A Cluster Grouping Technique for Texture Segmentation

Roberto Manduchi
Jet Propulsion Laboratory
California Institue of Technology
Pasadena, CA 91109
manduchi@jpl.nasa.gov

Abstract

We propose an algorithm for texture segmentation based on a divide-and-conquer strategy of statistical modeling. Selected sets of Gaussian clusters, estimated via Expectation Maximization on the texture features, are grouped together to form composite texture classes. Our cluster grouping technique exploits the inherent local spatial correlation among posterior distributions of clusters belonging to the same texture class. Despite its simplicity, this algorithm can model even very complex distributions, typical of natural outdoor images.

1 Introduction

This paper proposes a simple statistical parametric technique for texture segmentation. The statistical description of textures has received much attention in recent years. Texture features $\mathbf{c}(x)$ are typically extracted from the output of a set of scaled/oriented filters, which are supposed to capture the local salient information in the neighborhood of each image point. Parametric mixture models are the framework of choice for segmentation. These models assume that a feature $\mathbf{c}$ is generated by one of N possible processes ("components"). The probability density function of feature $\mathbf{c}$ can thus be expressed by a *mixture distribution*

$$p(\mathbf{c}) = \sum_{j=1}^{N} P(j)p(\mathbf{c}|j) \tag{1}$$

where $p(\mathbf{c}|j)$ is the conditional likelihood of the feature $\mathbf{c}$ generated by the component j and $P(j)$ is the prior probability of the component j (called *mixing parameter*). The posterior probabilities $P(j|\mathbf{c}(x))$ are derived straightforwardly from the mixture model using Bayes' rule, and are used for the final segmentation. Note that each component of the model corresponds to exactly one image segment.

Mixture models owe their popularity in part to the existence of an efficient technique (the Expectation-Maximization algorithm) for the maximum likelihood parameters estimation [4]. In its simplest formulation, the EM algorithm relies on two hypotheses: 1) a suitable model for the conditional likelihoods is known, and 2) the observed samples are statistically independent. Neither of these hypotheses are verified in typical textures. In this paper we tackle the first problem, the determination of a statistical model for feature generation within each texture class, originating our argument from the observation that simple Gaussian models are inadequate to describe "multimodal" textures, such as can be often encountered in practice.

Mixture of Gaussians are the most common instance of mixture models, one of the reasons being that Gaussian conditional likelihoods allow for the E- and M- steps of the EM algorithm to be solved in closed form [4]. Each Gaussian cluster represents a "mode" of the mixture distribution. Malik *et al.* [2] call the cluster centers "textons" and use them for compact texture representation (via vector quantization). Our main point here is that it is often necessary to use more than one Gaussian cluster to represent an homogeneous texture feature distribution. To deal with multimodal textures we propose a divide-and-conquer strategy. First, extract a suitable number of mixture components using the EM algorithm; then, group together those clusters which are likely to belong to the same texture.

How can we estimate the correct assignments cluster–texture? Our algorithm determines a cost function of the cluster grouping that takes *spatial coherence* into account. A simple, non-iterative technique allows us to determine the cluster groupings that minimize such a function, and the final Bayesian assignment is performed based on the new combined posterior distribution. Results on natural textured scene show the effectiveness of the proposed method.

2 Multimodal texture segmentation

2.1 Problem statement

Our strategy for segmenting multimodal textures is based on "grouping together" some of the components of a given mixture model. More precisely, we rewrite (1) as

$$p(\mathbf{c}) = \sum_{k=1}^{N_1} \sum_{i=1}^{N_2(k)} P_k(i) p_k(\mathbf{c}|i) = \qquad (2)$$

$$= \sum_{k=1}^{N_1} P^{(1)}(k) \left(\sum_{i=1}^{N_2(k)} P_k^{(2)}(i) p_k(\mathbf{c}|i) \right)$$

Note that (2) represents a purely formal operation; the conditional likelihoods that appear in (1) and (2) are the same (albeit with different indices) and the density $p(\mathbf{c})$ remains unchanged. The index k in (2) labels the different texture in the scene; the index i enumerates the clusters within each texture class. A feature $\mathbf{c}$ is assigned to the texture k that maximizes $\sum_{i=1}^{N_2(k)} P_k(i) p_k(\mathbf{c}|i)$.

