An Optimization Approach to Unsupervised Hierarchical Texture Segmentation*

T. Hofmann

Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, MA 02139

J. Puzicha and J.M. Buhmann

Institut für Informatik III
University of Bonn
53117 Bonn, Germany

Abstract

In this paper we introduce a novel optimization framework for hierarchical data clustering and apply it to the problem of unsupervised texture segmentation. The proposed objective function assesses the quality of an image partitioning simultaneously at different resolution levels and yields a sequence of consistently nested image segmentations. A novel model selection criterion to select significant image structures from various scales is proposed. As an efficient deterministic optimization heuristic a mean–field annealing algorithm is derived.

1 Introduction

Texture provides an important cue for the segmentation of a large class of images. While supervised methods rely on predefined texture classes and the strong notion of optimal texture *discrimination*, unsupervised methods require a general *similarity* or *homogeneity measure* and are founded on the weaker notion of *texture proximity*. The question of texture distinctiveness per definition does not arise in the supervised context but is one of the main challenges for any unsupervised approach, and has often been addressed in terms of cluster validation to identify the 'correct' number of textures [6].

We present an optimization approach to hierarchical image segmentation which yields a nested sequence of refined image partitionings and captures relevant image structure at multiple resolutions. To select a subset of meaningful levels of resolution we propose a model selection criterion to identify stable partitionings. The presented algorithm is a hierarchical extension of methods previously presented in [4, 5] and relies on *mean-field annealing* as an efficient optimization heuristics [1, 3].

2 Texture Dissimilarity

Since natural textures arise at a wide range of spatial frequencies we use a multi–scale image representation by convolution with a set of complex *Gabor filters* [6]. Inspired by [2], homogeneity between pairs of texture patches is measured by a *non–parametric statistical test* applied to the empirical feature distribution functions of locally sampled Gabor coefficients. For each image site $\vec{x}_i$ and Gabor filter r we define a squared window or *block* B_i^r of size $n(r)$ to capture the local texture characteristics. $n(r)$ is chosen proportional to the inverse center frequency of the Gabor filter. For each block B_i^r the empirical distribution of Gabor coefficients $(b_s^r)_{1\le s\le n(r)}$,

$$f_i^r(t) = \left|\, \{t_{u-1} < b_s^r \le t_u\} \,\right|, \quad t \in (t_{u-1}, t_u]\,, \qquad (1)$$

is calculated, where $(t_u)_{0\le u\le M}$ is an appropriate binning. The dissimilarity between two blocks (B_i, B_j) with respect to channel r is then given by the χ^2-statistic[1]

$$D_{ij}^{(r)} = \sum_{u=1}^{M} \frac{(f_i^r(t_u) - f_{ij}^r(t_u))^2}{f_{ij}^r(t_u)}, \quad f_{ij}^r = \frac{f_i^r + f_j^r}{2}. \qquad (2)$$

These values are finally combined by the L_1-norm, $D_{ij} = \sum_r D_{ij}^{(r)}$.

To guarantee computational efficiency the evaluation of $f_i^r(t)$ is restricted to N positions $\vec{x}_i$ on a regular image sub–lattice. Moreover, dissimilarities are only evaluated for a substantially reduced (irreflexive and symmetric) neighborhood $\mathcal{N}$. Following [2] the neighborhood of a site $\vec{x}_i$ consists of the four adjacent sites and a larger number of random neighbors.

3 Hierarchical Clustering

To partition an image a unique region label has to be assigned to each block B_i. Given N blocks and a maximal number of labels K an image segmentation is represented by a Boolean assignment matrix $\mathbf{M} \in \{0,1\}^{N\times K}$, where $M_{i\nu} = 1$ indicates that block B_i is assigned to the cluster of blocks with label ν. As an objective function the cluster compactness criterion

$$\mathcal{H}_K^{cc}(\mathbf{M}) = \sum_{\nu=1}^{K} P_\nu \cdot \mathcal{C}_\nu\,,\ \mathcal{C}_\nu = \frac{\sum_{(i,j)\in\mathcal{N}} M_{i\nu} M_{j\nu} D_{ij}}{\sum_{(i,j)\in\mathcal{N}} M_{i\nu} M_{j\nu}}. \qquad (3)$$

*Supported by a MIT Faculty Sponser's Discretionary Fund, by the German Federal Ministry for Education, Science and Technology (BMBF) under grant # 01 M 3021 A/4, and by the German Research Foundation (DFG) under grant # BU 914/3-1.

