Clutter Identification Based on Kernel Density Estimation and sparse-recovery

Haokun Wang[a], Yijian Xiang[b], Elise Dagois[a], Malia Kelsey[a], Satyabrata Sen[c], Arye Nehorai[b], and Murat Akcakaya[a]

[a]University of Pittsburgh, Pittsburgh, USA
[b]Washington University in St.Louis, St. Louis, USA
[c]Oak Ridge National Laboratory, Oak Ridge, USA

ABSTRACT

A cognitive radar framework is being developed to dynamically detect changes in the clutter characteristics, and to adapt to these changes by identifying the new clutter distribution. In our previous work, we have presented a sparse-recovery based clutter identification technique. In this technique, each column of the dictionary represents a specific distribution. More specifically, calibration radar clutter data corresponding to a specific distribution is transformed into a distribution through kernel density estimation. When the new batch of radar data arrives, the new data is transformed to a distribution through the same kernel density estimation method and its distribution characteristics is identified through sparse-recovery. In this paper, we extend our previous work to consider different kernels and kernel parameters for sparse-recovery-based clutter identification and the numerical results are presented as well. The impact of different kernels and kernel parameters are analyzed by comparing the identification accuracy of each scenario.

Keywords: Clutter identification, nonstationary clutter, sparse-recovery, batch orthogonal matching pursuit, kernel density estimation, kernel bandwidth

1. INTRODUCTION

Detecting and tracking targets in the presence of nonstationary clutter, noise, and interference have been the most pertinent and challenging problems in radar systems. In the practical scenarios, various issues, such as the terrain and weather conditions, dynamics of the targets, and hostile electronic environments, may fluctuate and alter the statistical characteristics of the environmental background (clutter) during the radar operation period.[1] These nonstationarities of the clutter, if not adaptively coped with, can significantly hinder the performance of the classical detection and tracking techniques.[2,3] For example, in the target detection problem, the fluctuation of the clutter distribution parameters may require to readjust the threshold of the detector, and even worse, a change of the family of the clutter distributions may require to redesign the detector altogether, in order to maintain the optimal or nearly optimal detection performance.

Traditionally, however, in most of the radar applications, the target detection and tracking techniques have been developed with a specific clutter distribution, which is assumed to be known a priori and be stationary throughout the entire processing period. Although the Gaussian distributions are used extensively to represent the clutter characteristics, it has been shown in the literature that the Gaussian representations suffer from performance deterioration when the the measured clutter data are heavy-tailed.[4,5] Instead, some compound-Gaussian distributions, such as K distribution and Student-t distribution, are proposed to accurately model the received clutter, for example, sea- or foliage-clutter when radars operate in high-resolution and/or low-grazing-angle modes.[6–10] Weibull and lognormal distributions are other two popular clutter distributions that achieve great fitness to the real data, and are capable of modeling the spiky nature of the clutter.[11,12] In nonstationary operating scenarios, however, a prior knowledge of the clutter characteristics represented using a fixed, parametric distribution does not hold true anymore as the clutter statistics may drastically alter during analysis. Therefore,

Further author information: Send correspondence to Murat Akcakaya
Murat Akcakaya: E-mail: akcakaya@pitt.edu, Telephone: 412-624 8622

having a capability of determining the clutter distribution on-the-fly would be crucial to maintain, or even to improve, the radar detection/tracking performance in nonstationary environments.

In order to create such an adaptive framework, recent advances in computational capabilities have allowed radar system designers to consider the design of more complex and intelligent systems, termed as cognitive radar systems.[13,14] Cognitive radars aim to early detect the changes of the environments (clutter), precisely learn the new distribution of clutter, and adaptively update the detection/tracking algorithms for maintaining or bettering the performances achieved by current (nonadaptive) state-of-the-art systems.

In our previous work,[1] we proposed a data-driven method and used the (extended) CUSUM algorithm to address the first issue in the cognitive radar framework, i.e., finding out whether the modeled/assumed clutter distribution has changed or not. For the second issue, we previously developed a sparse-recovery based clutter identification method,[15,16] that applies the kernel density estimation (KDE) and a batch orthogonal matching pursuit (BOMP) method to identify the distribution of the received clutter data based on a pre-learned dictionary of distributions. Earlier, another clutter identification method, namely the Ozturk algorithm,[17,18] was proposed to identify the received clutter distribution as the nearest neighbor to a dictionary of distributions after transforming each distribution to a point on two-dimensional plane. However, comparing the performances of the sparse-recovery based clutter identification approach (i.e., BOMP method) with that of the Ozturk algorithm, we have shown[16] that the BOMP method has (i) improved accuracy in identifying clutter distributions that have different parameters, but are from the same family; and (ii) robustness in terms of measurements used for dictionary generation and test distribution identification.

To further explore the potentials of the sparse-recovery based clutter identification method, in the paper, we investigate the effects of different kernels types and kernel parameters used in KDE on the clutter identification accuracy, while only the normal kernel with a default parameter was considered in our previous work. With respect to the four different kernels, namely the normal, triangle, box, and Epanechnikov kernels, we observe that the BOMP method is robust to the kernel bandwidth selection, and the Epanechnikov kernel is found to be the most suited kernel for the BOMP algorithm.

The rest of the paper is organized as follows. In Section 2, we describe the sparse-recovery based clutter identification method and provide the details of the kernel density estimation approach. In Section 3, we demonstrate various numerical results to objectively compare the effects of different types of kernels and kernel parameters. Section 4 provides a discussion of the observed results, and Section 5 concludes the paper.