As anticipated in the Introduction, we will determine the groupings in (2) by exploiting the spatial coherence of the class assignment function. More precisely, we observe that the posterior probabilities $P_k(i_1|\mathbf{c}(x))$ and $P_k(i_2|\mathbf{c}(x))$ for two clusters i_1 and i_2 belonging to the same texture k are typically *spatially correlated.* They can assume high values (≤ 1) only in image areas corresponding to the same texture; for homogeneous textures, it is reasonable to assume that, within a *window of observability* of suitable scale, we will always find pixels belonging to cluster i_1 and pixels belonging to cluster i_2. This notion is exploited in the context of the *maximum descriptiveness* principle for grouping "redundant" clusters in a mixture model. We first discuss the maximum descriptiveness principle, and then show its application in the context of this work.

2.2 Model descriptiveness

Consider a mixture model with density $p(\mathbf{c})$ expressed by (1), The *descriptiveness* of the model [3] is defined by

$$D = \sum_{j=1}^{N} \int p(\mathbf{c}|j) P(j|\mathbf{c}) d\mathbf{c} \qquad (3)$$

where the posterior probabilities $P(j|\mathbf{c})$ are derived from (1) using Bayes' rule. Let us examine each term of the sum in (3). The j-th cluster "describes" each feature $\mathbf{c}$ by means of the conditional likelihood $p(\mathbf{c}|j)$. The posterior probability $P(j|\mathbf{c})$ specifies in a "soft" fashion which features are actually assigned by the model to the j-th cluster. Thus, the integrals in the sum determine how well each cluster describes the features that are assigned to it. It is easily seen that models with "hard" assignment rules have the highest descriptiveness (which can only be less than or equal to N). "Redundant" models with highly overlapping densities $p(\mathbf{c}|j)$ have smaller descriptiveness for the same number of classes. The lowest value of the descriptiveness (D=1) is achieved when all of the conditional likelihoods are identical.

A very useful property of the descriptiveness is that it can be easily estimated: a simple application of Bayes' rule proves the following identity:

$$D = \sum_{j=1}^{N} D_j \ , \ D_j = \frac{E\left[P(j|\mathbf{c})^2\right]}{P(j)} \qquad (4)$$

where $E[\cdot]$ is the expectation computed with respect to the density $p(\mathbf{c})$. The numerator of each term in (4) can thus be estimated by simply averaging $P(j|\mathbf{c}(x))^2$ over the image.

For our purposes, the descriptiveness of a model is not used by itself; it is its *variation* when two or more clusters are grouped together which is of interest to us. Suppose that a new model is generated by grouping two clusters (say, clusters i and j) into a new cluster $i \cup j$ according to the following rules:

$$\begin{aligned} &P(i \cup j) = P(i) + P(j) \\ &P(i \cup j \ |\mathbf{c}) = P(i|\mathbf{c}) + P(j|\mathbf{c}) \\ &p(\mathbf{c}|i \cup j) = p(\mathbf{c}|i) \tfrac{P(i)}{P(i)+P(j)} + p(\mathbf{c}|j) \tfrac{P(j)}{P(i)+P(j)} \end{aligned} \qquad (5)$$

Note that the conditional likelihood defined in the last row of (5) is such that the density $p(\mathbf{c})$ defined by the model does not change: our grouping operation (which is equivalent to (2)) is purely formal and should not affect the unconditional likelihood $p(\mathbf{c})$. However, the model descriptiveness D will change (in general) as an effect of cluster grouping. Indeed, it can be shown that the model descriptiveness D may only decrease or remain unchanged when two or more clusters are grouped together. The descriptiveness decreases the most when the grouping involves clusters with well-separated conditional distributions, while highly overlapping distributions can be grouped with little descriptiveness loss. Thus, by studying the values assumed by the descriptiveness decrement ΔD, we can decide which sets of clusters are "redundant" and can be grouped together in order to reduce the number of clusters of the model. In other words, we will group together clusters by ensuring that the final model has the highest descriptiveness, i.e., by minimizing ΔD.

We will call this strategy the *maximum descriptiveness principle.* A fast sub-optimal technique for minimizing ΔD works by greedy merging two clusters at a time [3].