[1] Other statistical test have been empirically investigated in detail.

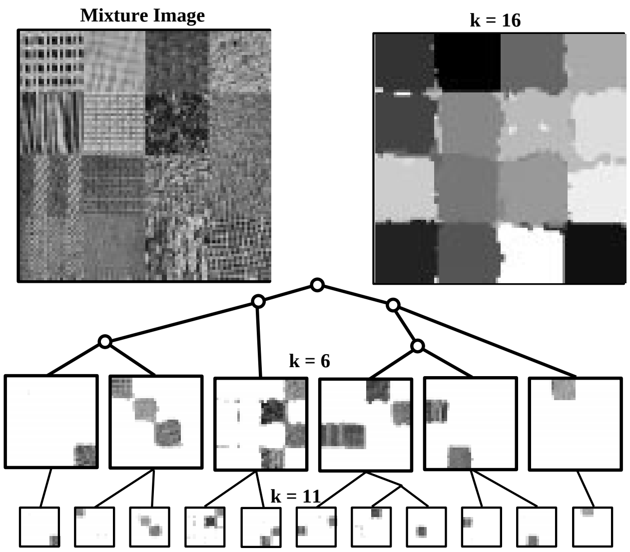

Figure 1: Mixture image with 16 Brodatz microtextures. For the segmentation 12 Gabor filters on three octaves were used. The resolution is $N = 64^2$, $K = 24$ and $|\mathcal{N}|/2 = 150 \cdot N$ evaluated dissimilarities.

is utilized. Here $P_\nu = \sum_{i=1}^{N} M_{i\nu}$ denotes the cluster size and $\mathcal{C}_\nu$ is the average cluster compactness. $\mathcal{H}_K^{cc}$ can be rigorously derived in an axiomatic approach based on fundamental invariance principles [5].

Restricting hierarchical representations to complete binary trees, we assume a tree $\mathcal{T}$ with $K-1$ inner nodes n_α and K leaf nodes l_ν to be given. Numbers of inner nodes represent the sequence of expansions, e.g., the immediate successors of the root n_1 represent a segmentation with $k = 2$, the successors of n_1 and n_2 define a $k = 3$ solution, and so on. The leaf layer constitutes the finest segmentation with $k = K$ textures. For a given leaf node assignment matrix $\mathbf{M}$ we denote by $\mathbf{M}(k)$ the coarse partition with $k \leq K$ clusters, obtained by pruning all nodes except inner nodes n_α, $\alpha < k$, and its immediate successors. We propose the following hierarchical objective function:

$$\mathcal{H}^{hi}(\mathbf{M}) \quad = \quad \sum_{k=2}^{K} w(k) \cdot \mathcal{H}_k^{cc}(\mathbf{M}(k)). \qquad (4)$$

$\mathcal{H}^{hi}$ sums cost contributions at all resolution levels k, weighted with $w(k) \in \mathrm{I\!R}_0^+$, for example, $w(k) = 1$. The presented optimization approach guarantees a consistent and strictly nested hierarchy of segmentations, since clusters formed at inner nodes are per definition given by the union of all clusters associated with its leaf node successors.

Not all k–partitionings can be considered as defining 'natural' solutions. Therefore we propose the following validation criterion in order to select informative levels of the hierarchy. To all k–partitioning costs $\mathcal{H}_k^{cc}(\mathbf{M}(k))$ we add *complexity costs* $\mathcal{H}^{cmp}(k) = \lambda N \log(k)$, $\lambda \in \mathrm{I\!R}^+$. Only those k–segmentations are chosen which possess a range of λ where they have the lowest total costs. Complexity costs proportional to $\log K$ are motivated from the expected cost decay for a random instance in the $N \to \infty$ limit. Notice, that this index does not identify a single 'true' partition, but only eliminates implausible segmentations, which are suboptimal for *all* values of λ.

4 Mean-field Annealing

In the preceding section we have arrived at a formulation of hierarchical clustering in an optimization framework. However, two problems remain so far unresolved: First, we have to specify an efficient optimization heuristic to minimize $\mathcal{H}^{hi}$ and second, we need a mechanism to generate a suitable tree topology. Both problems can be combined elegantly by a technique known as *deterministic annealing* (DA) [7] or, more specifically, as *mean-field annealing* (MFA) [1, 3].

