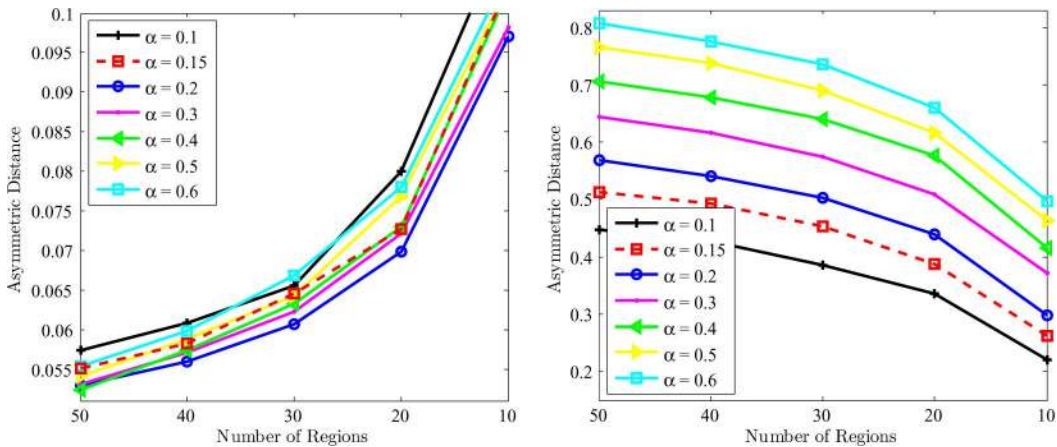
The conclusions are similar to those obtained in the previous experiment referred to the number of bins. The oversegmentation error monotonically decreases with the value of α. Nevertheless, there is not such a direct relation between the scale parameter and the degree of undersegmentation, presenting a minimum for values around 0.2. Justified by the observed compromise between under- and oversegmentation and by the subjective quality of the provided partitions, as

commented in Section IV, a value of $\alpha = 0.15$ was chosen in our experiments.

5) Experiment 5. Partition Selection Criterion Evaluation: In this experiment, the performance of the partition selection criterion proposed in Section V is evaluated for a specific purpose. Concretely, we study the ability of the most significant partition proposed by the significance index in (34) to correctly represent most of the semantic content of the image. Considering the

TABLE II

EVALUATION OF THE ABILITY OF THE MOST SIGNIFICANT PARTITION P_{N_1}, WHERE N_1 IS ITS NUMBER OF REGIONS, TO REPRESENT THE SEMANTIC CONTENT OF THE IMAGE FOR THE COREL SUBSET. FOR EACH METHOD, THE RESULTS IN EACH COLUMN (FROM LEFT TO RIGHT) SHOW: MEAN VALUE FOR THE ASYMMETRIC DISTANCE FROM THE PARTITION WITH ONE REGION MORE THAN THE MOST SIGNIFICANT PARTITION, P_{N_1+1}, TO THE HUMAN-CREATED PARTITION $Q (d_{\mathrm{asym}}(P_{N_1+1}, Q))$; MEAN VALUE FOR THE ASYMMETRIC DISTANCE FROM THE MOST SIGNIFICANT PARTITION TO THE HUMAN PARTITION $(d_{\mathrm{asym}}(P_{N_1}, Q))$; MEAN VALUE FOR THE ASYMMETRIC DISTANCE FROM THE PARTITION WITH ONE REGION LESS THAN THE MOST SIGNIFICANT PARTITION, P_{N_1-1}, TO THE HUMAN PARTITION $(d_{\mathrm{asym}}(P_{N_1-1}, Q))$; PERCENTAGE OF RELATIVE INCREASE IN ASYMMETRIC DISTANCE BETWEEN P_{N_1+1} AND $P_{N_1}(\Delta_{N_1,N_1+1}(\%))$; PERCENTAGE OF RELATIVE INCREASE IN ASYMMETRIC DISTANCE BETWEEN P_{N_1} AND $P_{N_1-1}(\Delta_{N_1-1,N_1}(\%))$

Method	$d_{\mathrm{asym}}(P_{N_1+1}, Q)$	$d_{\mathrm{asym}}(P_{N_1}, Q)$	$d_{\mathrm{asym}}(P_{N_1-1}, Q)$	Δ_{N_1,N_1+1} (%)	Δ_{N_1-1,N_1} (%)
1. KL area-weighted	0.1391	0.1570	0.2400	12.91	52.88
2. KL area-unweighted	0.1315	0.1649	0.2002	25.39	21.41
3. BHAT area-weighted.	0.1407	0.1659	0.2628	17.86	58.45
4. BHAT area-unweighted	0.1409	0.1614	0.2111	14.57	30.76
5. M-KL area-weighted	0.1523	0.1714	0.2692	12.58	57.08
6. M-KL area-unweighted	0.1226	0.1392	0.1736	13.54	24.67
7. M-BHAT area-weighted.	0.1433	0.1702	0.2831	18.76	66.28
8. M-BHAT area-unweighted	0.1231	0.1383	0.1844	12.34	33.38

human-created partitions as semantic reference, the asymmetric partition distance between these partitions and the first partition in significance proposed by the selection criterion is computed (see column $d_{\mathrm{asym}}(P_{N_1}, Q)$ in Table II).

