SCI INSTITUTE

TECHNICAL REPORT

Nonparametric Statistics of Image Neighborhoods for Unsupervised Texture Segmentation

Suyash P. Awate, Tolga Tasdizen, and Ross T. Whitaker

UUSCI-2005-003

Scientific Computing and Imaging Institute
University of Utah
Salt Lake City, UT 84112 USA

April 19, 2005

Abstract:

In this paper, we present a novel approach to unsupervised texture segmentation that is based on a very general statistical model of image neighborhoods. We treat image neighborhoods as samples from an underlying, high-dimensional probability density function (PDF). We obtain an optimal segmentation via the minimization of an entropy-based metric on the neighborhood PDFs conditioned on the classification. Unlike previous work in this area, we model image neighborhoods directly without preprocessing or the construction of intermediate features. We represent the underlying PDFs nonparametrically, using Parzen windowing, thus enabling the method to model a wide variety of textures. The entropy minimization drives a level-set evolution that provides a degree of spatial homogeneity. We show that the proposed approach easily generalizes, from the two class case, to an arbitrary number of regions by incorporating an efficient multi-phase level-set framework. This paper presents results on synthetic and real images from the literature, including segmentations of electron microscopy images of cellular structures.

THE UNIVERSITY OF UTAH

Nonparametric Statistics of Image Neighborhoods for Unsupervised Texture Segmentation

Suyash P. Awate   Tolga Tasdizen   Ross T. Whitaker
School of Computing, University of Utah, Salt Lake City, Utah 84112
{suyash,tolga,whitaker}@cs.utah.edu

Abstract

In this paper, we present a novel approach to unsupervised texture segmentation that is based on a very general statistical model of image neighborhoods. We treat image neighborhoods as samples from an underlying, high-dimensional probability density function (PDF). We obtain an optimal segmentation via the minimization of an entropy-based metric on the neighborhood PDFs conditioned on the classification. Unlike previous work in this area, we model image neighborhoods directly without preprocessing or the construction of intermediate features. We represent the underlying PDFs nonparametrically, using Parzen windowing, thus enabling the method to model a wide variety of textures. The entropy minimization drives a level-set evolution that provides a degree of spatial homogeneity. We show that the proposed approach easily generalizes, from the two-class case, to an arbitrary number of regions by incorporating an efficient multi-phase level-set framework. This paper presents results on synthetic and real images from the literature, including segmentations of electron microscopy images of cellular structures.

1. Introduction

Image segmentation is one of the most extensively studied problems in computer vision. Many different approaches have been proposed for the partitioning of images based on a variety of criteria including brightness (intensity), color, texture, depth, and motion. This paper addresses the problem of segmenting textured images. By *texture segmentation* we mean the partitioning of static, grayscale images, with regions that are not easily distinguished from one another by their average intensity values. Textured regions do not necessarily adhere to the piecewise-smooth or piecewise-constant assumptions that underly most intensity-based segmentation problems; they are defined by some kind of regularity in the higher-order statistics of their pixel neighborhoods.

In recent years researchers have pushed the state-of-the-art in texture segmentation in several important directions. The first direction concerns the mechanism used to model or quantify the regularity in image textures. Researchers have been developing progressively richer descriptions of local image structure and thereby capturing more complex and subtle distinctions between regions. Another direction has been in modeling the variability with textured regions, typically through more sophisticated, statistically-based metrics. Finally, the research in texture segmentation, like segmentation in general, has pursued more robust mechanisms for enforcing geometric regularity in texture segmentations, usually through the construction of a patchwork of regions that simultaneously minimize a set of geometric and statistical criteria.

This paper advances the state-of-the-art in texture segmentation by proposing a strategy that takes the richness of the image descriptors and the generality of the statistical representations to new levels. The proposed method relies on a complete representation of image neighborhoods that exist in a very high dimensional probability space. It relies on nonparametric description of image statistics, and therefore imposes very few assumptions on the statistical structure of neighborhoods. Therefore, it is easily applicable to a wide range of segmentation problems. The proposed method also incorporates relatively recent advances in the computation of level-set evolution, and hence offers a practical way of combining sufficient levels of geometric regularity with an extensive set of statistical computations.

