A ROBUST KERNEL DENSITY ESTIMATOR BASED MEAN-SHIFT ALGORITHM

Nevine Demitri and Abdelhak M. Zoubir

Signal Processing Group, Institute of Telecommunications
Technische Universität Darmstadt, Merckstr. 25, 64283
Darmstadt, Germany
{demitri,zoubir}@spg.tu-darmstadt.de

ABSTRACT

We propose a robustification of the mean-shift algorithm. We understand robustness in the statistical sense as the deviation from the nominal, distributional assumption. The derivation of the robust mean-shift vector is based on a robust version of the kernel density estimator (KDE), where the KDE is interpreted as an inner product in a higher dimensional feature space. The mean in this formulation is replaced by an M-estimate in order to robustify against outlying data points. We show the superiority of our algorithm compared to the standard mean-shift algorithm and to the median-shift algorithm using both simulated and real data in both contaminated and uncontaminated data. The real data stems from an image segmentation application for blood glucose measurement.

Index Terms— M-estimator, robust kernel density estimation, mean-shift, mode finding, glucose measurement

1. INTRODUCTION

The mean-shift algorithm is an unsupervised, nonparametric clustering approach that is based on finding the modes of an underlying distribution and assigning each data point to its nearest mode. It does not require *a priori* information on the shape or the number of clusters, as this is rather an output of the algorithm. Each data point in the feature space is moved by the mean-shift vector in the direction of the steepest ascent. The mean-shift vector can be intuitively understood as the difference between a data point and a weighted mean of its neighboring points. Basically, the mean-shift algorithm is a gradient ascent approach with the advantage of an adaptive step-size that is determined by the magnitude of the mean-shift vector. The mean-shift vector reaches convergence when its magnitude becomes zero, indicating a stationary point.

The mean-shift algorithm was proposed in 1975 by Fukanaga and Hostetler [1]. Later, it was readopted and generalized by Cheng [2] as a gradient ascent method. Both Cheng and Carreira-Perpinan [3] proved its convergence for different kernels. Comaniciu *et al.*, [4–8], applied it to low-level vision problems such as segmentation and tracking. The mean-shift algorithm has recently become a popular tool for segmentation of biomedical images [9–11], where neither the number of clusters nor their shape are known in advance.

However, real data is often contaminated, due to either lighting effects or artifacts in the scene of interest, as in our application, such as air bubbles, granularities in the chemical substance or small contaminating particles. For these cases a robust solution is required. Otherwise, the clustering behavior is influenced by outliers, leading to a degraded performance. There has been some work on the median-shift as a more robust alternative to the mean-shift that additionally achieves a significant speed-up [12, 13]. Nevertheless, it tends to under-segment the data [14], and therefore, does not always correctly assign each data point to the corresponding cluster.

In this contribution we propose a different approach for the robustification of the mean-shift algorithm that is based on the robust KDE (RKDE) of Kim and Scott [15]. We show that the robust mean-shift algorithm outperforms the standard mean-shift. Furthermore, we apply our concept to the median-shift to obtain a robust KDE based median-shift that compares favorably to the former one. For validation, we use both simulated 1-D data, as well as real data. The real data stems from an image segmentation application for blood glucose level estimation.

The remainder of the paper is organized as follows: Section 2 briefly revisits the concept of M-estimation as it is the basis of the robustification procedure, while Section 3 explains the robust kernel density estimator. The mean-shift algorithm is covered in Section 4, followed by a derivation of the robust formulation of the mean-shift algorithm. This is succeeded by the presentation of the data sets and a depiction of the results in Section 5. Finally, a conclusion is drawn in Section 6.