Additionally, we investigate the correlation between the proposed partition and a significant error in terms of semantic content into the next merging step. In other words, we evaluate the asymmetric distance increase due to the next region merging. Thus, a large increase in asymmetric distance will be interpreted as a significant semantic error of the partition containing one region less than the selected one. These results for the Corel image subset are outlined in Table II. For a correct evaluation, the relative distance increase caused by the creation of the selected partition is also shown (i.e., the asymmetric distance between ground truth partitions and partitions containing one region more than the selected one).

As it can be seen in Table II, the distance increase is very large for area-weighted methods (from 53% to 66%), specially when is compared with the increase in distance caused by the generation of the selected partition (from 13% to 19%). For the area-unweighted techniques the increase is still large but lower (a 25%–33% increase with respect to a previous 12%–15%), except for the i.i.d. version of the Kullback–Leibler criterion. For this last criterion the correlation between the selected partition and an imminent semantic error is not observed (a 21% increase with respect to a previous 25%). This fact seems to be associated with the low performance of this criterion in terms of undersegmentation error, compared to the other statistical methods, that was observed in Experiment 1 (see Fig. 6).

Moreover, some additional examples comparing the unsupervised results provided by the partition selection criterion (concretely, the most significant partition) to other unsupervised and supervised statistical techniques are presented in Fig. 12. The images and the human segmentations have been extracted from the Berkeley Segmentation Database [26]. The first method to compare (third column of Fig. 12) is a statistical segmentation technique based on the mean-shift procedure [22]. The results

for this method were computed using the software provided by the authors (http://www.caip.rutgers.edu/riul/research/code.html). The other two compared techniques are the statistical region merging algorithm in [27] (discussed in Section I) and a supervised version of this algorithm [31], respectively, the fourth and fifth columns of Fig. 12. The results for this two methods have been obtained from the authors web page (http://www.univ-ag.fr/~rnock/Articles/TPAMI03/). Particularly, the supervised algorithm is based on grouping with bias [54], i.e., the user points in the image some pixels that he/she thinks belong to identical/different objects and the segmentation is solved as a constrained grouping problem. The user markers are also shown in Fig. 12, different marker shapes and colors meaning different objects. It can be seen that the segmentations provided by the proposed algorithms working in an unsupervised manner is similar to the human partition, being comparable or superior to the other supervised and unsupervised methods (for instance, see the image in the fourth row of Fig. 12). Note that our approach does not require any parameter adjustment (for area-unweighted methods, the scale parameter, $\alpha = 0.15$, is never varied) as the other compared techniques do: the number of variables in the region model, Q, in [27] and [31]; and the bandwidth parameter for the mean-shift procedure, (h_s, h_r) and the minimum region size, in [22].

B. The Prague Texture Segmentation Datagenerator and Benchmark

A different evaluation is performed using now the benchmark system presented in [55]. It contains a set of 20 computer generated texture mosaics and benchmarks composed from the following real natural texture types: monospectral textures, multispectral textures, BTF (bidirectional texture function) textures, rotation invariant textures, scale invariant textures, and illumination invariant textures. The system allows 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 [55]. A complete description of these methods, as well as the results for
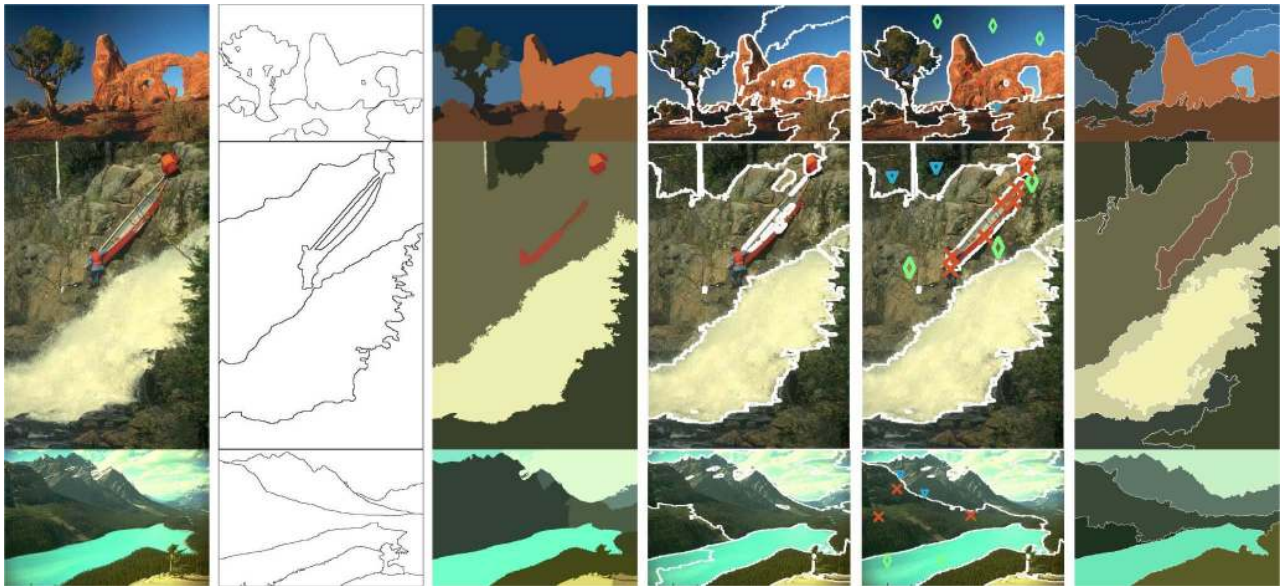
Fig. 12. Comparison between different supervised and unsupervised statistical segmentation techniques. First column: Original images extracted from the Berkeley Segmentation Dataset [26]. Second column: Human segmentations, also from the Berkeley Segmentation Database. Third column: Partitions computed using [22] with bandwidth parameters $(h_s, h_r) = (32, 16)$ and minimum region size of 100 pixels. Fourth column: Partitions obtained by [27] (from the web page of the authors). Fifth column: Partitions obtained by [31] (from the web page of the authors). Sixth column: Most relevant partitions using the significance index in (34) computed by M-BHAT (first row) and M-KL (second and third rows) merging criteria, both quantized to ten bins.