2 Related Work

Many texture segmentation algorithms rely on *signatures* to summarize the underlying spatial intensity patterns. This strategy reduces the dimensionality (and complexity) of neighborhood relationships, thereby making analysis more manageable. For instance, intensity (or grayscale) histograms, computed in local neighborhoods go a long way toward capturing interesting differences in textures [15, 14]. As in this paper, that work relies on nonparametric density estimation. However, the grayscale intensity statistics (i.e. 1D histograms), fail to capture the *geometric* structure of neighborhoods, which is essential for distinguishing textured regions with similar grayscale distributions. Thus, we have taken nonparametric estimation to a very large number of dimensions in order to capture local geometric structure. Much of the previous work in texture analysis relies on filter banks to described the local structure of images. For instance, Gabor filters [11] produce texture features that have been used to discriminate textured regions [20, 23, 24]. Gabor filters are a prominent example of a very large class of oriented, multiscale filters [4, 27].

Other authors have proposed even more compact representations. For instance, Bigun *et al.* [2] use the structure tensor (second-order moment matrix—used, for instance, to analyze flow-like patterns [30]), for texture segmentation. Rousson *et al.* [22] refine this strategy by using vector-valued anisotropic diffusion instead of Gaussian blurring to compute the structure tensor. However, this strategy requires that the structure tensors of the image has a sufficient degree of homogeneity within patches and sufficient degree of difference between patches. Not all images meet these criteria, and we propose to use the full statistics of image neighborhoods; in other words, the complete set of unfiltered pixel intensities.

Recently, researchers have investigated modeling image statistics more directly. For instance, Doretto *et al.* [6] use a hidden Markov process to model the relationships between all pixels within regions, and they apply their method to dynamic textures, capturing relationships in space and time. However, that method assumes a Gaussian process, a very strong assumption that cannot easily account for complex or subtle texture geometries. We take an alternative approach, which is to use a more general statistical model but limit this model to the analysis of small to medium image neighborhoods.

Researchers analyzing the statistics of natural images in terms of local neighborhoods have drawn conclusions that are consistent with Markov random field models of images [12]. For instance, Huang *et al.* [13] analyze single pixel statistics, two-point statistics and derivative statistics of natural images. They found that the mutual information between the intensities of two adjacent pixels in natural images is rather large and attributed this to the presence of spatial correlation in the images. Lee *et al.* [16] and Silva *et al.* [5] analyze the statistics of 3×3 high-contrast patches in optical images, in the corresponding high-dimensional spaces, and find the the data to be concentrated in clusters and low-dimensional manifolds exhibiting a nontrivial topologies.

Popat *et al.* [21] were among the first to use nonparametric Markov sampling in images. They attempt to capture the higher-order nonlinear image statistics via cluster-based nonparametric density estimation and apply their technique for image restoration, image compression and texture classification. However, their method takes a *supervised* approach for learning neighborhood relationships. The work in this paper also relies on the hypothesis that natural images exhibit some regularity in neighborhood structure, but this regularity is discovered for each image individually in a nonparametric manner. The proposed method builds on the work in [1], which lays down the foundations for unsupervised learning of higher-order image statistics. That work however, proposes reducing the entropy of image-neighborhood statistics as a method for *denoising* grayscale images.

The literature dealing with *texture synthesis* also sheds some light on the proposed method. Texture synthesis algorithms rely on image statistics from an input image to construct novel images that exhibit a qualitative resemblance to the input [3, 9, 29]. This paper describes a very different application, but the texture synthesis work demonstrates the power of neighborhood statistics in capturing essential aspects of image structure.

This paper also relies on a rather extensive body of work on variational methods for image segmentation. In particular the Mumford-Shah model [18], its extensions to motion, depth, and texture [18], and its implementation via level-set flows [28]. In particular we use the very fast approximation proposed by Esedoglu [10], and extend it to include multiple regions within a probabilistic framework.

3. Methodology

This section overviews the random field image model with the associated notation and then describes the optimal segmentation formulation based on an entropy minimization.

3.1. Random Field Image Model

A random field/process [7] is a family of random variables $X(\Omega; T)$, for an index set T, where, for each fixed $T = t$, the random variable $X(\Omega; t)$ is defined on the sample space Ω. If we let T be a set of points defined on a discrete Cartesian grid and fix $\Omega = \omega$, we have a realization of the random field called the *digital image*, $X(\omega, T)$. In this case $\{t\}_{t \in T}$ is the set of pixels in the image. For two-dimensional images t is a two-vector. If we fix $T = t$ and let ω vary then $X(t)$ is a random variable on the sample space. We denote a specific realization $X(\omega; t)$ (the image), as a deterministic function $x(t)$.