There is an interesting interpretation of the descriptiveness which will be useful in our work. Suppose we are grouping two clusters of indices i and j. Then, from (4) and (5) we have that

$$\Delta D = D_i \frac{P(j)}{P(i)+P(j)} + D_j \frac{P(i)}{P(i)+P(j)} - \frac{2E\left[P(i|\mathbf{c})P(j|\mathbf{c})\right]}{P(i)+P(j)} \tag{6}$$

The last term in this sum is the cross-correlation between the two distributions, normalized with respect to the average of the corresponding priors. Thus, for given cluster descriptiveness D_i,D_j and prior probabilities $P(i)$,$P(j)$, the two clusters will determine a large descriptiveness decrement when grouped together if the two corresponding distribution are uncorrelated. Since these distributions are actually a function of the spatial position x of the features $\mathbf{c}(x)$, we may use the signal processing definition of cross-correlation as a function of the displacement X:

$$C_{ij}(X) = E\left[P(i|\mathbf{c}(x))P(j|\mathbf{c}(x+X))\right] \tag{7}$$

and rewrite the last term of (6) as $-\frac{2C_{ij}(0)}{P(i)+P(j)}$.

2.3 Cluster-texture assignment

Our goal is to find a criterion that tells us when two clusters belong to the same texture, so that we can group them together as in (2). The maximum descriptiveness criterion described in the previous section is not helpful if applied directly on the posterior probabilities $P(j|\mathbf{c})$: two clusters belonging to the same texture may be well separated in feature space. Instead, we propose to apply the same criterion to the *spatially filtered* version of the posterior probabilities. The intuition behind this strategy is the following. As observed earlier, we expect that the posterior distributions for different clusters belonging to the same texture should be spatially correlated. By spatially smoothing these distributions, we expect that a point that was assigned with high probability to just one cluster will now be softly assigned to a number of clusters belonging to the same texture. Cluster grouping is then determined by applying the maximum descriptiveness algorithm to the smoothed posterior distributions. Note that this procedure is used only to find the correspondence cluster-texture: the final segmentation is operated using the model (2), i.e., based on non-filtered distributions.

We now give a more thorough justification of our method. Let $g(x)$ be an isotropic Gaussian kernel of suitable scale σ, normalized to unit area. Let $\bar{P}(j|x) = \int P(j|\mathbf{c}(t))g(x-t)\,dt$ be the filtered version of the posterior distribution $P(j|\mathbf{c}(x))$ (we dropped the dependency on $\mathbf{c}$ because now $\bar{P}(j|x)$ is a function of a whole ensemble of features in a neighborhood of x). Since $g(x)$ has unit area, it is easily proved that $\bar{P}(j|x)$ for $1 \leq j \leq N$ is still a mass distribution for each x. Furthermore, $\bar{P}(j) = E\left[\bar{P}(j|x)\right] = P(j)$.

Now, consider the cross-correlation function

$$\bar{C}_{ij}(X) = E\left[\bar{P}(i|x)\bar{P}(j|(x+X)\right] \tag{8}$$

It is easy to prove that

$$\bar{C}_{ij}(X) = \int C_{ij}(x)\bar{g}(X-x)dx \tag{9}$$

where $C_{ij}(x)$ is defined in (7) and $\bar{g}(x) = \int g(t)g(t-x)dx$ (note that $\bar{g}(x)$ is a unit-area Gaussian kernel with standard deviation $\bar{\sigma} = \sigma/2$). Therefore $\bar{C}_{ij}(0)$ is a weighted average of the cross-correlation between the i-th and the j-th posterior distributions within a neighborhood of radius proportional to $\sigma/2$ (which we will call the *observation window*).

Now consider the decrement of descriptiveness $\Delta\bar{D}$ consequent to grouping two clusters i and j after spatial smoothing:

$$\Delta\bar{D} = \bar{D}_i \frac{P(j)}{P(i)+P(j)} + \bar{D}_j \frac{P(i)}{P(i)+P(j)} - \frac{2\bar{C}_{i,j}(0)}{P(i)+P(j)} \tag{10}$$

From (10) we maintain that, for given $P(i|\mathbf{c}(x))$, $P(j|\mathbf{c}(x))$ and priors $P(i)$, $P(j)$, the value $\Delta\bar{D}$ depends on the degree of local spatial correlation between the two posterior distributions. Thus, the maximum descriptiveness algorithm applied on the smoothed distributions will correctly determine which cluster posterior distributions best correlate, and will group them together into texture classes.

2.4 Experiments

We present here the segmentation results using our method with the real-world 'Zebras" image (Figure 1(a)).