Annealing methods try to avoid unfavorable local minima by artificially introducing noise to an optimization problem. The noise amplitude is controlled by a parameter T, called the computational temperature. Starting from high temperatures, the noise is gradually decreased until the original optimization problem is recovered in the $T \to 0$ limit. The most prominent representative of annealing methods is *simulated annealing* (SA), where the temperature controls the degree of randomization utilized in combination with some local search procedure. In contrast, MFA does not involve Monte Carlo search but directly incorporates the noise into the objective function in terms of an additional entropy contribution. The combinatorial search problem in the space Ω of 'hard' data partitionings is replaced by the problem of identifying a probability distribution P over Ω which minimizes the *generalized free energy*

$$\mathcal{F}_T(P) = \sum_{\mathbf{M} \in \Omega} \left(\mathcal{H}^{hi}(\mathbf{M}) + T \log P(\mathbf{M})\right) P(\mathbf{M}) \, . \quad (5)$$

It is a well-known fact, that $\mathcal{F}_T$ is minimized by the *Gibbs distribution* at temperature T. Since a straightforward computation of Gibbs-averages is intractable, we make the further assumption that assignment probabilities at different sites are independent (*mean-field approximation*). This results in a variational problem with parameters $\langle M_{i\nu} \rangle$, denoting the marginal probability of assigning site $\vec{x}_i$ to cluster ν.

As a first step, we calculate the partial cost difference $\psi_{i\nu}$ between assigning or not assigning site $\vec{x}_i$ to a leaf l_ν of the tree, given the labels of all other sites. Denote by $\nu(k)$ the predecessor of l_ν which enters into the k–partition $\mathbf{M}(k)$. From (4) we obtain $\psi_{i\nu} = \sum_{k=2}^{K} w(k) \phi_{i\nu(k)}$, where

$$\phi_{i\alpha} = \mathcal{C}_\alpha^{+i} + (\mathcal{C}_\alpha^{+i} - \mathcal{C}_\alpha^{-i}) \sum_{j \neq i} M_{j\alpha}. \qquad (6)$$

$\mathcal{C}_\alpha^{+i}$ and $\mathcal{C}_\alpha^{-i}$ denote the compactness of (super–) cluster α after inclusion and exclusion of site $\vec{x}_i$, respectively (c.f. (3)). From $\psi_{i\nu}$ explicit solutions for assignment probabilities $\langle M_{i\nu} \rangle$ of site $\vec{x}_i$ given probabilistic

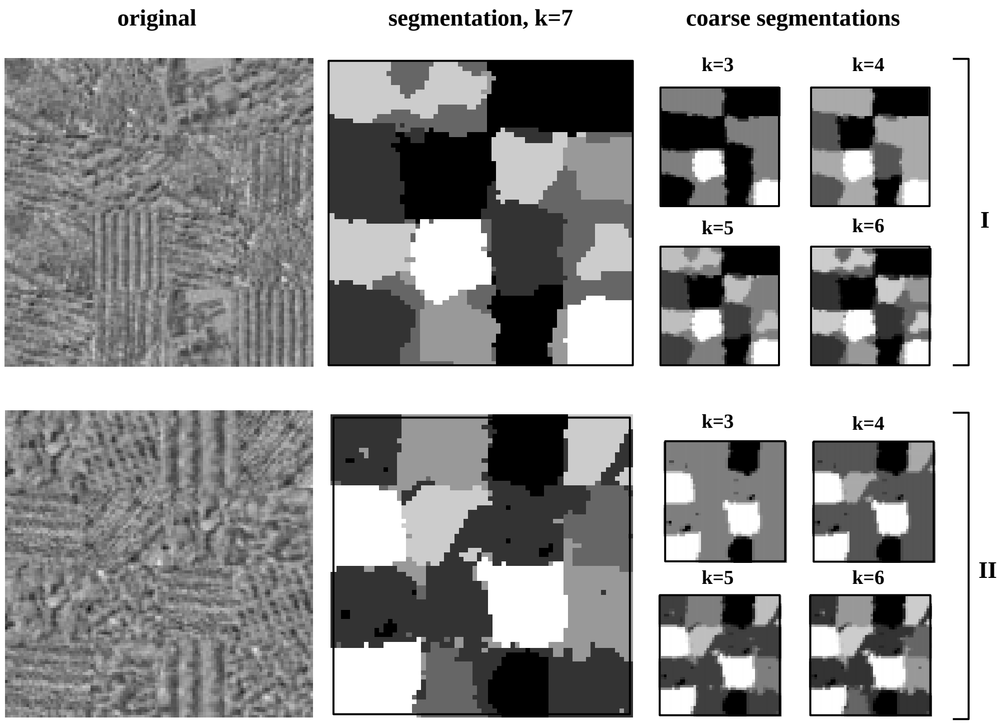

Figure 2: Two synthetical mixtures generated from aerial images of 7 different types of vegetation.