TABLE III
STATISTICAL REGION MERGING SUPERVISED EVALUATION ON THE PRAGUE TEXTURE SEGMENTATION BENCHMARK. RESULTS DIRECTLY OBTAINED FROM THE SYSTEM IN [55]. 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: CORRECT SEGMENTATION (CS), OVER-SEGMENTATION (OS), UNDER-SEGMENTATION (US), MISSED ERROR (ME), NOISE ERROR (NE), OMISSION ERROR (O), COMMISSION ERROR (C), CLASS ACCURACY (CA), RECALL—CORRECT ASSIGNMENT (CO), PRECISION—OBJECT ACCURACY (CC), TYPE I ERROR (I.), TYPE II ERROR (II.), MEAN CLASS ACCURACY ESTIMATE (EA), MAPPING SCORE (MS), ROOT MEAN SQUARE PROPORTION ESTIMATION ERROR (RM), COMPARISON INDEX (CI), GLOBAL CONSISTENCY ERROR (GCE), LOCAL CONSISTENCY ERROR (LCE), MIRKIN METRIC (dM), VAN DONGEN METRIC (dD), VARIATION OF INFORMATION (dVI). SEGMENTATION TECHNIQUES ARE IN DESCENDING ORDER ACCORDING TO THE CORRECT SEGMENTATION INDICATOR (CS)

	KL Area (supervised)	M-KL Area (supervised)	M-BHAT Area (supervised)	BHAT Area (supervised)	Method in [45] (unsupervised)	BHAT NoArea (supervised)	M-BHAT NoArea (supervised)	KL NoArea (supervised)	M-KL NoArea (supervised)
↑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

all the proposed methods and a comparison with other state-of-the-art texture segmentation techniques are available online at http://mosaic.utia.cas.cz/.

1) Experiment 6. Supervised Evaluation: The goal of this experiment is to evaluate the performance of the proposed criteria for texture segmentation purposes. In the presented benchmark, our statistical methods were compared with other state-of-the-art segmentation techniques. Here, the evaluation is per-

formed in a supervised manner, i.e., when the number of regions in the evaluated partitions is manually set to the number or regions in the ground truth partitions.

