

Regression based Bandwidth Selection for Segmentation using Parzen Windows

Maneesh Singh

ECE Department
University of Illinois
Urbana-Champaign, IL 61801
msingh@uiuc.edu

Narendra Ahuja

ECE Department
University of Illinois
Urbana-Champaign, IL 61801
n-ahuja@uiuc.edu

Abstract

We consider the problem of segmentation of images that can be modelled as piecewise continuous signals having unknown, non-stationary statistics. We propose a solution to this problem which first uses a regression framework to estimate the image PDF, and then mean-shift to find the modes of this PDF. The segmentation follows from mode identification wherein pixel clusters or image segments are identified with unique modes of the multi-modal PDF. Each pixel is mapped to a mode using a convergent, iterative process. The effectiveness of the approach depends upon the accuracy of the (implicit) estimate of the underlying multi-modal density function and thus on the bandwidth parameters used for its estimate using Parzen windows. Automatic selection of bandwidth parameters is a desired feature of the algorithm. We show that the proposed regression-based model admits a realistic framework to automatically choose bandwidth parameters which minimizes a global error criterion. We validate the theory presented with results on real images.

1. Introduction

In this paper, we consider the problem of segmentation of images that can be modelled as piecewise continuous signals having unknown, non-stationary statistics. Non-stationarity in the image can be modelled by the conditional density function $f(I|x, y)$ that varies with the location (x, y) . When $f(I|x, y)$ is not known, it can be estimated from the image data using Parzen windows (henceforth called kernels). The way the probability density function (PDF) of the image is related to segmentation of the image data is via the observation that the image PDF is multi-modal and that each mode of the PDF is associated with a cluster of pixels which forms a homogeneous (having the same central tendency, i.e., the mode) segment in the image. This model for the image data, as clusters of pixels in the combined range-domain (for e.g., Intensity-spatial) space, has been proposed before (see [1, 5] for recent papers).

For segmenting the image, it is not sufficient to estimate the underlying PDF and its modes. Also required is an algorithm that associates each image pixel with the mode associated with the segment it belongs to. If kernel-based PDF estimators are used, under mild conditions¹ on the kernels, the mean-shift algorithm (proposed by Fukunaga and Hostetler [7]) can be used to achieve this goal: it iteratively shifts each pixel to its respective mode. In [5], 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 iterative but simple, fast and gives visually good results. Consequently, mean-shift algorithm (based on the kernel-based PDF estimation framework) can be used to segment non-stationary image signals.

The main challenge of this approach is the estimation of bandwidth (scale) parameters associated with the estimating kernels. There are several approaches listed in the literature - these include (A) subjective choice (page 6, [5]); using asymptotic mean square criterion (B1) for a reference distribution, called plug-in estimators (page 86, [10]); (B2) data-driven cross-validation schemes (pages 160-180, [9]), and other (C) completely data-driven techniques (see for excellent discussion, Section 3.1, [5] and [3, 11] for two different techniques). These techniques evolve in complexity and reduce in domain-specific knowledge from (A) to (C) - this is the inherent tradeoff involved. Further, these schemes differ in estimating global and locally-adaptive bandwidths - locally-adaptive bandwidths perform better but involve a higher complexity.

We make two observations here: (1) The advantage of mean-shift approach is its simplicity and hence one would like as simple a bandwidth estimation technique as possible, and (2) All the bandwidth-selection methods listed above are for generic data distributions. However, image data is highly structured - in fact, it is a function (scalar or vector) from the domain of definition to intensity or color space. Further, as we show in Section 2, real images are seen to

¹see Section 4.1, Step 4

have data distributions that can justify a plug-in estimator. Thus, our scheme fits into category (B1) above. We exploit this fact to propose a novel bandwidth selection scheme in this paper. The proposed scheme is simple and fast ($O(nk)$ where k is the size of the spatial neighborhood ($k \ll n$).

Following are the main steps in our algorithm: (A) We model the image data using a regression framework; (B) Then, based on the spatial resolution at which segmentation is desired, we select a regression estimator; (C) The residues of the regression estimator are modelled using a parametric distribution which yields the bandwidth parameters for kernels; (D) These bandwidth parameters are used to estimate the underlying PDF that produced the image data and to apply the mean-shift algorithm to produce the end segmentation.

Following are the main contributions of this paper: We present an automatic bandwidth selection procedure for mean-shift based image segmentation. This selection procedure is fast and easy to compute. We show that the bandwidths selected are consistent and the particular choice of bandwidths bounds the asymptotic mean square error (for the estimated image PDF) which asymptotically goes to zero at the optimal rate². The selection procedure gives a central role to the spatial resolution at which the image is analyzed. Further, we validate the bandwidth selection procedure with results on real data.

