Mean-Shift Segmentation with Wavelet-based Bandwidth Selection

Maneesh K. Singh* and Narendra Ahuja
Beckman Institute and ECE Department
University of Illinois at Urbana-Champaign
{msingh,n-ahuja}@uiuc.edu

Abstract

Recently, various non-linear techniques for segmentation have been proposed based on non-parametric density estimation. These approaches model image data as clusters of pixels in the combined range-domain space, using kernel based techniques to represent the underlying, multi-modal Probability Density Function (PDF). In Mean-shift based segmentation, pixel clusters or image segments are identified with unique modes of the multi-modal PDF by mapping each pixel to a mode using a convergent, iterative process. The advantages of such approaches include flexible modeling of the image and noise processes and consequent robustness in segmentation. An important issue is the automatic selection of scale parameters - a problem far from satisfactorily addressed. In this paper, we propose a regression-based model which admits a realistic framework to choose scale parameters. Results on real images are presented.

1 Introduction

A popular segmentation framework is to model image data as clusters of pixels in the combined range-domain space [1, 3]. If the image pixels are assumed to be drawn independently from an underlying multi-modal probability density function (PDF) and different modes of the PDF are identified with different segments, the segmentation algorithm would just need to identify each pixel with a unique mode. Mean-shift procedure (proposed by Fukunaga [5]) iteratively shifts each pixel to it's respective mode. In [3], Comaniciu and Meer analyzed the properties of the mean-shift algorithm and proved its convergence for a specific class of kernels (that includes the Gaussian and Epanechnikov kernels). The algorithm is non-linear but simple, fast and gives visually good results. Due to the underlying flexible model, it can be applied to a variety of images and noise processes.

Our paper is based on two related observations: Firstly ,the knowledge that images are functions defined on the spatial domain is not used by Comaniciu and Meer [3]. Secondly, an automatic and appropriate choice of scale parameters is a problem that is far from satisfactorily addressed. We propose to address both the problems by modeling the image in the wavelet domain - this leads to a good choice of scale parameters and also admits a dyadic multi-scale segmentation framework.

In the next Section, we present the motivation for the proposed model to be used for segmentation. In Section 3, we present the kernel-based density estimator for the proposed model and derive expressions for asymptotically optimal scale parameters. In Section 4, we present our algorithms for scale selection and the consequent mean-shift segmentation process. In Section 5, we present results on real images and round off with discussions about future work in Section 6.

2 Model

In [3], the image is modeled in joint range-domain space; image pixels are assumed to be drawn independently from an underlying PDF model defined on the joint space. The mean-shift procedure shifts each pixel to one of the modes of the underlying PDF. Naturally, the parameters for the mean-shift algorithm need to be selected for an optimal estimate of the underlying PDF. The asymptotically optimal scale parameters, in the Mean Integrated Square Error (MISE) sense, depend upon the yet-to-be-estimated PDF, or at least it's average properties. One popular approach (for it's relative simplicity) is to use a plug-in estimate, where optimal bandwidth parameters are found with respect to a (plug-in) family of PDFs like the multivariate Gaussian. In the statistics domain, several other methods (more sophisticated and computationally intensive) have been suggested but plug-in estimates remain an attractive choice. We refer the reader to [7, 8] for further details.

Image data is highly non-Gaussian - as an example, we consider the Sailboat image in Figure 2(a) and depict the marginal PDF for the intensity variable (Fig-

*Send correspondence to first author

Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision (WACV'02)
0-7695-1858-3/02 $17.00 © 2002 IEEE