Table III outlines the performance of the proposed methods, using as reference the best texture segmentation technique into the benchmark [45]. The complete evaluation results are available online at http://mosaic.utia.cas.cz/. The conclusions from the benchmark results are that, for most indicators,
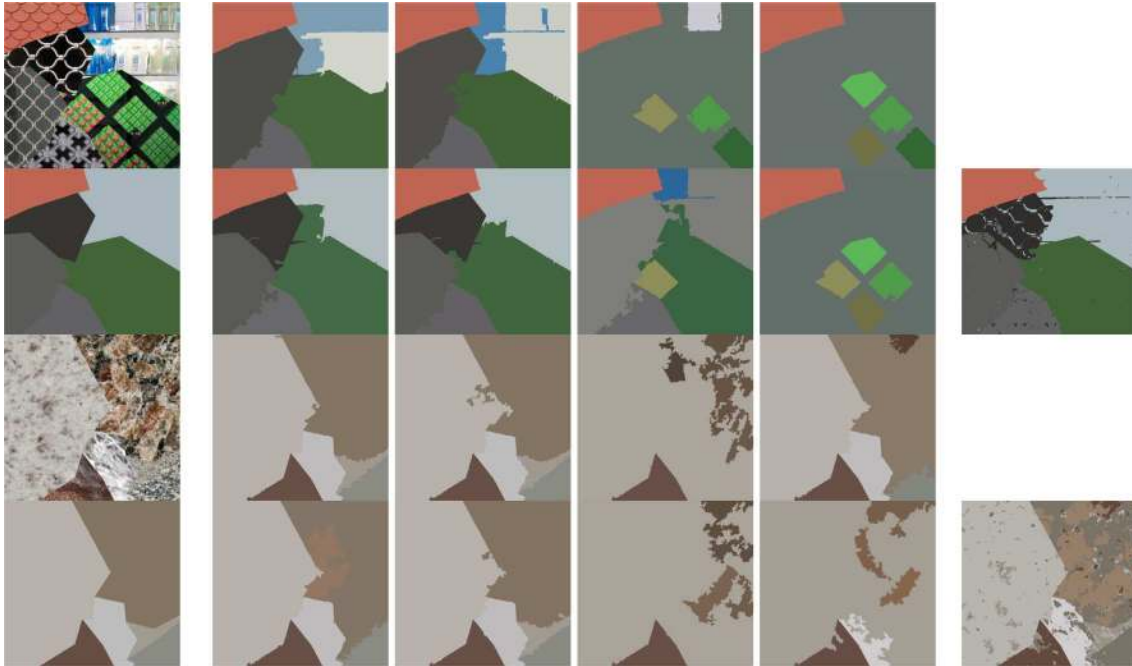
Fig. 13. Merging criteria comparison for the texture database in [55]. For each example, columns from left to right: original image (1st row), ground truth partition (2nd row), KL area-weighted (1st row), M-KL area-weighted (2nd row); BHAT area-weighted (1st row), M-BHAT area-weighted (2nd row); KL area-unweighted (1st row), M-KL area-unweighted (2nd row); BHAT area-unweighted (1st row), M-BHAT area-unweighted (2nd row); method in [45] (2nd row). Types and empirical transition matrices quantized to 10 bins per dimension.

area-weighted statistical approaches outperform the rest of techniques in the benchmark. The area-weighted i.i.d. Kullback–Leibler criterion outperform all other techniques for most indicators, closely followed by the area-weighted Markov-based Bhattacharyya criterion. However, note that similarly to the results shown in the object-oriented evaluation, both area-weighted Bhattacharyya versions provide the best results in terms of undersegmentation error (US indicator in Table III).

Due to the large size of some fundamental parts of the textures (for instance, observe the texture formed by large green squares with a black background in the third example of Fig. 13) and for the general purpose value of the scale threshold proposed in Section IV, area-unweighted methods present a lower performance in most indicators, specially for correct detection (OS), undersegmentation (US) and all pixel-wise average criteria. In this particular case, these methods are not able to compact such large parts into a single textured region, due to its size and heterogeneity. Anyway, if some prior knowledge is available about the granularity of the textures present into the regions, the scale threshold could be specifically set to optimize the performance for a particular type of images. Nevertheless, note that area-unweighted Bhattacharyya criteria exhibit an excellent performance in terms of oversegmentation (OS) indicator (similarly to the object-oriented evaluation), consistency error criteria (GCE and LCE) and variation of information (dVI).

Some segmentation examples are shown in Fig. 13. It can be seen that partitions provided by area-weighted statistical methods have larger quality than partitions obtained by [45], in terms of the correct detection of the textures in the mosaic and into the region connectivity and regularity. In general,

Markov-based and i.i.d. based area-weighted methods provide similar segmentation results, although in some cases the Markov model improves the obtained partition. For instance, see the third and fourth example in Fig. 13, where the Markov versions are able to compact into a single region such complex and heterogenous regions as those formed by the white and blue shampoo bottles and the cactus, respectively.

The stability under various noise types and degradation levels has also been evaluated for the technique providing the best performance in terms of correct segmentation (CS) indicator in supervised mode: the Kullback–Leibler version based on the i.i.d. model. The benchmark system automatically provides a noise corrupted version of the dataset for three noise types (Gaussian, Poisson, and Salt&Pepper) and under different levels of degradation (only for Gaussian and Salt&Pepper). For the sake of conciseness, the results have not been included here, but they are available online at http://mosaic.utia.cas.cz/.

As expected, the performance of the IT-RM technique decays with the SNR for Gaussian noise (SNR values of 10, 5, 0, −5, and −10 dB provide CS values of 56.61, 47.86, 46.68, 36.25, and 15.05, respectively). The performance for an SNR of 0 dB is still high, especially if it is compared to the performance in absence of noise of other state-of-the-art techniques in the database. Similar results are observed for the Poisson corrupted images. For the Salt&Pepper noise, the IT-RM approach is particularly robust, showing an impressive level of resilience to the highest levels of degradation available in the database. For a probability of noise presence in the image of 0.5, the CS value is over 63, which outperforms the performance in absence of noise of the other state-of-the-art techniques in the system.