This paper is organized as follows: In Section 2, we present the image model and its justification; then, in Section 3, we present the kernel based density estimator for the image model and derive expressions for the regression-based bandwidth parameters. In Section 4, we present the algorithm for bandwidth estimation and the consequent mean-shift segmentation process. In Section 5, we present results on real images. Finally, in Section 6, we present our concluding remarks.

2. Model

In [5], images are modelled using a PDF distribution defined on the joint domain-range space and pixels are assumed to be drawn independently from this distribution. 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 its average properties.

As discussed in Section 1, several methods have been suggested in the literature (refer to pages 160-180, [9] and [3, 5, 6] for details). However, plug-in estimators, where optimal bandwidth parameters are found with respect to a

²Refer to Equation (9) and the following reference.

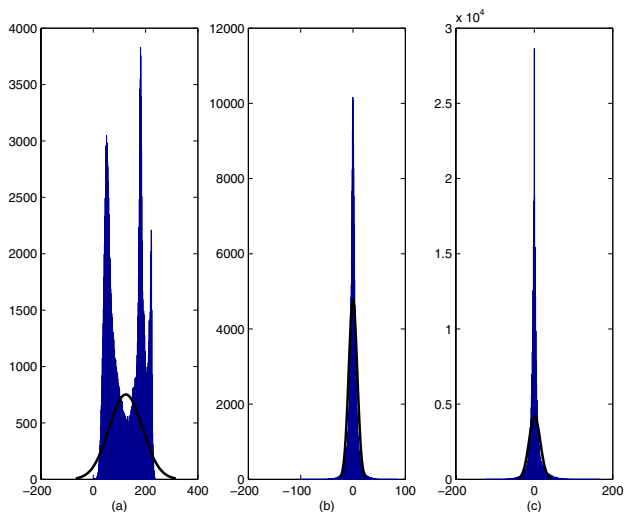


Figure 1: This figure shows that image data does not have a simple distribution while residues after regression do. For Sailboat image 4(a), size 512×512 , marginal histogram of (a) raw intensity values; (b) residuals of LL projection using db2 wavelet; (c) residuals after 7×7 median filtering.

(plug-in) family of PDF's like the multivariate Gaussian, remain popular (because of their simplicity). Hitherto, only multivariate Gaussian was used as a plug-in estimator, either globally [4] or locally [6].

However, image data is highly non-Gaussian. Consider, for example, the Sailboat image in Figure 4(a). Figure 1(a) shows the marginal PDF for the intensity variable along with the best-fit Gaussian curve which is indeed a poor fit. In fact, the data cannot be easily modelled using parametric distributions. Contrast this with Figures 1(b) and (c) where the distribution is highly structured and can be modelled through a Generalized Gaussian Distribution (GGD)³. In these images we plot the histogram of the spatial residuals when the Sailboat image is projected onto the "LL" sub-band using Daubechies' db2 wavelet (1(b)), and spatial residues after median filtering the image (1(c)), respectively.

The fact that GGD is a good model for residues in Figure 1(b) is already shown in [8]. We empirically observed that it is true for median filtering as well. This shows that plug-in estimators can be used for a wide class of regression residuals, though not (at least globally) for images themselves. This brings us to our second observation: Since images are defined on a lattice, hence $f(I, x, y) = f(I|x, y)f(x, y) = c \times f(I|x, y)$ for some constant c . In other words, the probability of finding a pixel in the x - y plane is constant over the domain of definition. Thus, only the estimation of $f(I|x, y)$ is required.

³refer to first paragraph, Section 4.2.

In brief, we make the following observations: (1) One just needs to model and estimate $f(I|x, y)$, (2) good estimators like wavelets and median filters are available for regression estimates (denoted by $r(x, y)$) for an image, and (3) the residuals ($\epsilon(x, y) = I(x, y) - r(x, y)$) can be modelled using a plug-in estimator. This suggests that rather than modelling the image data in the joint domain-range space (as in [3-6]), one might model it specifically as a function defined over a domain. Thus, we use the following image model where $\mathbf{t} = (x, y)$ is the spatial location:

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

It is worth noting that for bandwidth 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 - thus, it is not a chicken-egg problem.

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

In this section, we introduce the kernel estimator for $f(I|x, y)$ and analyze its asymptotic MSE (mean square error) properties. We do this to derive optimal bandwidth estimates which are critical for segmentation using kernel density estimation.