If we associate with T a family of pixel neighborhoods $N = \{N_t\}_{t \in T}$ such that $N_t \subset T$, and $u \in N_t$ if and only if $t \in N_u$, then N is called a neighborhood system for the set T and points in N_t are called neighbors of t. We define a random vector $Z(t) = \{X(t)\}_{t \in N_t}$, denoting its realization by $y(t)$, corresponding to the set of intensities at the neighbors of pixel t. For the formulation in this paper, we assume that the intensities in each texture patch arise out of a stationary ergodic process. For notational simplicity we use the short hand for random variables $X(t)$ as X and their realizations $x(t)$ as x, dropping the index t.

3.2. Texture Segmentation via Entropy Minimization

Let $p_k(Z = z)$ be the probability of observing the image neighborhood z given that the center pixel of the neighborhood belongs to the texture class k. The total entropy associated with a set of K texture probability distributions is

$$h = -\sum_{k=1}^{K} \int_{\Re^m} p_k(Z = z) \log p_k(Z = z) dz \qquad (1)$$

where $m = |N_t|$ is the neighborhood size. Let $\{T_k\}_{k=1}^{K}$ denote a mutually exclusive and exhaustive decomposition of the image domain into regions generated by the K texture classes. Let $R_k : T \to [0, 1]$ denote the indicator function for T_k, i.e. $R_k(t) = 1$ for $t \in T_k$ and $R_k(t) = 0$ otherwise. The total entropy generated by a particular region decomposition is

$$h = -\sum_{k=1}^{K} \int_{T} R_k(t) p_k(z(t)) \log p_k(z(t)) dt \qquad (2)$$

We consider the optimal decomposition to be the set of functions R_k where h attains a minimum. The strategy in this paper is, therefore, to minimize the total entropy given in (2) by manipulating the regions defined by R_k. This rather large nonlinear optimization problem has, potentially, many local minima. Furthermore, as a practical consideration, textures incorporate some degree of randomness, and regions that are too small will not generate a sufficient number of samples to estimate their statistics. To regularize the solution, and alleviate these problems, variational formulations typically penalize the boundary length of the segmentation [18]. With this modification the objective function becomes

$$E = h + \alpha \sum_{k=1}^{K} \int_{T} \parallel \nabla R_k(t) \parallel dt, \qquad (3)$$

where α is a regularization parameter.

3.3. Level Set Formulation

The level set framework [19] is an attractive option for solving the variational problem defined by (3), because it does not restrict the shapes or topologies of regions. However, the method carries some significant computational costs—in particular the CFL condition for stability limits the motion of the moving wavefront (patch boundaries) to one pixel per iteration.

Recently, Esedoglu and Tsai introduced a fast level set algorithm based on threshold dynamics for minimizing Mumford-Shah type energies [10]. In this paper, we adopt their approach to implement the speed term given in (4), but rely on a *multiphase extension* of the basic formulation to enable multiple regions [17, 28]. We now let $\{R_k\}_{k=1}^{K}$ be a set of level-set functions. The segmentation for texture k is then defined as $T_k = \{t \in T | R_k(t) > R_j(t), \forall j \neq k\}$. The level set speed term for minimizing the energy defined in (3) is therefore

$$\frac{\partial R_k(t)}{\partial \tau} = p_k(z(t)) \log p_k(z(t)) + \alpha \nabla \cdot \left(\frac{\nabla R_k(t)}{\parallel \nabla R_k(t) \parallel} \right), \qquad (4)$$

where τ denotes the time evolution variable.

In accordance with Esedoglu and Tsai's scheme [10], given an initialization for texture regions $\{R_k\}_{k=1}^{K}$, the algorithm iterates over the following steps:

1. Estimate $p_k(z(t))$
2. $R'_k = R_k^m + \beta p_k(z(t)) \log p_k(z(t))$
3. $R''_k = K_\epsilon \otimes R'_k$, where K_ϵ is a Gaussian kernel with a standard deviation ϵ and $\otimes$ denotes convolution.
4. Set $R_k^{m+1}(t) = 1$ if $R_k(t) > R_j(t)$ for all $j \neq k$, set $R_k^{m+1}(t) = 0$ otherwise.
5. Stop when the change in the segmentation, i.e. the sets $\{R_k\}$, falls below a small threshold.