TABLE IV
STATISTICAL REGION MERGING UNSUPERVISED EVALUATION ON THE PRAGUE TEXTURE SEGMENTATION BENCHMARK. RESULTS DIRECTLY OBTAINED FROM THE SYSTEM IN [55]. 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. FOR BENCHMARK CRITERIA SEE TABLE III

	M-KL Area (unsupervised)	KL Area (unsupervised)	Method in [45] (unsupervised)	BHAT Area (unsupervised)	M-BHAT Area (unsupervised)	BHAT NoArea (unsupervised)	M-BHAT NoArea (unsupervised)	KL NoArea (unsupervised)	M-KL NoArea (unsupervised)
↑CS	54.70	54.05	51.25	47.60	47.31	25.77	19.40	18.26	16.79
↓OS	15.14	15.88	5.84	6.58	24.59	17.91	8.96	14.61	5.50
↓US	23.71	14.45	7.16	27.13	29.10	50.06	58.24	51.60	50.36
↓ME	5.49	13.61	31.64	18.67	3.31	9.64	14.89	12.91	27.57
↓NE	4.55	14.03	31.38	17.86	3.46	9.47	13.72	13.92	27.67
↓O	24.79	23.44	19.65	34.05	31.65	56.13	66.84	61.07	58.99
↓C	31.43	26.47	9.67	14.93	36.53	64.91	45.37	45.19	43.96
↑CA	63.98	66.75	67.45	63.26	59.88	43.56	36.85	39.43	35.43
↑CO	71.55	73.67	76.40	73.21	67.49	56.00	51.99	52.75	51.13
↑CC	75.39	81.41	81.12	70.52	76.27	65.32	57.98	58.72	61.73
↓I.	28.45	26.33	23.60	26.79	32.51	44.00	48.01	47.25	48.87
↓II.	4.53	3.35	4.09	6.19	5.40	13.60	17.42	15.42	18.40
↑EA	69.79	73.14	75.80	69.16	66.84	51.43	44.18	47.32	43.72
↑ MS	61.55	65.75	65.19	60.87	57.99	39.01	30.20	33.11	28.22
↓RM	8.77	6.76	7.21	9.91	9.47	11.78	14.90	15.61	15.34
↑CI	71.46	75.09	77.21	70.44	69.10	54.84	47.83	50.79	48.20
↓GCE	6.75	8.87	20.35	8.88	5.59	10.31	11.56	11.19	15.99
↓LCE	4.82	5.98	14.36	5.76	3.95	5.40	6.66	5.77	9.27
↓dM	17.61	12.36	12.64	17.44	17.83	32.09	39.77	37.41	41.07
↓dD	16.16	15.95	18.01	15.83	17.84	23.84	26.36	25.50	27.84
↓dVI	13.90	14.87	14.06	12.89	14.63	13.55	12.25	12.96	12.35

2) Experiment 7. Unsupervised Evaluation: In this case, the goal of the experiment was to evaluate, into the texture segmentation context, the performance of the proposed statistical region merging techniques working in an unsupervised manner. For that reason, the number of regions of the evaluated partitions is automatically selected by the partition selection criteria presented in Section V, concretely, the first partition in importance order with respect to the significance order in (34).

The results for the joint evaluation of the merging techniques and the selection criterion are outlined in Table IV. As before, all results are available online at http://mosaic.utia.cas.cz/. Despite the decrease in the performance due to the unsupervised approach, the area-weighted methods still outperform most texture segmentation techniques present into the benchmark, showing a performance comparable to [45], the best texture segmentation technique into the benchmark. Although for most indicators area-weighted techniques present similar values, the Kullback–Leibler version based on the i.i.d. model, working together with the selection criteria, presents a good performance in most of them.

VII. CONCLUSION

In this paper, we have proposed a practical solution for the segmentation of generic images. Summarizing, the main characteristics of the presented approach are as follows:

- *generality*, as it can be directly used with images of different complexities and characteristics;
- *accuracy*, as it provides high quality partitions in terms of objective and subjective quality measures;
- *efficiency*, as the region models and criteria are not significantly complex and, hence, the merging procedure can be implemented using fast graph-based algorithms in linear time.

From our evaluation of the proposed region merging techniques, the area-weighted methods exhibit a better performance in terms of minimizing the merging error or undersegmentation, but, in general, suffer from a larger degree of oversegmentation. The area-unweighted extensions have shown a better tradeoff between under- and oversegmentation. In the texture segmentation context, area-unweighted methods are sensible to the characteristics of textures into the regions, namely its size and heterogeneity. In this case, any prior information about the granularity of the textures can be incorporated into the scale-based merging order to optimize the performance of these techniques. When no specific knowledge is available, area-weighted merging strategies are preferred for texture images to obtain a correct and reliable segmentation.

For natural images, merging criteria based on the Bhattacharyya coefficient are slightly superior to those based on the Kullback–Leibler divergence. Particularly the Markov area-weighted Bhattacharyya criterion presents an excellent performance in terms of both under- and oversegmentation. This difference is not appreciated for texture images. However, experiments for both types of images agree that, at least for the area-unweighted extensions, the Bhattacharyya criteria are preferred to the Kullback–Leibler criteria.

Hence, we can conclude that applications where undersegmentation errors are crucial (or where oversegmentation errors are not a problem) can find a valuable solution into area-weighted criteria combined with a classical merging order. The use of a more complex statistical model, such as the proposed first-order Markov model, can further improve the results in terms of undersegmentation. On the contrary, applications being more error tolerant may find into a simple i.i.d. statistical model a more practical solution.