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

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

where H is a non-singular 3×3 bandwidth matrix and $K : \mathcal{R}^3 \rightarrow \mathcal{R}$ 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 ($\int_{\mathcal{R}^3} \mathbf{z}\mathbf{z}^T K(\mathbf{z}) d\mathbf{z} = I_3$). For data defined on a regular grid, such as for images, $\mathcal{D} = \frac{1}{m^2|H|} \sum_{i=1}^m \sum_{j=1}^m \int_{\mathcal{R}} K(H^{-1}(\mathbf{z}_{ij} - \mathbf{z})) dI$ can be treated as a normalization constant.

For computational tractability in MSE calculations, it is standard practice to do an asymptotic analysis using just second order statistics. To carry out an asymptotic analysis, we assume that the number of samples, $n = m^2$, tends to infinity via 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, identically distributed irrespective of the grid size. For the ease of bandwidth estimation, we also assume that $H = \text{diag}(h_I, h_x, h_y)$.

The asymptotic mean square error at any point (I, x, y) is given as the sum of the square of the asymptotic bias (AB) and the asymptotic variance (AV),

$$\text{AMSE}(I, x, y) = (\text{AB}(I, x, y))^2 + \text{AV}(I, x, y) \quad (3)$$

where,

$$\begin{aligned} \text{AB}(I, x, y) &= \text{AB}\{\hat{f}_{I|x, y}(I|\mathbf{t})\} \\ E[\hat{f}_{I|\mathbf{t}}(I|\mathbf{t})] - f_{I|\mathbf{t}}(I|\mathbf{t}) &= \frac{1}{2} \text{tr}\{HH^T \nabla^2 f_{I|\mathbf{t}}\} \\ &= \{h_I^2 a^2 + h_x^2 b^2 + h_y^2 c^2 + (\text{higher orders of } H)\} \end{aligned}$$

such that,

$$\begin{aligned} 2a(I, x, y) &\triangleq \frac{\partial^2 f_{\epsilon|\mathbf{t}}}{\partial \epsilon^2} \\ 2b(I, x, y) &\triangleq \frac{\partial^2 f}{\partial \epsilon^2} \left(\frac{\partial r}{\partial x}\right)^2 - 2 \frac{\partial^2 f}{\partial \epsilon \partial x} \frac{\partial r}{\partial x} - \frac{\partial f}{\partial \epsilon} \frac{\partial^2 r}{\partial x^2} + \frac{\partial^2 f}{\partial x^2} \\ 2c(I, x, y) &\triangleq \frac{\partial^2 f}{\partial \epsilon^2} \left(\frac{\partial r}{\partial y}\right)^2 - 2 \frac{\partial^2 f}{\partial \epsilon \partial y} \frac{\partial r}{\partial y} - \frac{\partial f}{\partial \epsilon} \frac{\partial^2 r}{\partial y^2} + \frac{\partial^2 f}{\partial y^2} \end{aligned} \quad (4)$$

and,

$$\begin{aligned} \text{AV}(I, x, y) &= \text{AV}(\hat{f}_{I|\mathbf{t}}(I|\mathbf{t})) \\ &= \left(\frac{(\Delta x)(\Delta y) f_{\epsilon|x|\mathbf{t}} \int_{\mathcal{R}^3} K(\mathbf{z})^2 d\mathbf{z}}{n|H|} \right. \\ &\quad \left. \triangleq \frac{d(I, x, y)}{nh_I h_x h_y} \right) + (\text{higher orders of } n, H) \end{aligned} \quad (5)$$

Equations 4 and 5 bring out the explicit dependence of the bias and variance on the derivatives of the regression curve $r(x, y)$, and on the derivatives of the conditional PDF for the residual, $f_{\epsilon|\mathbf{t}} \equiv f_{\epsilon|\mathbf{t}}(I - r(\mathbf{t}))$. In the next section, we find bandwidths in terms of these expressions.