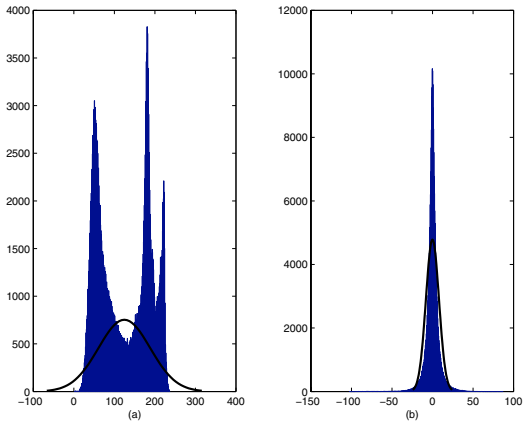

Figure 1: *For Sailboat image, marginal histogram of (a) raw intensity values, and, (b) residuals of LL projection using the db2 wavelet.*

ure 1(a)). Evidently, we should be considering the joint PDF instead of the marginal PDF since the data is not spatially stationary. Even so, this example and the histogram is typical of the usual image data and suffices for the current discussion. Non-gaussianity is evidently the norm, reflected in the fact that the image PDF is often estimated to be a multi-modal density function. Any estimate based on the Gaussian plug-in is likely to yield unsatisfactory results. Comaniciu and Meer [4] propose an empirical estimate using a locally Gaussian assumption.

In this paper, we take a different approach. We first note that in our domain of application, images, we are dealing with functions (scalar, or vector as in color images) defined on the 2-D spatial domain. Hence, the data should be modeled as such. Excellent non-parametric frameworks, like wavelets, are already available. Good approximations can be generated for the underlying regression function using these representations. Secondly, deviations from the estimated regression function can be used to develop a model for the noise process. If the first observation is valid, then the noise model should be more accurate and the noise PDF easier to estimate.

As an example, consider Figure 1(b), where we plot the (intensity) histogram of the residual image when the Sailboat image is projected onto the LL sub-band using Daubechies' db2 wavelet. Clearly, the residuals can be more easily and realistically modeled than the complete Sailboat image in the spatial domain (Figure 1(a)). It has been shown that the Generalized Gaussian Distribution (GGD) is a good model for the residuals (in Figure 1(b)) [6].

Hence, we model the image data as,

$$I(\mathbf{t}) = r(\mathbf{t}) + \epsilon(\mathbf{t}) \tag{1}$$

where $\mathbf{t}$ is a vector representing the spatial location, $I(\cdot)$, and $\epsilon(\cdot)$ are noisy image and the additive noise respectively while $r(\cdot)$ is the *clean* image irradiance function that we seek to model using the regression framework. We note that for parameter estimation purposes, it is sufficient if $r(\cdot)$ can be *approximated* by the regression model and likewise, the residuals can be approximated by GGD. In other words, the regression need not yield a piecewise smooth curve separated by jumps at the segment boundaries.

In the next Section, we propose a kernel-based density estimator for the above model and derive expressions for asymptotically optimal scale parameters.

3 Analysis

Let us define a 3-tuple $\mathbf{z} = [v, \mathbf{t}^T]^T = [v, x, y]^T \in \mathcal{R}^3$. Then, we define a kernel based estimator for the conditional PDF of I given the spatial location $\mathbf{t} = [x, y]^T$ as,

$$\hat{f}_{I|\mathbf{t}}(v|\mathbf{t}) = \frac{1}{m^2|H|\mathcal{D}} \sum_{i=1}^{m}\sum_{j=1}^{m} K(H^{-1}(\mathbf{z}_{ij} - \mathbf{z})) \tag{2}$$

where H is a non-singular 3×3 bandwidth matrix and $K : \mathcal{R}^3 \to \mathcal{R}$ is a kernel such that it is non-negative, has a unit area ($\int_{\mathcal{R}^3} K(\mathbf{z})\, d\mathbf{z} = 1$), zero mean ($\int_{\mathcal{R}^3} \mathbf{z}K(\mathbf{z})\, d\mathbf{z} = 0$), and, unit covariance ($\int_{\mathcal{R}^3} \mathbf{z}\mathbf{z}^T K(\mathbf{z})\, d\mathbf{z} = I_3$). $\mathcal{D} = \frac{1}{m^2|H|}\sum_{i=1}^{m}\sum_{j=1}^{m}\int_{\mathcal{R}} K(H^{-1}(\mathbf{z}_{ij} - \mathbf{z}))\, dI$ can be treated as a normalization constant.