When oversegmentation is also considered as a source of error, area-unweighted criteria combined with a scale-based merging order may be more appropriate, directly for images with no coarse heterogeneous textures (such as generic real images).

Last but not least, these unsupervised techniques can serve as a generic analysis tool in image processing, as they can provide to the user or to the next steps of the processing chain a reduced subset of the most relevant partitions found into the whole hierarchy, without any prior knowledge on the type of analysis to be performed. Particularly, the most significant partition can be directly used as a rough approximation of the semantic content of the image. In turn, when the proposed methods work in a supervised manner, they clearly outperform state-of-the-art object-based and texture segmentation approaches.

Our current work aims at improving the region model quantization using a data-dependent quantization strategy to obtain a more accurate estimation of the discrete distributions characterizing the regions [41]. At the moment, we are also investigating on the fusion of several of the proposed techniques to improve the accuracy and the robustness of the segmentation results. As future guidelines, we would like to extend and validate these techniques to other problems of generic image segmentation such as segmentation of compressed images, segmentation of low resolution versions or extensions to video and 3-D images.

APPENDIX
CHERNOFF BOUND FOR FIRST ORDER MARKOV SEQUENCES

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 [46], 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

$$P_e = \sum_{A^C} \pi_1 P_1 + \sum_A \pi_2 P_2 \qquad (35)$$

$$= \sum_{x \in \mathcal{X}} \min\{\pi_1 P_1, \pi_2 P_2\} \qquad (36)$$

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

$$\min\{a, b\} \le a^\lambda b^{1-\lambda} \text{ for all } 0 \le \lambda \le 1. \qquad (37)$$

Using this to continue the chain, we have

$$P_e \le \sum_{x \in \mathcal{X}} (\pi_1 P_1)^\lambda (\pi_2 P_2)^{1-\lambda} \qquad (38)$$

$$\le \sum_{x \in \mathcal{X}} P_1^\lambda P_2^{1-\lambda}. \qquad (39)$$

For a sequence of observations, we have

$$P_e^{(n)} \le \sum_{x^n \in \mathcal{X}^n} \left(P_1^{(n)}\right)^\lambda \left(P_2^{(n)}\right)^{1-\lambda}. \qquad (40)$$

Particularizing for first-order Markov sequences,

$$P_e^{(n)} \le \sum_{x^n \in \mathcal{X}^n} P_1^\lambda(x_1) P_2^{1-\lambda}(x_2)$$
$$\cdot \prod_{k=2}^n P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1}). \qquad (41)$$

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

$$P_e^{(n)} \le \sum_{x^n \in \mathcal{X}} \prod_{k=2}^n P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1})$$
$$\le \prod_{k=2}^n \sum_{x_k, x_{k-1}} P_1^\lambda(x_k|x_{k-1}) P_2^{1-\lambda}(x_k|x_{k-1})$$
$$\le \prod_{k=2}^n \sum_{i,j} \left(p_{ij}^1\right)^\lambda \left(p_{ij}^2\right)^{1-\lambda}$$
$$= \left(\sum_{i,j} \left(p_{ij}^1\right)^\lambda \left(p_{ij}^2\right)^{1-\lambda} \right)^{n-1} \qquad (42)$$

where $\mathbf{P}_1 = \left(p_{ij}^1\right)$, $\mathbf{P}_2 = \left(p_{ij}^2\right)$ correspond to the transition matrices of the processes, respectively.

Hence, the exponent of the error probability is bounded by

$$\log P_e^{(n)} \le (n-1) \log \left(\sum_{i,j} \left(p_{ij}^1\right)^\lambda \left(p_{ij}^2\right)^{1-\lambda} \right). \qquad (43)$$

Although the whole derivation has been done for the case of $n_1 = n_2 = n$, note that 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 (43)

$$C(\mathbf{P}_1, \mathbf{P}_2) \triangleq - \min_{0 \le \lambda \le 1} \log \left(\sum_{i,j} \left(p_{ij}^1\right)^\lambda \left(p_{ij}^2\right)^{1-\lambda} \right). \qquad (44)$$

REFERENCES

[1] D. Ballard and C. Brown, *Computer Vision*. Englewood Cliffs, NJ: Prentice-Hall, 1982.

[2] D. Forsyth and J. Ponce, *Computer Vision—A Modern Approach*. Englewood Cliffs, NJ: Prentice-Hall, 2003.

[3] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," *Proc. IEEE*, vol. 76, pp. 869–889, Aug. 1988.

[4] P. Salembier and F. Marques, "Region-based representations of image and video: Segmentation tools for multimedia services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1147–1167, 1999.

[5] C. Brice and C. Fennema, "Scene analysis using regions," *Artif. Intell.*, vol. 1, no. 3–4, pp. 205–226, 1970.

[6] P. Salembier and L. Garrido, "Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval," *IEEE Trans. Image Process.*, vol. 9, pp. 561–576, 2000.