3.1. Global Bandwidth Estimate

AIMSE, the asymptotic mean-square error integrated over the whole domain-range space is given by,

$$\begin{aligned} \text{AIMSE} &= \int \text{AB}^2(I, x, y) + \int \text{AV}(I, x, y) = \\ &\left(\begin{array}{c} h_I^2 \\ h_x^2 \\ h_y^2 \end{array} \right)^T \left(\begin{array}{ccc} \|a\|^2 & \langle a, b \rangle & \langle a, c \rangle \\ \langle a, b \rangle & \|b\|^2 & \langle b, c \rangle \\ \langle a, c \rangle & \langle b, c \rangle & \|c\|^2 \end{array} \right) \left(\begin{array}{c} h_I^2 \\ h_x^2 \\ h_y^2 \end{array} \right) \\ &\quad + \frac{\Delta x \Delta y R(K)}{nh_I h_x h_y} \end{aligned} \quad (6)$$

where $R(g) \triangleq \int g(\mathbf{z})^2 d\mathbf{z}$. Optimal bandwidth matrix is sought by minimizing the above expression, which is a non-trivial problem. Hence, we suggest an upper-bound

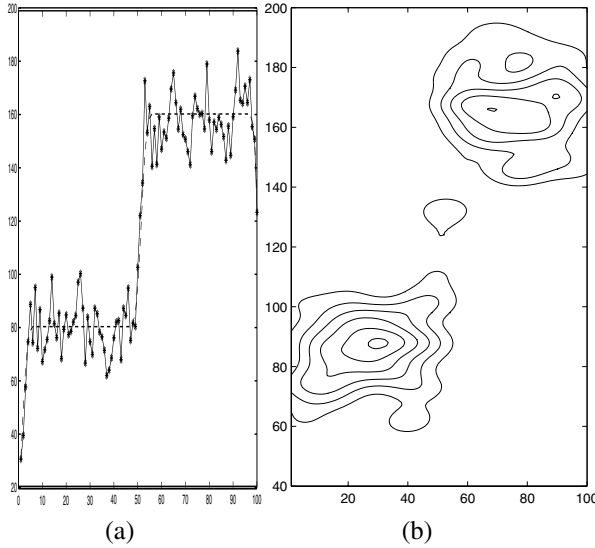


Figure 2: This figure depicts the need for locally adaptive bandwidth estimates. (a) 1-d signal with an edge (b) PDF contours showing a PDF mode at the edge which translates to a step-like region being discovered there.

for AIMSE and compute parameters to minimize the upper bound. By applying Cauchy-Schwartz, we get,

$$\text{AIMSE} \leq (h_y^2 \|a\| + h_x^2 \|b\| + h_y^2 \|c\|)^2 + \frac{\Delta x \Delta y R(K)}{nh_x h_y} \quad (7)$$

The bandwidth parameters that minimize the upper bound in Equation (7) are,

$$\begin{aligned} h_I^* &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|b\| \|c\|}}{12 \|a\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \\ h_x^* &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \|c\|}}{12 \|b\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \\ h_y^* &= \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \|b\|}}{12 \|c\|^3} \right]^{\frac{1}{7}} n^{-\frac{1}{7}} \end{aligned} \quad (8)$$

and consequently,

$$\begin{aligned} \text{AIMSE}^*(I, x, y) \\ \leq 21 \left[\frac{\Delta x \Delta y R(K) \sqrt{\|a\| \|b\| \|c\|}}{12n} \right]^{\frac{4}{7}} \end{aligned} \quad (9)$$

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

In Figure 2, we show a 1-dimensional signal and the estimated PDF using the globally asymptotic bandwidths estimated using Equation (8). The PDF contours show that a mode is discovered on the blurred boundary - thus a segment is discovered on the edge. This is highly undesirable since a large edge with substantial blur can be broken up into multiple segments - all present on the boundary between two regions. This is because of the fact that the global bandwidth reflects the average properties of the image residues. Since regression residues at large edges lie typically on the tail of the distribution, this suggests that global bandwidths may be inadequate to deal with edges having different heights and blurs. Hence, locally adaptive bandwidths are needed to adapt to this image structure.

3.2. Local Bandwidth Estimates

We adapt the PDF estimate to local image structure by using sample point estimators (see pages 185-190, [9]) since they reduce the asymptotic bias and yield bona fide PDF estimates. This estimator, give below, has a different bandwidth parameter for each image pixel.

$$\tilde{f}_{I|t}(I|t) = \frac{1}{m^2 \mathcal{D}} \sum_{i=1}^m \sum_{j=1}^m \frac{1}{|H_{ij}|} K(H_{ij}^{-1}(\mathbf{z}_{ij} - \mathbf{z})) \quad (10)$$

There are several ways to estimate local bandwidths (refer to Section 3.1, [5]). However, most are computationally intensive and would imply that the best features of the mean-shift algorithm, its simplicity and speed, would be compromised. Hence, we use a very simple estimator, called Abramson's square root law [9] which modifies the local bandwidths H_i using the following rule:

$$H_{ij} = H \times \left\{ \frac{\lambda}{f(\mathbf{z}_{ij})} \right\}^{\gamma} \quad (11)$$

where $\gamma \in \{0, 1\}$ and λ , the geometric mean of $\{f(\mathbf{z}_{ij})\}$'s, is a normalization factor. Abramson suggested taking $\gamma = 0.5$. The most attractive property of this law is that it is simple and it considerably reduces the bias (by 2 orders in the bandwidth parameters) [6].