The non-linear nature of the estimate does not permit an exact analysis. To carry out an asymptotic analysis, we assume that the number of samples, $n = m^2$, tends to infinity via a successive refinement of the sampling grid. Consequently, $\{r(\mathbf{t}_{ij})\}$ represent the underlying function $r(\mathbf{t})$ more and more accurately. We assume that the noise samples are independent and identically distributed irrespective of the grid size. For the ease of bandwidth estimation, we also assume that $H = \mathrm{diag}(h_I, h_x, h_y)$.

Imposing the condition that $\|\mathrm{diag}(h_I, h_x, h_y)\| \to 0$ and $n\det(H) \to \infty$ as $n \to \infty$, the estimate $\hat{f}(\cdot)$ can be made consistent i.e. the asymptotic integrated mean square error (AIMSE) goes to zero as the number of samples approaches infinity (via refinement of the grid). It can be verified that,

$$\mathrm{AIMSE} = \int \mathrm{bias}^2(v, x, y) + \int \mathrm{var}(v, x, y) =$$

$$\begin{pmatrix} h_I^2 \\ h_x^2 \\ h_y^2 \end{pmatrix}^T \begin{pmatrix} \|a\|^2 & < a,b > & < a,c > \\ < a,b > & \|b\|^2 & < b,c > \\ < a,c > & < b,c > & \|c\|^2 \end{pmatrix} \begin{pmatrix} h_I^2 \\ h_x^2 \\ h_y^2 \end{pmatrix} + \frac{\Delta x \Delta y R(K)}{n h_I h_x h_y} \quad (3)$$

where $R(g) \triangleq \int g(\mathbf{z})^2 \, d\mathbf{z}$ and,

$$\begin{aligned} 2a(v,x,y) &\triangleq \partial_\epsilon^2 f_{\epsilon|x,y}(v - r(x,y)|x,y) \\ 2b(v,x,y) &\triangleq \partial_x^2 f_{\epsilon|x,y}(v - r(x,y)|x,y) \\ &= \partial_\epsilon^2 f \cdot (\partial_x r)^2 - 2(\partial_\epsilon \partial_x f) \cdot \partial_x r - \partial_\epsilon f \cdot \partial_x^2 r + \partial_x^2 f \\ 2c(v,x,y) &\triangleq \partial_y^2 f_{\epsilon|x,y}(v - r(x,y)|x,y) \\ &= \partial_\epsilon^2 f \cdot (\partial_y r)^2 - 2(\partial_\epsilon \partial_y f) \cdot \partial_y r - \partial_\epsilon f \cdot \partial_y^2 r + \partial_y^2 f \quad (4) \end{aligned}$$

Optimal bandwidth matrix is sought by minimizing the above expression, which is a non-trivial problem. Hence, we suggest an upper-bound for AIMSE and compute parameters to minimize the upper bound. By applying Cauchy-Schwartz, we get,

$$\begin{aligned} \text{AIMSE} &\leq (h_I^2\|a\| + h_x^2\|b\| + h_y^2\|c\|)^2 \\ &\quad + \frac{\Delta x \Delta y R(K)}{n h_I h_x h_y} \quad (5) \end{aligned}$$

The bandwidth parameters that minimize the upper bound in Equation (5) are,

$$\begin{aligned} h_I^\star(v,x,y) &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|b\| \, \|c\|}}{12\|a\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \\ h_x^\star(v,x,y) &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \, \|c\|}}{12\|b\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \\ h_y^\star(v,x,y) &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \, \|b\|}}{12\|c\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \end{aligned}$$

and consequently,

$$\begin{aligned} &\text{AIMSE}^\star(v,x,y) \\ &\leq 21 \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \, \|b\| \, \|c\|}}{12} \right]^{\frac{4}{7}} n^{-\frac{4}{7}} \quad (6) \end{aligned}$$

This particular choice of bandwidth bounds the error, which goes to zero at the optimal rate as the number of samples increase to ∞.