assignments of all other sites can be calculated, because for factorial distributions P

$$\frac{\partial \mathcal{F}_T}{\partial \langle M_{i\nu} \rangle} = \psi_{i\nu}(\langle \mathbf{M} \rangle) + T \log \langle M_{i\nu} \rangle + \lambda_i \ , \qquad (7)$$

where $\psi_{i\nu}(\langle \mathbf{M} \rangle)$ are the expected partial costs, obtained from $\psi_{i\nu}(\mathbf{M})$ by replacing Boolean assignments $M_{i\nu}$ with their expectations $\langle M_{i\nu} \rangle$. λ_i is a Lagrange parameter to enforce the normalization $\sum_\nu \langle M_{i\nu} \rangle = 1$. Setting (7) equal to zero results in the so-called mean-field equations

$$\langle M_{i\nu} \rangle = \frac{\exp\left[-\frac{1}{T}\psi_{i\nu}(\langle \mathbf{M} \rangle)\right]}{\sum_{\mu=1}^{K} \exp\left[-\frac{1}{T}\psi_{i\mu}(\langle \mathbf{M} \rangle)\right]}. \qquad (8)$$

The $K \cdot N$ equations in (8) constitute a system of coupled transcendental equations which can be solved by a convergent single-site update procedure. We refer to [3, 5] for more details.

As a major advantage MFA offers the possibility to grow a meaningful tree topology by tracking the phase transitions (cluster splits) during the annealing process [7]. At high temperatures the entropy maximization dominates and the only solution to the mean-field equations is the trivial uniform distribution, $\langle M_{i\nu} \rangle = 1/K$. This solution becomes unstable at the first phase transition, where two top-level clusters get separated. The remaining degeneracy in the left and right subtrees vanishes successively at subsequent bifurcation points, until no two leaf clusters are identical. Notice that the optimization of assignments is non-greedy, i.e., the split at the first bifurcation may differ from the final $k = 2$ solution of the hierarchy; it is only the topology which is 'grown' incrementally. The described tree growing technique is a unique feature of annealing methods and has been applied in all our experiments.

5 Results

The result of a hierarchical segmentation on a test image generated from 16 Brodatz textures is depicted in Fig. 1. All textures have been correctly identified, borders are localized precisely. We would like to stress that the result has been obtained without incorporating prior knowledge about the spatial relationship of different sites. The clustering procedure treats each site equally without referring to its original position in the image. Further improvements are possible by including typical MRF priors. The application of the validation criterion lead to the detection of two stable solutions for $k = 11$ and the correct $k = 16$ (c.f. Fig 4 (a)).

Fig. 2 shows the result of a segmentation of 2 mixture images which have been artificially generated from aerial images. Each mixture consists of 16 patches from 7 distinctive textures. Most of the computed segments correspond to correct texture patches. In particular, unconnected components of the same texture are correctly identified. The true $k = 7$ has been recovered by the validation criterion.

A segmentation example for an aerial image with the same set of parameters is shown in Fig. 3. The