2. THE M-ESTIMATOR

Given the location model

$$\mathbf{x}_l = \boldsymbol{\mu} + \mathbf{v}_l, \quad l = 1, \cdots, L, \tag{1}$$

where $\mathbf{v}_l$ and herewith, $\mathbf{x}_l$, are realizations of i.i.d. random variables, drawn from the distributions with density function

978-1-4799-2893-4/14/$31.00 ©2014 IEEE

$f_{\mathbf{V}}$ and $f_{\mathbf{X}}$, respectively. A good model for uncontaminated data is $f_{\mathbf{X}} \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$. The aim is to estimate $\hat{\boldsymbol{\mu}} = \hat{\boldsymbol{\mu}}(\mathbf{x})$ even when the data is strongly contaminated. Contamination can be modelled as follows [16]

$$f_{\mathbf{X}} = (1-\varepsilon)f_N + \varepsilon f_C, \tag{2}$$

where f_N is the nominal distribution, f_C is the contaminating distribution and ε is the degree of contamination. An M-estimate [17] of $\boldsymbol{\mu}$ reads

$$\hat{\boldsymbol{\mu}} = \arg\min_{\boldsymbol{\mu}} \sum_{l=1}^{L} \rho(\mathbf{x}_l - \boldsymbol{\mu}), \tag{3}$$

where ρ is a differentiable loss function. Taking the derivative w.r.t. $\boldsymbol{\mu}$ and setting it to zero leads to

$$\sum_{l=1}^{L} \psi(\mathbf{x}_l - \boldsymbol{\mu}) = 0, \tag{4}$$

where $\psi = \rho'$. For the Maximum Likelihood (ML) estimate ψ is a linear function, *i.e.* $\psi(\mathbf{x}) = \mathbf{x}$. Thus, $\hat{\boldsymbol{\mu}}_{\mathrm{ML}}(\mathbf{x}) = \frac{1}{L}\sum_{l=1}^{L}\mathbf{x}_l$. Huber proposed the use of different loss functions [18], that perform in a nearly optimal manner when $f_{\mathbf{X}}$ is exactly normal as well as when $f_{\mathbf{X}}$ deviates from the nominal assumption. In this way, M-estimators can be seen as a generalization of the ML estimator. Intuitively, they can be understood as a weighted average

$$W(\mathbf{x}_l - \boldsymbol{\mu}) = \begin{cases} \frac{\psi(\mathbf{x}_l - \boldsymbol{\mu})}{\mathbf{x}_l - \boldsymbol{\mu}} & \text{if } (\mathbf{x}_l - \boldsymbol{\mu}) \neq 0 \\ \psi'(0) & \text{if } (\mathbf{x}_l - \boldsymbol{\mu}) = 0. \end{cases} \tag{5}$$

For the majority of the data the weights are close to one, while the outliers are down-weighted. Furthermore, the weights $w_l = W(\mathbf{x}_l - \boldsymbol{\mu})$ fulfill $w_l \geq 0$ and $\sum_{l=1}^{L} w_l = 1$. The location estimate, thus, reads

$$\hat{\boldsymbol{\mu}}(\mathbf{x}) = \sum_{l=1}^{L} w_l \mathbf{x}_l.$$

Usually, a closed-form solution cannot be found and the weights have to be calculated numerically, *e.g.* using an iterative reweighting algorithm [17].

If the problem is that of location estimation with unknown scale, a robust initialization of the scale $\hat{\sigma}$ needs to be found to ensure convergence to a good solution. For that, it can be sufficient to use [19]

$$\hat{\sigma}(\mathbf{x}) = 1.483 \cdot \mathrm{median}(|\mathbf{x} - \mathrm{median}(\mathbf{x})|). \tag{6}$$

Different loss functions exist in the literature [20]. They can be mainly divided into the family of monotone loss functions and the family of redescending loss functions, which tend to zero when residuals grow very large. Fig. 1 illustrates commonly used loss functions.

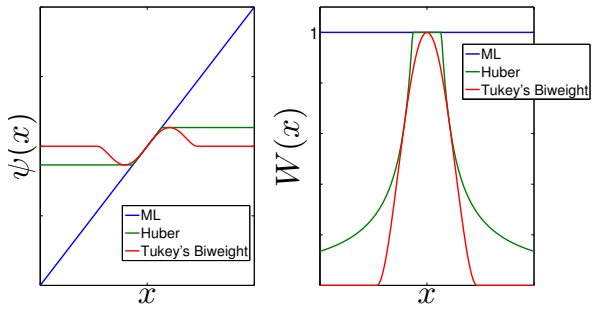

Fig. 1. Different score (a) and weight (b) functions: Tukey's Biweight belongs to the family of redescenders, while Huber belongs to the monotone family. The Maximum Likelihood ψ-function is shown for comparison.