4 Scale-guided Segmentation

Humans view signals and the information they convey at various scales - but not simultaneously. When the signal is processed to reveal information at a certain scale, analysis at a larger scale is done, and signal information at smaller scales is viewed as finer details or noise for the analysis at the chosen scale. We adopt this philosophy for the purpose of segmentation.

Thus, at any spatial resolution, an estimate of the regression function is obtained, as also an estimate of the noise realization. From these estimates, we obtain the bandwidth parameters for the kernel-based PDF estimator in Equation (2). Mean-shift procedure is used to map each pixel to the mode of the estimated multi-modal PDF, thereby yielding the transformed data. This transformed data is then segmented. Below, we give the proposed algorithm.

4.1 Algorithm

1. *Regression Estimate*: Let $\theta(\mathbf{t})$ be a *smoothing function* (integral equal to 1 and converges to 0 at infinity). We denote the smoothing function at scale s as $\theta_s(\mathbf{t}) = \frac{1}{s}\theta(\frac{\mathbf{t}}{s})$. Then, we approximate the smoothed image by $\hat{r}_s(\mathbf{t}) = I_s(\mathbf{t}) = I * \theta_s(\mathbf{t})$. The derivative of the regression estimate are given by convolution of $I(\mathbf{t})$ with wavelets that are components of $\nabla_{\mathbf{t}}\theta_s(\mathbf{t})$ and $\nabla_{\mathbf{t}}\nabla_{\mathbf{t}}^T\theta_s(\mathbf{t})$. As an example, we take $\theta_s(\mathbf{t}) = \mathcal{N}(0, s^2 I)$, the symmetric Gaussian function.

2. *Bandwidth Estimation*: We estimate the noise at scale s by $\hat{\epsilon}_s(\mathbf{t}) = I(\mathbf{t}) - \hat{r}_s(\mathbf{t}) = I * (\delta - \theta_s)(\mathbf{t})$. It has been noted (and as is evident from Figure (1)) that the difference signal can be modeled as a Generalized Gaussian Distribution (GGD). GGD is a parameterized family of distributions. The functions ($a(\cdot)$, $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$) are easily computed in terms of these parameters as explained in Section 4.2.

3. *Mean Shift*: We consider kernels such that $K(x) = k(\|x\|^2)$ where $k()$ is convex. Then, defining a transformation $M : \mathcal{R}^3 \to \mathcal{R}^3$ such that for any $\mathbf{p} = [v,x,y]^T$,

$$M(\mathbf{p}) = \frac{\sum_{i,j} \mathbf{z}_{ij} k'(\|H^{-1}(\mathbf{z}_{ij} - \mathbf{p})\|^2)}{\sum_{i,j} k'(\|H^{-1}(\mathbf{z}_{ij} - \mathbf{p})\|^2)} \quad (7)$$

It follows from (Theorem 2 in [3]) that the sequence $\{\mathbf{p}\}_k$ defined by $\mathbf{p}_{k+1} = M(\mathbf{p}_k)$ in Equation (7) converges to a local mode of the PDF defined in Equation (2). Thus, the mean-shift

(a) (b)

Figure 2: *(a) Sailboat image: size* 512×512 *(b) segmented image with overlayed boundaries* $(h_I, h_x, h_y, \gamma) = (6.2, 8, 8, 64)$

process when applied to each image pixel, maps it to it's corresponding mode.

4. *Segmentation*: Edges between two pixels are detected if the normalized distance between the two pixels $\mathbf{p}_i$ and $\mathbf{p}_j$, $H^{-\frac{1}{2}}(M(\mathbf{p}_i) - M(\mathbf{p}_j)) > \frac{1}{2}$. Eventually and optionally, we discard small regions of size less than γ. A reasonable choice for $\gamma = h_x^\star \times h_y^\star$.

4.2 Bandwidth Estimation

To estimate the bandwidths in Equations (6), we need to compute $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$.