For a discussion of the relationship of the new parameters, namely ϵ and β, to the parameter α in the traditional level set framework, we refer the reader to [10]. For this work, the critical problem lies in the estimation of $p_k(z(t))$, which is addressed in the next section. These updates represent an approximation to a gradient descent algorithm on the image entropy. We use an initial segmentation of randomly generated regions and segment the image in an unsupervised manner, as shown in Section 4.

3.4. Nonparametric Density Estimation

Entropy optimization entails the estimation of higher-order conditional PDFs. This introduces the challenge of high-dimensional, scattered-data interpolation, even for modest sized image neighborhoods. High-dimensional spaces are notoriously challenging for data analysis (regarded as the *the curse of dimensionality* [26, 25]), because they are so sparsely populated. Despite theoretical arguments suggesting that density estimation beyond a few dimensions is impractical, the empirical evidence from the literature is more optimistic [25, 21]. The results in this paper confirm that observation. Furthermore, stationarity implies that the random vector Z exhibits identical marginal PDFs, and thereby lends itself to more accurate density estimates [25, 26]. We also rely on the neighborhoods in natural images having a lower-dimensional topology in the multi-dimensional feature space [16, 5]. Therefore, *locally* (in the feature space) the PDFs of images are lower dimensional entities that lend themselves to better density estimation.

We use the Parzen-window nonparametric density estimation technique [8] with an n-dimensional Gaussian kernel $G_n(\tilde{z}, \Psi_n)$. We have no a priori information on the structure of the PDFs, and therefore we choose an isotropic Gaussian, i.e. $\Psi_n = \sigma I_n$, where I_n is the $n \times n$ identity matrix. For a stationary ergodic process, the multivariate Parzen-window estimate is

$$p_k(Z = z) \approx \frac{1}{|A_k|} \sum_{t_j \in A_k} G_n(z - z_j, \Psi_n) \qquad (5)$$

where the set A_k is a small subset of T_k, chosen randomly and z_j is shorthand for $z(t_j)$.

Using optimal values of the Parzen-window parameters is critical for success, and that can be especially difficult in such high-dimensional spaces. The best choice depends on a variety of factors including the sample size $|A_k|$ and the natural variability in the data. To address this problem we have developed a method for automatically choosing optimal value of this parameter [1]. We choose σ to minimize the entropy of the associated PDF via a Newton-Raphson optimization scheme. This entropy minimizing choice for σ is consistent with the entropy minimization segmentation formulation. Our experiments show that for sufficiently large $|A_k|$ additional samples do not significantly affect the estimates of entropy and σ, and thus $|A_k|$ can also be generated automatically from the input data.

The quality of the results also depends on the neighborhood size—bigger neighborhoods are generally more effective but take longer to compute. Typically 9×9 neighborhoods suffice. To obtain rotational invariance we use a metric in the feature space that controls the influence of each neighborhood pixel so that the distances in this space are less sensitive to neighborhood rotations [1]. In this way, feature space dimensions corresponding to corners of the square neighborhood collapse so that they do not influence the filtering. Likewise image boundaries are handled through such anisotropic metrics so that they do not distort the neighborhood statistics of the image.

4. Results

In this section, we discuss results from experiments with real and artificial data. For the level-set initialization we used a checkerboard image, with K different labeled checks. The number K is a user parameter and should be chosen to match the desired number of texture classes. Figure 1(a) is an electron microscopy image of a cell. This image is challenging to segment using edge or intensity information because it is not piecewise homogeneous. The discriminating feature of these two cell types (type A: upper left and bottom right, type B: middle) is their textures formed by the arrangements of sub-cellular structures. Figure 1(b) illustrates a checkerboard pattern with two classes used to initialize the algorithm. Figure 1(c) shows the result of the proposed algorithm starting from this initialization. The two cell types are segmented with a high degree of accuracy; however, notice that the membranes between the cells are grouped together with the middle cell. A third texture class could be used for the membrane, but this is not a trivial extension due to the thin, elongated geometric structure of the membrane and the difficulties of sampling associated with such structures. Figure 1(d) and (e) show another initialization with a finer scale checkerboard pattern and the segmentation results, respectively. The final segmentation is almost the same as before demonstrating the robustness of the algorithm to initializations.