3. THE ROBUST KERNEL DENSITY ESTIMATOR

In [15], Kim and Scott derived a robust kernel density estimator (RKDE) based on the M-estimator. The standard KDE for the samples $\mathbf{x}_l$ is given by

$$\hat{f}(\mathbf{x}) = \frac{1}{L}\sum_{l=1}^{L} K\left(||\mathbf{x} - \mathbf{x}_l||^2\right), \tag{7}$$

where $K(\cdot)$ is the Gaussian kernel function. Using the so-called *kernel trick* [21], the kernel function can be expressed as an inner product in the Hilbert space $\mathcal{H}$, such that $K\left(||\mathbf{x} - \mathbf{x}_l||^2\right) = \langle\Phi(\mathbf{x}), \Phi(\mathbf{x}_l)\rangle$, where Φ is the mapping function: $\Phi : \mathbb{R}^d \rightarrow \mathcal{H}$ and $\langle\cdot\rangle$ denotes the inner product. This yields

$$\begin{aligned}\hat{f}(\mathbf{x}) &= \frac{1}{L}\sum_{l=1}^{L}\langle\Phi(\mathbf{x}), \Phi(\mathbf{x}_l)\rangle \\ &= \langle\Phi(\mathbf{x}), \frac{1}{L}\sum_{l=1}^{L}\Phi(\mathbf{x}_l)\rangle = \langle\Phi(\mathbf{x}), \hat{\boldsymbol{\mu}}_{\Phi,\mathrm{ML}}\rangle,\end{aligned} \tag{8}$$

which is the inner product between $\Phi(\mathbf{x})$ and the sample mean of $\Phi(\mathbf{x}_l)$. Replacing the sample mean by a robust M-estimate

$$\hat{\boldsymbol{\mu}}_{\Phi} = \arg\min_{\boldsymbol{\mu}_{\Phi}} \sum_{l=1}^{L} \rho(||\Phi(\mathbf{x}_l) - \boldsymbol{\mu}_{\Phi}||), \tag{9}$$

leads to the robust KDE

$$\begin{aligned}\hat{f}(\mathbf{x}) &= \langle\Phi(\mathbf{x}), \hat{\boldsymbol{\mu}}_{\Phi}\rangle = \left\langle\Phi(\mathbf{x}), \sum_{l=1}^{L} w_l\Phi(\mathbf{x}_l)\right\rangle \\ &= \sum_{l=1}^{L} w_l K\left(||\mathbf{x} - \mathbf{x}_l||^2\right).\end{aligned} \tag{10}$$

They can be obtained using Iteratively ReWeighted Least Squares (IRWLS) [17]. The IRWLS for the RKDE weight

Table 1. IRWLS for calculating the weights of the RKDE.

1.	Initialize weights $\mathbf{w}^0$ and iteration variable k = 0.
2.	Compute $\|\|\Phi(\mathbf{x}_l) - \boldsymbol{\mu}_\Phi^k\|\|$ as in (11).
3.	Calculate $\hat{\sigma}(\|\|\Phi(\mathbf{x}_l) - \boldsymbol{\mu}_\Phi^k\|\|)$ robustly using (6).
4.	Update $\tilde{\mathbf{w}}^k = \frac{\psi(\|\|\Phi(\mathbf{x}_l)-\boldsymbol{\mu}_\Phi^k\|\|/\hat{\sigma})}{\|\|\Phi(\mathbf{x}_l)-\boldsymbol{\mu}_\Phi^k\|\|/\hat{\sigma}}$ and normalize to $\mathbf{w}^k$.
5.	Stop when $\|\hat{\boldsymbol{\mu}}_\Phi^{k+1} - \hat{\boldsymbol{\mu}}_\Phi^k\| < \hat{\sigma}\epsilon$.