- $a(\cdot)$: GGD for ϵ is given by, $f_\epsilon(v) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \exp(-|v|/\alpha)^\beta$ where parameters α and β model the variance and shape of the distribution ($\beta = 1, 2$ give double-sided exponential and Normal distributions respectively). These parameters [6] and $a(\cdot)$ are computed directly using the moments of the histogram of residuals.

- $b(\cdot)$ and $c(\cdot)$: In this paper we assume that the noise is i.i.d. and is independent of the signal. Hence, in the expressions for $b(\cdot)$ and $c(\cdot)$ in Equations (4), second and fourth terms reduce to zero. Further, the third term is negligible (confirmed experimentally) as compared to the first. Hence, $b(v, x, y) \approx a(v)(I(x, y) \star \nabla_x \theta_s(x, y))^2$ and $c(v, x, y) \approx a(v)(I(x, y) \star \nabla_y \theta_s(x, y))^2$.

5 Results

We present results on two real images. Depending upon the resolution (spatial scale) at which the regression function is estimated, the parameters take different values. This is akin to choosing the resolution at which the image is viewed. In the MRA framework, we can choose dyadic spatial scales and compute corresponding bandwidths.

To compute the bandwidths, we use robust estimators. Thus, we use mean absolute deviation (MAD) instead of standard deviation for computing α. More importantly, to compute the mean of the absolute values of derivatives (to fourth powers) required in computing $\|b\|$ and $\|c\|$, we only use values up to percentile 75. We do so as we do not want large edges to significantly alter the mean image gradient and curvature information.

We present estimated bandwidths for the two images in Table (1) and segmentation results for $s = 4$ in Figures (2) and (3). The segmentation results are obtained by using the EDISON software [2]. The parameters closely match those obtained by hand-selection (for best visual results) in [3]. Since the EDISON software assumes $h_x = h_y$, we choose the largest of the two bandwidths for our results. This has an insignificant influence on the quality of the results.

6 Discussions

We note that by modeling the image as proposed in this paper, we can design algorithms to automatically compute bandwidths for the mean-shift algo-

(a) (b)

Figure 3: *(a) Peppers image: size* 512×512 *(b) segmented image with overlayed boundaries* $(h_I, h_x, h_y, \gamma) = (4.8, 7, 7, 49)$

	Sailboat			Peppers		
s	h_I	h_x	h_y	h_I	h_x	h_y
1	4.4	2.9	2.9	3.1	3.0	3.5
2	5.2	4.5	4.6	3.7	3.9	4.9
4	6.2	7.5	7.7	4.8	5.6	7.1
16	7.1	13. 5	13.0	6.2	9.3	11. 6
32	7.6	24. 1	25.1	7.4	17.7	21. 6

Table 1: *Estimated bandwidth parameters*

rithm. There are two obvious directions in which the authors are extending this work. Firstly, in evolving adaptive bandwidth selection schemes that are sensitive to local image information (like local noise variance (heteroscedastic case) and local image gradients). Secondly, in developing a multi-scale segmentation algorithm that determines and links structurally relevant image segmentations at all image resolutions.

References

[1] N. Ahuja. A transform for multiscale image segmentation by integrated edge and region detection. *P AMI, IEEE Trans.*, 18(12):1211–1235, 1996.

[2] C. M. Christoudias, B. Georgescu, and P . Meer. Synergism in low level vision. *ICPR, 16th Intl Conf.*, accepted, 2002.

[3] D. Comaniciu and P. Meer. Mean shift: a robust approach tow ardfeature space analysis. *PAMI, IEEE Trans*, 24(5):603–619, May 2002.

[4] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. *Computer Vision, Eighth Intl. Conf. on*, pages 438–445, 2001.

[5] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Info. Theory, IEEE T ans.*, 21:32–40, 1975.

[6] S. G. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *PAMI, IEEE Trans.*, 11(7):674–693, July 1989.

[7] D. W. Scott. *Multivatiate density estimation: theory, pr actic e and visualizaion* Wiley-Interscience, 1992.

[8] J. S. Simonoff. *Smoothing methods in statistics.* Springer, 1996.