4. Resolution-Guided Segmentation

In the pervious sections, we presented the kernel PDF estimator and a scheme to estimate bandwidth parameters. In the current section, we present our algorithm which is dependent on the (spatial) resolution at which the image is analyzed.

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

analysis at the chosen scale. For the purpose of segmentation (or edge detection), we adopt this philosophy.

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 to estimate the multi-modal image PDF. 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.

The intensity component of the mapping yields refined estimates for regression and the noise realization. These estimates, along with the PDF estimate allows us to estimate the adaptive scale parameters thereby yielding a locally adaptive mean-shift segmentation. This locally adaptive step can be repeated a number of times to define better estimates but it has been noticed that practically no difference in PDF estimates occurs by repeated iterations in contrast to the criticality of the first two steps. 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). Denote the smoothing function at scale s as $\theta_s(\mathbf{t}) = \frac{1}{s}\theta(\frac{\mathbf{t}}{s})$. The regression estimate is given by $\hat{r}_s(\mathbf{t}) = I_s(\mathbf{t}) = I * \theta_s(\mathbf{t})$. Derivatives 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 $\theta_s(\mathbf{t}) = \mathcal{N}(0, s^2I)$, the Gaussian function.
2. *Global Bandwidth Computation:* 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 modelled as a Generalized Gaussian Distribution (GGD). GGD is a parameterized family of distributions. The functions $(a(\cdot), b(\cdot)$ and $c(\cdot))$ are easily computed in terms of these parameters as explained in Section 4.2.
3. *Local Bandwidth Computation:* Once we have the globally estimated scale parameters, we use these to estimate the probability at every pixel using Equation (2) and locally adapt the scale parameters using Abramson's law (Equation (11)).
4. *Mean Shift:* We consider kernels such that $K(x) = k(\|x\|^2)$ where $k(\cdot)$ is convex and decreasing. Then, defining a transformation $M : \mathcal{R}^3 \rightarrow \mathcal{R}^3$ such that for any $\mathbf{p} = [I, x, y]^T$,

$$M(\mathbf{p}) = \frac{\sum_{i,j} \mathbf{z}_{ij} k'(\|H_{ij}^{-1}(\mathbf{z}_{ij} - \mathbf{p})\|^2)/|H_{ij}|}{\sum_{i,j} k'(\|H_{ij}^{-1}(\mathbf{z}_{ij} - \mathbf{p})\|^2)/|H_{ij}|} \quad (12)$$

It follows from (Theorem 2 in [5]) that the sequence $\{\mathbf{p}\}_k$ defined by $\mathbf{p}_{k+1} = M(\mathbf{p}_k)$ in Equation (12) converges to a local mode of the PDF defined in Equation (2). This convergent process, called mean-shift, when applied to each image pixel, maps it to its corresponding mode.

5. *Segmentation:* A boundary between two pixels is detected if the normalized distance between the two pixels $\mathbf{p}_{i,j}$ and $\mathbf{p}_{k,l}$, $\|M(\mathbf{p}_i) - M(\mathbf{p}_j)\|_{H_{ij;kl}^{-1}} > \frac{1}{2}$ where $H_{ij;kl} := \frac{H_{ij} + H_{kl}}{2}$ and $\|x\|_A := x^T A x$. Eventually and optionally, we discard small regions of size less than α . A reasonable choice for $\alpha = h_x^* \times h_y^*$.

4.2. Bandwidth Estimation

To estimate the bandwidths in Equations (8), we need to compute $a(\cdot)$, $b(\cdot)$ and $c(\cdot)$.

- $a(\cdot)$: GGD for ϵ is given by, $f_\epsilon(I) = \frac{\beta}{2\alpha\Gamma(\frac{\beta}{2})} \exp(-|I|/\alpha)^\beta$ 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 and $a(\cdot)$ are computed directly using the moments of the histogram of residuals ([8]).
- $b(\cdot)$ and $c(\cdot)$: In this paper we assume that the noise is i.i.d. and is independent of the signal. Hence, in Equation (4) (and (5)), second and fourth terms reduce to zero. Further, the third term is negligible (confirmed experimentally) as compared to the first. Hence, $b(I, x, y) \approx a(I)(I(x, y) * \nabla_x \theta_s(x, y))^2$ and $c(I, x, y) \approx a(I)(I(x, y) * \nabla_y \theta_s(x, y))^2$.