computation is summarized in Table 1.
Using $\hat{\boldsymbol{\mu}}_\Phi = \sum_{l=1}^{L} w_l \Phi(\mathbf{x}_l)$, we obtain for $||\Phi(\mathbf{x}_l) - \hat{\boldsymbol{\mu}}_\Phi||$

$$\begin{aligned}||\Phi(\mathbf{x}_l) - \hat{\boldsymbol{\mu}}_\Phi||^2 &= \langle \Phi(\mathbf{x}_l) - \hat{\boldsymbol{\mu}}_\Phi, \Phi(\mathbf{x}_l) - \hat{\boldsymbol{\mu}}_\Phi \rangle \\ &= \langle \Phi(\mathbf{x}_l), \Phi(\mathbf{x}_l) \rangle - 2\langle \Phi(\mathbf{x}_l), \hat{\boldsymbol{\mu}}_\Phi \rangle + \langle \hat{\boldsymbol{\mu}}_\Phi, \hat{\boldsymbol{\mu}}_\Phi \rangle \\ &= K\left(||\mathbf{x}_l - \mathbf{x}_l||^2\right) - 2\sum_{j=1}^{L} w_j K\left(||\mathbf{x}_j - \mathbf{x}_l||^2\right) \\ &+ \sum_{i=1}^{L}\sum_{j=1}^{L} w_i w_j K\left(||\mathbf{x}_i - \mathbf{x}_j||^2\right). \qquad (11)\end{aligned}$$

4. THE ROBUST MEAN-SHIFT ALGORITHM

Let us first revisit the mean-shift algorithm before deriving its RKDE-based formulation. For a given set of L data points $\mathbf{x}_l, l = 1, \ldots, L$ in a d-dimensional space $\mathbb{R}^d$, the underlying density is estimated using a KDE

$$\hat{f}_K(\mathbf{x}) = \frac{1}{L}\sum_{l=1}^{L} \frac{1}{h^d} K\left(\left(\frac{\mathbf{x} - \mathbf{x}_l}{h}\right)^2\right), \qquad (12)$$

where $K(\mathbf{x})$ is a radially symmetric kernel function with bandwidth h. To locate the modes of the estimated density the zeros of the gradient need to be found,

$$\nabla \hat{f}_K(\mathbf{x}) = \frac{2}{L}\sum_{l=1}^{L} \frac{(\mathbf{x} - \mathbf{x}_l)}{h^{d+2}} K'\left(\left(\frac{\mathbf{x} - \mathbf{x}_l}{h}\right)^2\right) = 0. \qquad (13)$$

Rearranging (13) leads to the so-called mean-shift vector

$$\mathbf{m}_{h,K'}(\mathbf{x}) = \left[\frac{\sum_{l=1}^{L} \mathbf{x}_l K'\left(\left(\frac{\mathbf{x}-\mathbf{x}_l}{h}\right)^2\right)}{\sum_{l=1}^{L} K'\left(\left(\frac{\mathbf{x}-\mathbf{x}_l}{h}\right)^2\right)} - \mathbf{x}\right], \qquad (14)$$

We pick up the derivation of the RKDE-based mean-shift algorithm at (10) and set the derivative of the RKDE to zero to obtain the robust mean-shift vector as

$$\mathbf{m}_{h,K'}(\mathbf{x}) = \left[\frac{\sum_{l=1}^{L} w_l \mathbf{x}_l K'\left(\left(\frac{\mathbf{x}-\mathbf{x}_l}{h}\right)^2\right)}{\sum_{l=1}^{L} w_l K'\left(\left(\frac{\mathbf{x}-\mathbf{x}_l}{h}\right)^2\right)} - \mathbf{x}\right]. \qquad (15)$$

The robust mean-shift algorithm iterates through each data point $\mathbf{x}_l$ in the data set, performing the following steps:

- Calculate the robust weights $w_l^k, l = 1, ..., L$ as in Table 1.
- Calculate the mean-shift vector $\mathbf{m}_{\mathbf{h},K'}(\mathbf{x}_l^k)$ of the current data point $\mathbf{x}_l^k$.
- Shift $\mathbf{x}_l^k$ towards $\mathbf{m}_{\mathbf{h},K'}(\mathbf{x}_l^k)$, hereby calculating the next iteration point as $\mathbf{x}_l^{k+1} = \mathbf{x}_l^k + \mathbf{m}_{\mathbf{h},K'}(\mathbf{x}_l^k)$.
- Stop when convergence is reached, *i.e.* $|\mathbf{x}_l^{k+1} - \mathbf{x}_l^k| < \epsilon$.

Note that the difference between (14) and (15) are the weights w_l. The formulation of the robust mean-shift vector, thus, results in a separation of the problem, where the KDE is, first, expressed robustly, hereby down-weighting the outliers. Then, the modes of the RKDE are found using the mean-shift concept. The proof for the convergence of the sequence $\mathbf{x}_l^k$ given in [5] can be easily generalized to include w_l.

We also apply this concept to the median-shift proposed in [13]. In this, the authors propose choosing the weighted medoid to be the new position of the data point

$$\mathbf{x}_l^{k+1,\text{med}} = \underset{\mathbf{x}\in\{\mathbf{x}_l^0\}}{\arg\min} \sum_{l=1}^{L} \frac{(\mathbf{x} - \mathbf{x}_l)}{h^{d+2}} K'\left(\left(\frac{\mathbf{x} - \mathbf{x}_l}{h}\right)^2\right). \qquad (16)$$

Based on (16), we suggest the introduction of w_l, leading to

$$\mathbf{x}_l^{k+1,\text{med}} = \underset{\mathbf{x}\in\{\mathbf{x}_l^0\}}{\arg\min} \sum_{l=1}^{L} \frac{(\mathbf{x} - \mathbf{x}_l)}{h^{d+2}} w_l K'\left(\left(\frac{\mathbf{x} - \mathbf{x}_l}{h}\right)^2\right). \qquad (17)$$

5. EXPERIMENTS AND DISCUSSION

For the validation of the proposed algorithm, we use both simulated and real data. In the following, a Gaussian kernel is used with a data-driven bandwidth selection as suggested by Comaniciu *et al.* [4]. We calculate the weights w_l using Huber's loss function. The parameter of the loss function is tuned to achieve 95% asymptotic efficiency in the Gaussian case [20] and the IRWLS is initialized with uniform weights.

5.1. Simulation Data

First, we test the robustness of our algorithm by means of 1-dimensional unimodal and bimodal contaminated data modelled as in (2). For the unimodal case f_N follows $\mathcal{N}(\mu_1, \sigma^2)$, while f_C follows $\mathcal{N}(\mu_{C1}, \kappa\sigma^2)$, where κ is the strength of the impulsive component. For the bimodal data, f_N follows $\frac{1}{2}\mathcal{N}(\mu_1, \sigma^2) + \frac{1}{2}\mathcal{N}(\mu_2, \sigma^2)$, while f_C follows $\mathcal{N}(\mu_{C2}, \kappa\sigma^2)$. ε is varied between $0 \leq \varepsilon \leq 0.5$ and κ is varied between $1 \leq \kappa \leq 10$. We evaluate the mean squared error (MSE) between the actual mode locations (μ_1, μ_2) and the modes estimated using the respective mean-shift algorithm $(\hat{\mu}_{1,\text{method}}, \hat{\mu}_{2,\text{method}})$ over 500 Monte Carlo runs for different levels of ε and κ. The results of the robust mean-shift (R-MS) algorithm and the comparison to the standard

mean-shift (MS) algorithm can be seen in Fig. 2. We observe that the overall performance of R-MS is better than MS. For $\varepsilon = 0$ both R-MS and MS perform in a similar manner, with R-MS slightly outperforming MS. The higher the contamination level the better the performance of R-MS, compared to MS.