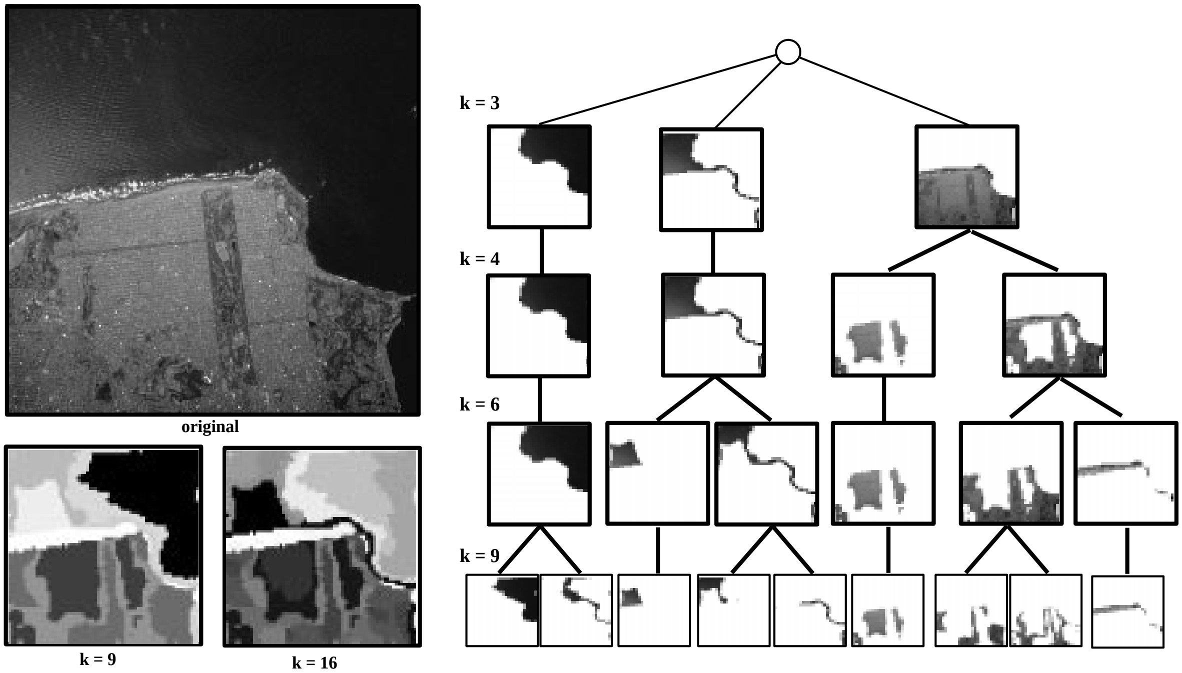

Figure 3: Aerial image and hierarchical segmentation of a section of San Francisco.

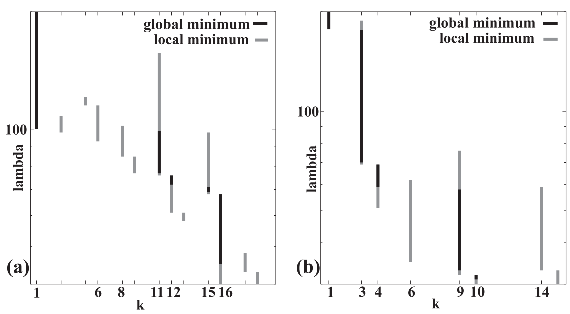

Figure 4: Validation criterion for (a) mixture image and (b) aerial image.

proposed validation criterion selected the segmentations with $k = 3, 4$ and 9. $k = 6$ possesses significant local stability (c.f. Fig. 4 (b)). The hierarchical organization is very intuitive: the first split separates land and ocean. At later stages homogeneously tilled areas are distinguished from vegetation and the surf.

6 Summary and Conclusions

A fully unsupervised hierarchical texture segmentation algorithm based on clustering of dissimilarity data has been presented which is able to identify a consistent hierarchy of successively refined segmentations. The algorithm has a sound foundation in optimization theory and does not require parameter tuning. Moreover it has proven to scale in a benign fashion with the number of textures.

Acknowledgments

The authors are grateful to J.M.H. du Buf for providing the image data depicted in Fig. 2.

[1] G. Bilbro, W. Snyder, S. Garnier, and J. Gault. Mean field annealing: A formalism for constructing GNC–like algorithms. *IEEE Trans. on Neural Networks*, 3(1), 1992.

[2] D. Geman, S. Geman, C. Graffigne, and P. Dong. Boundary detection by constrained optimization. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):609–628, July 1990.

[3] T. Hofmann and J. Buhmann. Pairwise data clustering by deterministic annealing. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 19(1):1–14, 1997.

[4] T. Hofmann, J. Puzicha, and J. Buhmann. Unsupervised segmentation of textured images by pairwise data clustering. In *Proceedings of the IEEE International Conference on Image Processing*, volume III, pages 137–140, 1996.

[5] T. Hofmann, J. Puzicha, and J. Buhmann. Deterministic annealing for unsupervised texture segmentation. In *EMMCVPR, Venice*, 1997.

[6] A. Jain and F. Farrokhnia. Unsupervised texture segmentation using Gabor filters. *Pattern Recognition*, 24(12):1167–1186, 1991.

[7] K. Rose, E. Gurewitz, and G. Fox. Statistical mechanics and phase transitions in clustering. *Physical Review Letters*, 65(8):945–948, 1990.