The vectors formed by the magnitude of the output of complex Gabor filters at three scales and four orientations have been used as texture features. The images were 134×222 pixels in size; the Gaussian filter used to smooth the posterior distributions for cluster-texture assignment had standard deviation $\sigma = 20$. In both cases, we started with a mixture model composed of eight Gaussian clusters. The EM algorithm was bootstraped by choosing initial parameter values with K-means clustering, and was stopped after

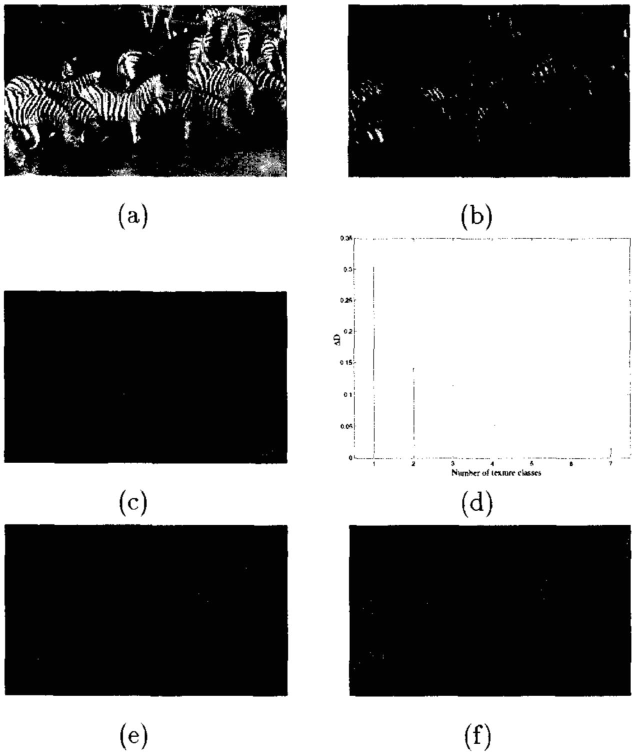

Figure 1: (a): "Zebras" image. (b): Segmentation with eight clusters. (c) Segmentation into two texture classes by cluster grouping. (d): Model descriptiveness decrement as a function of the number of texture classes. (e),(f): Segmentation into three and four texture classes by cluster grouping.

twenty iterations. In passing, we noticed that increasing the number of clusters reduces the risk of missing global minima in the EM iterations. A simple post-processing technique [5] was used to enforce spatial coherence on the resulting multimodal posterior distribution. This algorithm is in essence a "soft" version of Besag's Iterated Conditional Modes [1]; its relation to the mean field theory is discussed in [6].

Note that the whole set of striped shapes (six clusters) has been grouped into one segment. That the final segments *are not* the union of segments found with the 8-clusters segmentation: in other words, cluster grouping determines new Bayesian assignments that are not trivially derived from the original ones. using the greedy strategy of [3].

3 Conclusions

We presented a divide-and-conquer strategy for texture segmentation. The behavior of the texture features in the scene is first modeled by a number of Gaussian clusters, estimated via Expectation Maximization. Then, selected cluster sets are grouped together to form texture classes. Spatial correlation of the posterior cluster distributions is at the basis of the cluster grouping criterion.

Despite its simplicity, this algorithm can model even very complex and multimodal distributions, such as typically appear in natural outdoor images. Future work will be devoted to incorporating other visual features within the same modeling framework.

References

[1] J. Besag. On the statistical analysis of dirty pictures. *J. R. Statist. Soc. B*, 48(3):259–302, 1986.

[2] J. Malik, S. Belongie, J. Shi, T. Leung. Textons, contours and regions: cue integration in image segmentation. *Proc. ICCV*, 918–925, Kerkyra, 1999.

[3] R. Manduchi. Bayesian fusion of texture and color segmentations. *Proc. ICCV*, 956–962, Kerkyra, 1999.

[4] G.J. McLachlan, T. Krishnan. *The EM algorithm and extensions.* John Wiley and Sons, 1997.

[5] J. Zhang, J.W. Modestino, and D.A. Langan. Maximum-likelihood parameter estimation for unsupervised stochastic model-based image segmentation. *IEEE Trans. Image Proc.*, 3(4), 404–420, July 1994.

[6] J. Zhang and J.W. Modestino. The mean-field theory in EM procedures for Markov random fields. *Proc. IEEE ISIT*, Budapest, 1991.