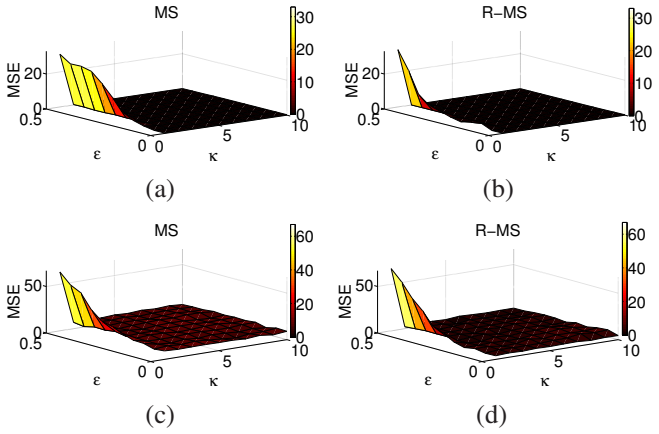

Fig. 2. Results of (a) MS, (b) R-MS for the unimodal distribution and (c) MS, (d) R-MS for the bimodal distribution.

5.2. Real Data

We validate our algorithm in an image segmentation scenario for blood glucose concentration estimation. For this purpose, the chemical reaction between a blood glucose sample and a chemical substance is obtained using a photometric measurement principle that is observed by a camera. The measurement procedure results in a set of frames describing the chemical reaction at the different time instances. We aim at estimating the intensity of the region of interest, *i.e.* the region where the blood glucose is actually reacting with the chemical substance, as it is directly related to the underlying glucose concentration at each time instant. In [11], we showed that the mean-shift algorithm provided good results for estimating the intensity values of the region of interest. In some cases however, the intensity estimate was not accurate due to occurring artifacts such as granularity of the chemical substance, air bubbles or particles contaminating the measurement area. We show, here, the robust mean-shift algorithm results in an improved estimation accuracy.

We present, first, an illustrative example in Fig. 3 to compare the performance of the standard mean-shift to the robust mean-shift algorithm. Clearly, the image is disturbed by a low intensity contamination in the upper image part. This leads to the standard mean-shift producing a mode estimate that is biased towards the low intensity region. The robust mean-shift algorithm manages to down-weight the area of contamination and place a more accurate mode estimate.

A real data set consisting of 47 measurements is used for a more comprehensive validation. The measurements are

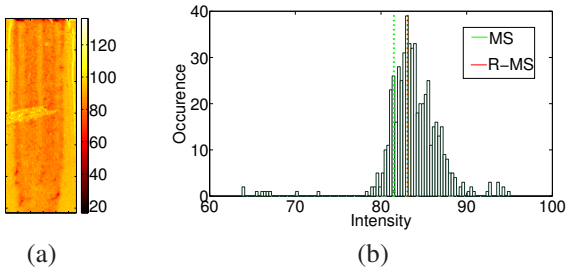

Fig. 3. Example of a contaminated image (a) of the chemical reaction for a blood glucose sample of 150 mg/dl and (b) its histogram, as well as mode estimates for MS and R-MS.

obtained from whole blood samples of five different known glucose levels: 30 mg/dl, 90 mg/dl, 150 mg/dl, 350 mg/dl, 550 mg/dl. Each measurement consists of a set of $N_{\text{frames}} = 605$ frames, corresponding to a testing time of $T_{\text{test}} = 20$s. The level of contamination varies among the frames. As an evaluation measure we use the intergroup variance $\hat{\sigma}^2_{\hat{R}}$ of the intensity estimates $\hat{R}$ at each time instant $0 \leq t \leq T_{\text{test}}$ for a measurement group of the same glucose concentration. We compare the results of the standard mean-shift algorithm (MS) to these of the RKDE mean-shift algorithm (R-MS), as well as the results of the median-shift (MedS) to those of the RKDE median-shift (R-MedS). Table 2 depicts the results. We can see that in most cases the robust formulations of the algorithms perform better than their non-robust counterparts. Even when the contamination is not as strong as in the example of Fig. 3, the robust formulations outperform. This is due to the finite data size, where the asymptotic assumption cannot be met.