5. Results

We validate the algorithm on real images by considering its two aspects separately, (1) global bandwidth selection and (2) adapting the global bandwidths using Abramson's square-root law. Both aspects are validated with respect to ground truth - i.e., we take real images, point out those features that pose problems to any segmentation tool and show how our algorithm performs. We validate the global bandwidths obtained by our algorithm using the segmentation software called EDISON [2]. EDISON is chosen since it is a publicly accessible software and it allows us to evaluate our results keeping the post-processing same. Since the EDISON software assumes $h_x = h_y$, we choose the largest of the two bandwidths for our results. This provides us with a lower bound performance for the bandwidth estimation algorithm. Locally adapted bandwidths are validated on our own implementation of mean-shift segmentation algorithm.

s	House			Tree		
	h_I	h_x	h_y	h_I	h_x	h_y
1	2.8	4.8	4.0	6.2	3.8	2.0
2	4.0	8.0	5.7	7.7	6.7	3.2
4	4.7	10.9	7.3	9.3	10.6	5.6
8	5.9	16.1	10.3	10.6	19.12	9.8

s	Sailboat			Peppers		
	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
8	7.1	13.5	13.0	6.2	9.3	11.6

Table 1: Bandwidth parameters, (h_I, h_x, h_y) , for the four images. The bandwidths are discovered at each spatial resolution (wavelet analysis scale) $s = 1, 2, 4, 8$.

5.1. Global Bandwidth Selection

In the regression-based framework we presented, the resolution at which the image is viewed plays a central role. Depending upon the resolution (spatial scale) chosen, the regression function and subsequently the bandwidth parameters are estimated. Using such an approach, a multi-resolution segmentation can be defined based on multi-resolution analysis of a signal. This, however, is the subject of developing papers: here we study the behavior of the proposed algorithm at one scale. Hence, we present results using a single spatial resolution.

We first present two outdoor images - that of a tree in Figure 3(a) and that of a house in Figure 3(b) in the top row. These images were chosen to be representative of real world images. The edges in the house image are straight lines typical of man-made objects. The edges in the tree image are typical of natural scenes. We depict the estimated bandwidths for these images at each of the wavelet analysis scale (the s column) in Table 1. We see that as the analysis scale increases, both the spatial bandwidth and the photometric bandwidth increases - this implies that a larger analysis scale captures properties that are *more* global and should lead to larger features getting detected. In this paper we present results at one analysis scale only ($s = 4$ for all images).

Another factor in the choice of these images is that they have all the complexities that make segmentation difficult - namely, blurred edges, corners and multiple regions being present close by. For example, there is considerable blur between the roof and the sky and the tree and the sky on the right top corner of image in Figure 3(b). Similarly, in the house image, one can see sharp corners and places where more than one region converge, for example where the slanted roof, chimney and the shadow of the roof converge.