[7] O. J. Morris, J. M. Lee, and A. G. Constantinides, "Graph theory for image analysis: An approach based on the shortest spanning tree," *Proc. Inst. Electr. Eng.—F*, vol. 133, no. 2, pp. 146–152, 1986.

[8] S. H. Kwok, A. Constantinides, and W.-C. Siu, "An efficient recursive shortest spanning tree algorithm using linking properties," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 6, pp. 852–863, Jun. 2004.

[9] S. Kwok and A. Constantinides, "A fast recursive shortest spanning tree for image segmentation and edge detection," *IEEE Trans. Image Process.*, vol. 6, no. 2, pp. 328–332, Feb. 1997.

[10] D. Mumford and J. Shah, "Optimal approximations by piecewise smooth functions and associated variational problems," *Commun. Pure Appl. Math.*, vol. 42, no. 4, pp. 577–685, 1989.

[11] L. Garrido, P. Salembier, and D. Garcia, "Extensive operators in partition lattices for image sequence analysis," *Signal Process.*, vol. 66, no. 2, pp. 157–180, 1998.

[12] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, no. 6, pp. 721–741, 1984.

[13] M. Unser, "Texture classification and segmentation using wavelet frames," *IEEE Trans. Image Process.*, vol. 4, no. 11, pp. 1549–1560, Nov. 1995.

[14] T. Wan, N. Canagarajah, and A. Achim, "Statistical multiscale image segmentation via alpha-stable modeling," in *IEEE Int. Conf. Image Processing (ICIP)*, Sep.–Oct. 16-19, 2007, vol. 4, pp. IV-357–IV-360.

[15] A. K. Jain and F. Farrokhnia, "Unsupervised texture segmentation using gabor filters," *Pattern Recogn.*, vol. 24, no. 12, pp. 1167–1186, 1991.

[16] P. Vautrot, N. Bonnet, and M. Herbin, "Comparative study of different spatial/spatial-frequency methods (Gabor filters, wavelets, wavelets packets) for texture segmentation/classification," in *Proc. Int. Conf. Image Processing*, Sep. 16–19, 1996, vol. 3, pp. 145–148.

[17] R. Duda, P. Hart, and D. Stork, *Pattern Classification*, 2nd ed. New York: Wiley, 2000.

[18] M. Pardas and P. Salembier, "Time-recursive segmentation of image sequences," in *Proc. EUSIPCO*, 1994, pp. 18–21.

[19] C. Bennstrom and J. Casas, "Object representation using colour, shape and structure criteria in a binary partition tree," in *Proc. ICIP*, Sep. 11–14, 2005, vol. 3, pp. 1144–7.

[20] T. Adamek and N. O'Connor, "Using Dempster-Shafer theory to fuse multiple information sources in region-based segmentation," in *Proc. ICIP*, Sep. 16–19, 2007, vol. 2, pp. 269–272.

[21] R. Chellappa and A. Jain, *Markov Random Fields: Theory and Application*. Boston, MA: Academic, 1993.

[22] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

[23] D. J. Crisp and G. N. Newsam, "A fast, efficient segmentation algorithm based on region merging," in *Proc. Image Vision Computing New Zealand 2000*, 2000, pp. 180–185.

[24] S. LaValle and S. M. Hutchinson, "Bayesian region merging probability for parametric image models," in *Proc. IEEE Computer Soc. Conf. Computer Vision Pattern Recognition*, Jun. 15–17, 1993, pp. 778–779.

[25] V. Gies and T. Bernard, "Statistical solution to watershed over-segmentation," in *Proc. ICIP*, Oct. 24–27, 2004, vol. 3, pp. 1863–1866.

[26] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int. Conf. Computer Vision*, Jul. 2001, vol. 2, pp. 416–423.

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

[28] F. Nielsen, *Visual Computing: Geometry, Graphics, and Vision*. Rockland, MA: Charles River Media/Thomson Delmar Learning, 2005.

[29] F. Nielsen and R. Nock, "On region merging: The statistical soundness of fast sorting, with applications," in *Proc. IEEE Computer Soc. Conf. Computer Vision Pattern Recognition*, Jun. 18–20, 2003, vol. 2, pp. II-19–II-26.

[30] R. Nock and V. Page, "Grouping with bias for distribution-free mixture model estimation," in *Proc. 17th Int. Conf. Pattern Recognition (ICPR)*, Aug. 23–26, 2004, vol. 2, pp. 44–47.

[31] R. Nock and F. Nielsen, "Grouping with bias revisited," in *Proc. IEEE Computer Soc. Conf. Computer Vision Pattern Recognition (CVPR)*, Jun. 27–Jul. 2, 2004, vol. 2, pp. II-460–II-465.

[32] R. Nock and F. Nielsen, "Semi-supervised statistical region refinement for color image segmentation," *Pattern Recogn.*, vol. 38, no. 6, pp. 835–846, 2005.

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

[34] V. Vilaplana, F. Marques, and P. Salembier, "Binary partition trees for object detection," *IEEE Trans. Image Process.*, vol. 17, no. 11, pp. 2201–2216, Nov. 2008.