Table 2. Comparison of methods for the real data set.

Gluc. [mg/dl]	# frames	$\hat{\sigma}^2_{\hat{R},\text{MS}}$	$\hat{\sigma}^2_{\hat{R},\text{R-MS}}$	$\hat{\sigma}^2_{\hat{R},\text{MedS}}$	$\hat{\sigma}^2_{\hat{R},\text{R-MedS}}$
30	6050	1.75	**1.64**	0.57	**0.14**
90	5445	0.65	**0.61**	0.30	**0.22**
150	5445	1.47	**0.96**	1.31	**0.98**
350	6050	1.05	**0.79**	0.76	**0.6**
550	5445	1.54	**1.01**	**0.89**	1.3

6. CONCLUSION

We have derived a robust formulation of the mean-shift algorithm based on a robust KDE. The mean-shift vector in our formulation is a calculated using the robustly weighted mean, where the weights are calculated using an M-estimator. We have validated our algorithm both on simulation as well as real data for a glucose level measurement application. In both cases our algorithm compares favorably to the standard mean-shift. We are able to achieve more accurate estimates of the glucose levels, even for uncontaminated data.

7. REFERENCES

[1] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Transactions on Information Theory*, vol. 21, no. 1, pp. 32–40, 1975.

[2] Y. Cheng, "Mean shift, mode seeking, and clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 8, pp. 790–799, 1995.

[3] M. A. Carreira-Perpinan, "Mode-finding for mixtures of gaussian distributions," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1318–1323, 2000.

[4] D. Comaniciu, V. Ramesh, and P. Meer, "The variable bandwidth mean shift and data-driven scale selection," in *Proc. Eighth IEEE Int. Conf. Computer Vision ICCV 2001*, 2001, vol. 1, pp. 438–445.

[5] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 603–619, 2002.

[6] D. Comaniciu, "An algorithm for data-driven bandwidth selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 281–288, 2003.

[7] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition, 2000*, vol. 2, pp. 142–149.

[8] D. Comaniciu and V. Ramesh, "Mean shift and optimal prediction for efficient object tracking," in *Proceedings International Conference on Image Processing, 2000.*, vol. 3, pp. 70–73.

[9] A. Mayer and H. Greenspan, "An adaptive mean-shift framework for MRI brain segmentation," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1238–1250, 2009.

[10] F. Sahba and A. Venetsanopoulos, "Mean shift based algorithm for mammographic breast mass detection," in *Proc. 17th IEEE Int Image Processing (ICIP) Conf*, 2010, pp. 3629–3632.

[11] N. Demitri and A. M. Zoubir, "Mean-shift based algorithm for the measurement of blood glucose in handheld devices," in *Proceedings of the 21st European Signal Processing Conference (EUSIPCO) 2013*.

[12] L. Shapira, S. Avidan, and A. Shamir, "Mode-detection via median-shift," in *IEEE 12th International Conference on Computer Vision, 2009*, pp. 1909–1916.

[13] Y. A. Sheikh, E. A. Khan, and T. Kanade, "Mode-seeking by medoidshifts," in *IEEE 11th International Conference on Computer Vision, 2007.*, pp. 1–8.

[14] J. Jordan and E. Angelopoulou, "Mean-shift clustering for interactive multispectral image analysis," *20th IEEE International Conference on Image Processing (ICIP)*, 2013.

[15] J. S. Kim and C. Scott, "Robust kernel density estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008.*, pp. 3381–3384.

[16] P. J. Huber, *Robust Statistics*, Springer, 2011.

[17] R. A. Maronna, R. D. Martin, and V. J. Yohai, *Robust Statistics*, J. Wiley, 2006.

[18] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.

[19] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

[20] P. J. Besl, J. B. Birch, and L. T. Watson, "Robust window operators," *Machine Vision and Applications*, vol. 2, no. 4, pp. 179–191, 1989.

[21] B. Schölkopf and A. J. Smola, *Learning with Kernels*, The MIT Press, 2002.