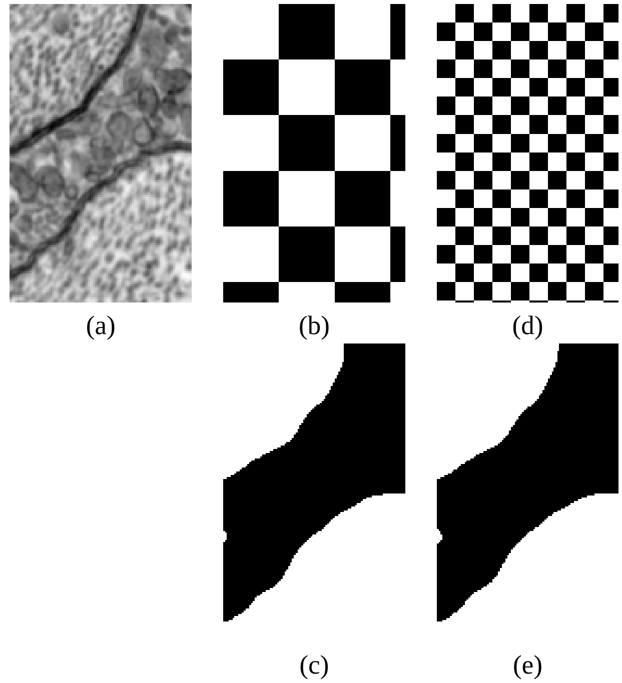
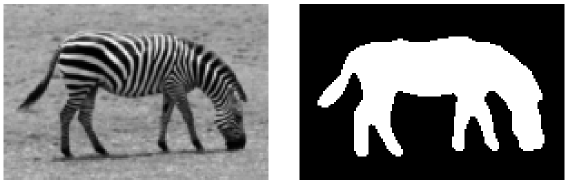


Figure 1: (a) An electron microscope image of a cell, (b) initialization with checkerboard pattern, (c) result of 2-class segmentation, (d) different initialization with smaller checkerboard pattern, and (e) result.

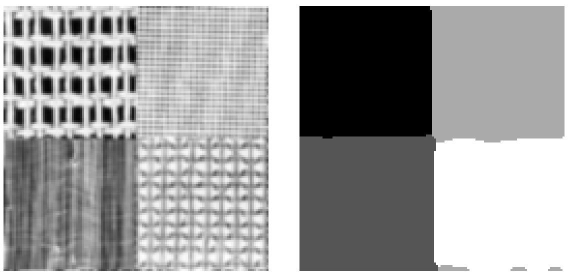
Figure 2(a) is a kind of image often used in the texture segmentation literature. Figure 2(b) demonstrates a successful segmentation of this image using the proposed algorithm.

Figure 3 depicts an image having multiple classes and a successful segmentation with the proposed algorithm using the multiphase level-set framework.



(a) (b)

Figure 2: (a) Zebra image and (b) final segmentation.



(a) (b)

Figure 3: (a) An image of 4 different Brodatz textures. (b) A segmentation into 4 regions using the multiphase version of the level-set algorithm.

5. Conclusions

We have presented a novel approach toward texture segmentation that exploits higher-order image statistics in an entropy-minimizing framework. The method automatically learns the image statistics via nonparametric density estimation and, unlike typical prevalent techniques, does not impose an adhoc image model. We incorporate an efficient multiphase level-set evolution framework [10] to obtain an optimal segmentation. The method relies on the information content of input data for setting important parameters, and does not require significant parameter tuning. Hence it is easily applicable to a wide spectrum of images.

The computational complexity of the proposed method is significant: $O(|T||A_k|E^D)$ where D is the image dimension and E is the extent of the neighborhood along a dimension. This is exponential in E, and our current results are limited to 2D images. The literature suggests some potential improvements (e.g. [31]). However, the purpose of this paper is to introduce the theory and methodology—algorithmic improvements are the subject of future work.

References

[1] S. Awate and R. Whitaker. Higher-order image statistics for unsupervised, information-theoretic, adaptive, image filtering. *To appear in Proc. IEEE Int. Conf. Computer Vision Pattern Recog. 2005*.

[2] J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE PAMI*, 13(8):775–790, 1991.

[3] J. D. Bonet and P. Viola. A non-parametric multi-scale statistical model for natural images. In *Proc. of the 1997 conf. on Advances in Neural Information Processing Systems*, pages 773–779. MIT Press, 1998.