[35] C. Dorea, M. Pardas, and F. Marques, "Trajectory tree as an object-oriented hierarchical representation for video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 4, pp. 547–560, Apr. 2009.

[36] F. Calderero and F. Marques, "General region merging approaches based on information theory statistical measures," in *Proc. 15th IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2008, pp. 3016–3019.

[37] F. Calderero and F. Marques, "General region merging based on first order Markov information theory statistical measures," in *Proc. 16th Eur. Signal Processing Conf. (EUSIPCO)*, 2008, pp. 5–5.

[38] L. M. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming," *USSR Comput. Math. and Math. Phys.*, vol. 7, pp. 200–217, 1967.

[39] I. Ciszar, "Information type measures of difference of probability distributions and indirect observations," *Stud. Sci. Math. Hung.*, no. 2, pp. 299–318, 1967.

[40] T. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.

[41] Q. Wang, S. Kulkarni, and S. Verdu, "Divergence estimation of continuous distributions based on data-dependent partitions," *IEEE Trans. Inf. Theory*, vol. 51, no. 9, pp. 3064–3074, Sep. 2005.

[42] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[43] R. G. Gallager, *Information Theory and Reliable Communication*. New York: Wiley, 1968.

[44] R. Haralick, "Statistical and structural approaches to texture," *Proc. IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[45] G. Scarpa, M. Haindl, and J. Zerubia, "A hierarchical finite-state model for texture segmentation," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Apr. 15–20, 2007, vol. 1, pp. 1209–1212.

[46] Z. Rached, ""Information measures for sources with memory and their application to hypothesis testing and source coding,"," Ph.D. dissertation, Queen's Univ., Ontario, Canada, 2002.

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

[48] A. Alatan, L. Onural, M. Wollborn, R. Mech, E. Tuncel, and T. Sikora, "Image sequence analysis for emerging interactive multimedia services—The European cost 211 framework," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 8, no. 7, pp. 802–813, Nov. 1998.

[49] T. Adamek and N. O'Connor, "Stopping region-based image segmentation at meaningful partitions," *Semantic Multimedia, Lecture Notes Comput. Sci.*, vol. 4816, pp. 15–27, 2007.

[50] P. Salembier, J. Llach, and L. Garrido, "Visual segment tree creation for mpeg-7 description schemes," *Pattern Recogn.*, vol. 35, no. 3, pp. 563–579, Mar. 2002.

[51] H. Lu, J. Woods, and M. Ghanbari, "Binary partition tree analysis based on region evolution and its application to tree simplification," *IEEE Trans. Image Process.*, vol. 16, no. 4, pp. 1131–1138, Apr. 2007.

[52] L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 6, pp. 583–598, Jun. 1991.

[53] J. Barthelemy and B. Leclerc, "The median procedure for partitions," *Partitioning Data Sets, DIMACS Series in Descrete Mathematics*, vol. 19, pp. 3–34, 1995.

[54] S. X. Yu and J. Shi, "Grouping with bias," in *Proc. NIPS*, 2001, pp. 1327–1334.

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

Felipe Calderero (S'02) received the M.S. degree in electrical engineering degree from the Technical University of Catalonia (UPC), Barcelona, Spain, in 2004, completing his Master's thesis (with honors) at Northeastern University, Boston, MA, in September 2004. He received the Ph.D. degree (*cum laude*) from the Technical University of Catalonia in February 2010.

Since September 2009, he has held a Marie Curie Grant from the European Commission, working as a Scientific Researcher at the Ultrasound Division of Siemens Healthcare, Erlangen, Germany. He is also collaborating with the Fetal and Perinatal Medicine Group at the Hospital Clínic, Barcelona, Spain, researching in ultrasound and magnetic resonance imaging. His research interests include image segmentation, region-based image analysis, information and modality fusion, medical image analysis, and brain connectomics.

Dr. Calderero received one of the IEEE 125th Anniversary Student Awards.

Ferran Marques (S'91–M'93–SM'09) received the electrical engineering degree and the Ph.D. degree from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1988 and 1992, respectively.

From April 1989 to June 1990, he was at the Swiss Federal Institute of Technology, Lausanne (EPFL), Switzerland, and in 1991, he joined the Signal and Image Processing Institute, University of Southern California, Los Angeles, where he became an Associate Professor in 1995. Since 2003, he has been a Full Professor with the TSC Department, UPC, where he is lecturing in the area of digital signal and image processing. He is the author or coauthor of more than 100 publications that have appeared as journal papers and proceeding articles, as well as two books, five book chapters, and four international patents.

Dr. Marques served as Associate Dean for International Relations of the Telecommunication School (ETSETB) at UPC (1997–2000) and as President of the European Association for Signal Processing EURASIP (2002–2004). He is Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING and has been Special Sessions Co-Chairman in the International Conference on Image Processing (ICIP), Barcelona, Spain, September 2003, and Technical Chairman of the Fourth International Workshop on Content-Based Multimedia Indexing (CBMI), Riga, Latvia, June 2005. He won the Spanish Best Ph.D. Thesis in Electrical Engineering Award in 1992.