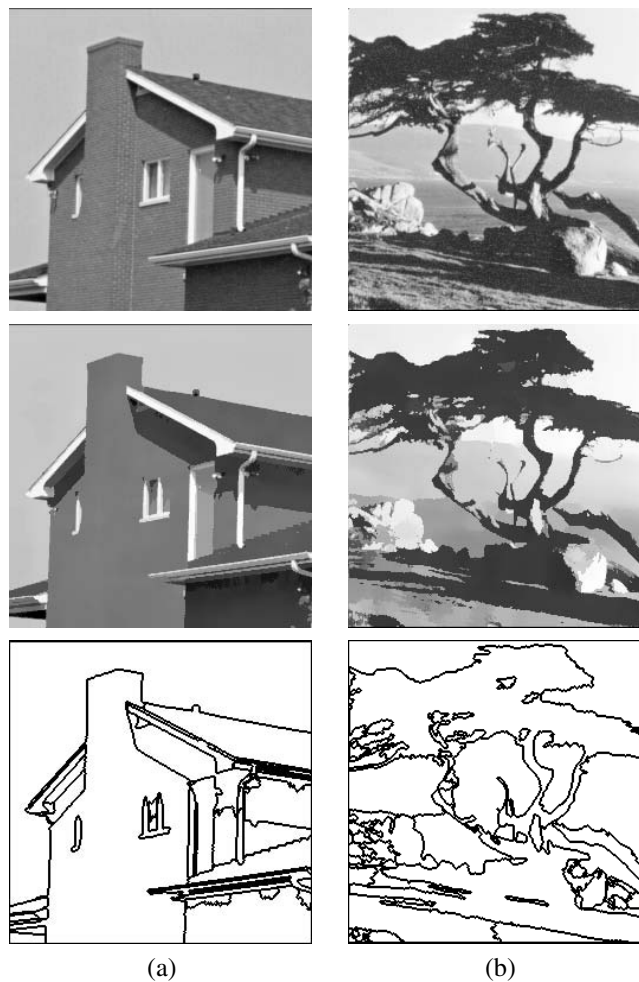


Figure 3: Results for (a) House and (b) Tree images with estimated bandwidths $(h_I, h_x, h_y, \alpha) = (4.7, 11, 11, 121)$ and $(7.7, 7, 7, 49)$ respectively. The original images are in the first row, second row contains the mean-shift filtered output and the bottom row shows segment boundaries.

In spite of these recognized difficulties, the filtered outputs, based on estimated modes at the bandwidths discovered by us, retained features like corners and multiple edges without rounding them. This is clear from the images presented in the second row of Figure 3 (a) and (b). Also note that in the filtered image, presented in middle row of Figure 3 (a), the blurred boundary between the tree and the sky is sharpened. Consequently, we can say that the bandwidths estimated are successful in detecting the segments present in the image.

We chose the Sailboat image in Figure 4(a) to compare the performance of our algorithm with automatic scale selection to that in [5] where hand-optimized bandwidths are chosen. One can see that our result, in Figure 4(b), compares favorably (compare with Figure 8 in [5]). We are able



Figure 4: Results for Sailboat image - size 512×512 (a). (b) depicts the segmented image with overlaid boundaries for global bandwidths $(h_I, h_x, h_y, \alpha) = (6.2, 8, 8, 64)$. (c) shows the result for adaptive bandwidths using Abramson's Law (Equation 11)

to discover all the image structure with automatic bandwidth selection that was discovered in [5]. This includes small structures like the boat, trees at the far side of the lake as well as large structures like the lake, clouds and the sky.

We chose the Peppers image in Figure 5(a) as it posed different challenges. It has some regions like the highlights on the peppers which are very blurry and are difficult to segment. Moreover, this image brings to fore the limitations of global bandwidth estimation and other assumptions of our algorithm. Thus, this image clearly delimits the performance of the algorithm while pointing us in the direction of how to proceed further.

We discuss these issues now: (1) Firstly, our algorithm is able to delineate the highlight regions (Figure 5(b)), however we get several segments while we perceive one. We would like to state though that we empirically observed that the highlights could be segmented out at other spatial resolutions. This implies that either (a) An approach that chooses different bandwidths adaptively in the image is required, or, (b) A multi-resolution segmentation scheme is required that detects regions at several resolution levels, and combines this information together. (2) The algorithm is not able to handle ramp-like structures. For example, look at the right boundary of the tall pepper, the bottom boundary of the pepper in the foreground and the left boundary of the pepper on the right above the really bright pepper. All these boundaries/ regions have been broken into steps (see Figure 5(b)). This is due to the fact that the modes of the PDF are in the spatial-intensity space. Consequently, there is no structure to handle ramp-like features. A formulation that deals with these linear structures is thus required.

5.2. Local Bandwidth Selection

In this section, we test whether locally adapting bandwidths by using Abramson's law is sufficient to address the problems mentioned at the end of the last section. These results are presented in Figures 4(c) and (5)(c). Where as no appreciable difference can be noticed for the Sailboat image (between Figures 4 (b) and (c)), some difference between Figures 5 (b) and (c) is indeed visible for peppers image. Specifically, in Figure 5(c), small-scale changes (small regions; small local contrast between the two halves of the pepper in the foreground) are better estimated but problems due to piecewise linear (not constant) regions and large blur remains.

6. Discussions and Conclusions

In this paper, we presented an automatic bandwidth selection procedure for mean-shift based image segmentation. The regression-based model we used enabled us to develop a simple plug-in bandwidth estimator which is fast and easy to compute. We also showed that the bandwidths selected are consistent and the particular choice of bandwidths bounds the asymptotic mean square error for the estimated image PDF. This error asymptotically goes to zero at the optimal rate. We also validated the bandwidth selection procedure with results on real data.

Further, the bandwidth selection procedure gives a central role to the spatial resolution at which the image is analyzed which we feel is a natural way of doing things as humans tend to view signals at different resolutions too. This feature of our algorithm can be used a central engine of a multi-resolution segmentation algorithm. This is an interesting issue for further research.