[4] J. S. de Bonet and P. Viola. Texture recognition using a non-parametric multi-scale statistical model. In *CVPR '98: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, page 641. IEEE Computer Society, 1998.

[5] V. de Silva and G. Carlsson. Topological estimation using witness complexes. *Symposium on Point-Based Graphics*, 2004.

[6] G. Doretto, D. Cremers, P. Favaro, and S. Soatto. Dynamic texture segmentation. In *Proc. of Int. Conf. on Computer Vision (ICCV'03)*, 2003.

[7] E. R. Dougherty. *Random Processes for Image and Signal Processing*. Wiley, 1998.

[8] R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification*. Wiley, 2001.

[9] A. A. Efros and T. K. Leung. Texture synthesis by non-parametric sampling. In *Int. Conf. Computer Vision*, page 1033, 1999.

[10] S. esedoglu and Y.-H. R. Tsai. Threshold dynamics for the piecewise constant mumford-shah functional. Technical Report CAM-04-63, Dept. Mathematics, UCLA, 2004.

[11] D. Gabor. Theory of communication. *J. IEEE*, 93:429–459, 1946.

[12] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions and the bayesian restoration of images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6:721–741, 1984.

[13] J. Huang and D. Mumford. Statistics of natural images and models. In *Int. Conf. Comp. Vision*, pages 541–547, 1999.

[14] T. Kadir and M. Brady. Unsupervised non-parametric region segmentation using level sets. In *Proc. of IEEE Int. Conference on Computer Vision*, pages 1267–1274, 2003.

[15] J. Kim, J. W. F. III, A. Y. Jr., M. Cetin, and A. S. Willsky. Nonparametric methods for image segmentation using information theory and curve evolution. In *Proc. of IEEE Int. Conference on Image Processing*, volume III, pages 797–800, 2002.

[16] A. Lee, K. Pedersen, and D. Mumford. The nonlinear statistics of high-contrast patches in natural images. *Int. J. Comput. Vision*, 54(1-3):83–103, 2003.

[17] B. Merriman, J. K. Bence, and S. Osher. Motion of multiple junctions: A level set approach,. Technical Report CAM-93-19, Dept. Mathematics, UCLA, 1993.

[18] D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Com. of Pure and Applied Math.*, 42:577–685, 1989.

[19] S. Osher and J. Sethian. Fronts propagating with curvature-dependent speed: algorithms based on hamilton-jacobi formulations. *Journal of Computational Physics*, 79(1):12–49, 1988.

[20] N. Paragios and R. Deriche. Geodesic active regions and level set methods for supervised texture segmentation. *IJCV*, 46(3), 2002.

[21] K. Popat and R. Picard. Cluster based probability model and its application to image and texture processing. *IEEE Trans. Image Processing*, 6(2):268–284, 1997.

[22] M. Rousson, T. Brox, and R. Deriche. Active unsupervised texture segmentation on a diffusion based feature space. In *CVPR'03: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 699–706. IEEE Computer Society, 2003.

[23] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi. Texture segmentation via a diffusion segmentation scheme in the gabor feature space. In *2nd Int. Workshop on Texture Analysis and Synthesis, Proc. Texture 2002*, 2002.

[24] B. Sandberg, T. Chan, and L. Vese. A level-set and gabor-based active contour algorithm for segmenting textured images. Technical Report CAM-02-39, Dept. Mathematics, UCLA, 2002.

[25] D. W. Scott. *Multivariate Density Estimation*. Wiley, 1992.

[26] B. Silverman. *Density Estimation for Statistics and Data Analysis*. Chapman and Hall, 1986.

[27] R. van den Boomgaard and J. van de Weijer. Robust estimation of orientation for texture analysis. In *2nd Int. Workshop on Texture Analysis and Synthesis, Proc. Texture 2002*, 2002.

[28] L. Vese and T. Chan. A multiphase level set framework for image segmentation using the mumford and shah model. Technical Report CAM-01-25, Dept. Mathematics, UCLA, 2001.

[29] L. Wei and M. Levoy. Order-independent texture synthesis. *Stanford University Computer Science Department Tech. Report TR-2002-01*, 2002.

[30] J. Weickert. Coherence-enhancing diffusion filtering. *Int. J. Computer Vision*, 31:111–127, April 1999.

[31] Yang, Duraiswami, Gumerov, and Davis. Improved fast gauss transform and efficient kernel density estimation. In *Int. Conf. Comp. Vision*, pages 464–471, 2003.