We also noticed certain drawbacks of our algorithm: (1)

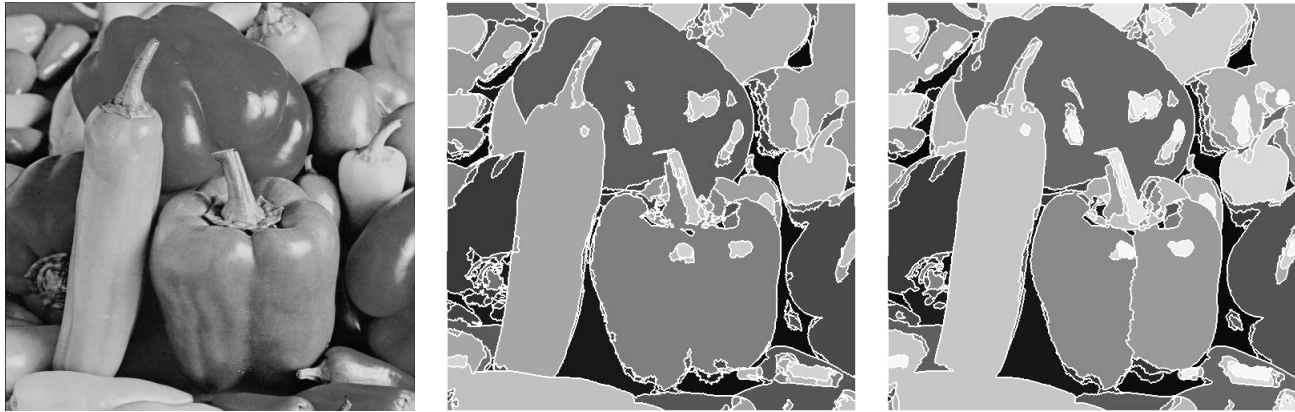


Figure 5: Results for Peppers image - size 512×512 (a): (b) depicts the segmented image with overlaid boundaries for global bandwidths $(h_I, h_x, h_y, \alpha) = (4.8, 7, 7, 49)$. (c) shows the result for adaptive bandwidths using Abramson's Law (Equation 11)

Firstly, large blur could not be handled, and (2) Secondly, ramp-like structures are broken into steps. We explored whether these could be addressed by using locally adaptive bandwidths generated by applying Abramson's square root law. Our results were negative. These lead us to believe that (A) problem (1) needs to be addressed either using a multi-resolution segmentation approach, or by a more sophisticated locally adaptive bandwidth estimation approach (we think this is an open research issue), and (B) Ramp-like structures can only be addressed by bringing in explicit information about image gradients not just in bandwidth estimation but in the theoretical formulation of image PDF as well.

Acknowledgements

The support of the Office of Naval Research under grant N00014-03-1-0107 is gratefully acknowledged. We also thank Ashish Jagmohan for his invaluable suggestions for preparing this manuscript.

References

- [1] N. Ahuja. A transform for multiscale image segmentation by integrated edge and region detection. *PAMI, 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. An algorithm for data-driven bandwidth selection. *PAMI, IEEE Trans*, 25(2):281–288, Feb. 2003.
- [4] D. Comaniciu and P. Meer. Distribution-free decomposition of multivariate data. *Pattern Analysis and Applications*, 2:22–30, 1999.
- [5] D. Comaniciu and P. Meer. Mean shift: a robust approach toward feature space analysis. *PAMI, IEEE Trans*, 24(5):603–619, May 2002.
- [6] D. Comaniciu, V. Ramesh, and P. Meer. The variable bandwidth mean shift and data-driven scale selection. *ICCV, Eighth Intl Conf.*, 1:438–445, 2001.
- [7] K. Fukunaga and L. D. Hostetler. The estimation of the gradient of a density function, with applications in pattern recognition. *Info. Theory, IEEE Trans.*, 21:32–40, 1975.
- [8] S. G. Mallat. A theory of multiresolution signal decomposition: the wavelet representation. *PAMI, IEEE Trans.*, 11(7):674–693, July 1989.
- [9] D. W. Scott. *Multivariate density estimation: theory, practice and visualization*. Wiley-Interscience, 1992.
- [10] B. W. Silverman. *Density estimation for statistics and data analysis*. Chapman and Hall, 1986.
- [11] M. Tabb and N. Ahuja. Unsupervised multiscale image segmentation by integrated edge and region detection. *IP, IEEE Trans.*, 6(5):642–655, May 1997.