2. CLUTTER IDENTIFICATION METHOD

In this section, we introduce the sparse-recovery based clutter identification method, including the impact of kernel density estimation, specifically the effects of the kernel-type and kernel-bandwidth on the estimation procedure.

2.1 BOMP Method

sparse-recovery algorithms aim to estimate a signal by linearly adding columns from a dictionary of predefined waveforms. Typically, representing the dictionary as a matrix $\boldsymbol{D} = \{\boldsymbol{\phi}_\omega : \omega \in \Omega\}$, whose each member ϕ_ω is called as atom and coefficient ω is obtained from a index set Ω, sparse-recovery techniques solve for $\boldsymbol{\gamma}$ using $\boldsymbol{D\gamma} = \boldsymbol{s}$, where $\boldsymbol{s}$ is the original signal and $\boldsymbol{\gamma}$ is a coefficient vector. In general, the objective is to estimate the signal with m atoms, where m is much smaller than the size of dictionary N,[19–21] and hence it is referred to as a sparsity-based estimation technique.

In general, dictionaries $\boldsymbol{D}$ are designed to be fat matrices, meaning a single exact solution of $\boldsymbol{\gamma}$ does not exist. Instead, greedy approaches are used to solve for the signal as an approximation. The most popular greedy approaches fall under the category of Matching Pursuit (MP) algorithms, one of which is the orthogonal matching pursuit (OMP) algorithm that reconstructs the input signal with the least number of atoms, implying the sparsest recovery. For exact-sparse problem, the exact recovery condition (ERC) for OMP method is given as $\max_{\boldsymbol{\phi}} ||\boldsymbol{a} \in \text{span}\{\boldsymbol{\phi}_\lambda : \lambda \in \Lambda_j\}|| < 1$, where the maximum reaches over the atoms that are not be part in the optimal representation of the signal, meaning that the sparest signal reconstruction is unique.[20] Starting with the initial approximation $\boldsymbol{a}_0 = \boldsymbol{0}$ and initial residual $\boldsymbol{r}_0 = \boldsymbol{s}$, OMP method at each step tries to find an

atom which correlates most perfectly with the residual; for example, at step j, the atom index λ_j is calculated by solving the optimization problem:

$$\lambda_j \in \arg\max_{\omega \in \Omega} |\langle \boldsymbol{r}_j, \boldsymbol{\phi}_\omega \rangle|$$

Subsequently, the jth approximation is computed as

$$\boldsymbol{a}_j = \arg\min_{\boldsymbol{a}} \|\boldsymbol{s} - \boldsymbol{a}\|_2 , \quad \text{subject to } \ \boldsymbol{a} \in \text{span}\{\boldsymbol{\phi}_\lambda : \lambda \in \Lambda_j\} ,$$

where $\Lambda_j = \{\lambda_1, ...\lambda_j\}$ denotes the atom-indexes selected till the jth step. Because the residuals are orthogonal to the atoms which have already been chosen, OMP never chooses the same atom twice, resulting to a zero residual after d steps.[22,23]

As a variant of OMP algorithm, BOMP still aims to solve the reconstruction problem by finding the local minimum solution of an undetermined linear system, while reducing the computational complexity by introducing the Cholesky factorization for residual calculation.[21,24] Denoting $\boldsymbol{\alpha} = \boldsymbol{D}^T\boldsymbol{r}$, $\boldsymbol{\alpha}^0 = \boldsymbol{D}^T\boldsymbol{s}$, $\boldsymbol{G} = \boldsymbol{D}^T\boldsymbol{D}$, and the sub-matrix $\boldsymbol{D}_\Lambda$ containing the columns indexed by Λ, we can write a new equation involving the pseudo inverse as

$$\boldsymbol{\alpha} = \boldsymbol{D}^T \left(\boldsymbol{s} - \boldsymbol{D}_\Lambda (\boldsymbol{D}_\Lambda)^+ \boldsymbol{s}\right) = \boldsymbol{\alpha}^0 - \boldsymbol{G}_\Lambda \left({\boldsymbol{G}_{\Lambda,\Lambda}}^{-1} \boldsymbol{\alpha}^0_\Lambda\right)$$

Therefore, with the pre-calculated $\boldsymbol{G}$ and $\boldsymbol{\alpha}^0$, it only needs to compute $\boldsymbol{\alpha}$ instead of $\boldsymbol{r}$ at each iteration. Also, the new multiplier $\boldsymbol{G}_{\Lambda,\Lambda}$ replaces the dictionary $\boldsymbol{D}$, where $\boldsymbol{G}_{\Lambda,\Lambda}$ denotes the progressive Cholesky factorization result.[19,21,24]

2.2 Kernel Density Estimation

Kernel density estimation (KDE) is commonly used to estimate the pdf of a random variable, which could be viewed as an update of histogram, where weight function becomes the kernel function with bandwidth. For example, given a set of random samples $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ from an unknown distribution $f_X(x)$, the kernel density estimator is represented as

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) ,$$

where $K(\cdot)$ is a kernel function determining the shape of weight function, and h is the kernel bandwidth determining the amount of smoothing applied in the estimation process.[25,26] The kernel function could be any symmetric pdf since it meets the following properties: $\int K(t)dt = 1$ and $K(t) \geq 1$. Furthermore, given sufficient number of samples, the kernel density estimator $\hat{f}(x)$ would asymptotically converge to any density function $f_X(x)$, and therefore KDE is applicable for almost every distribution.[25]

Now, it is obvious that the choice of kernel type and bandwidth would critically affect the estimation performance. To evaluate the estimation accuracy, let us define the mean squared error (MSE) as $\text{MSE}(\hat{f}(x)) = \text{E}(\hat{f}(x) - f_X(x))^2 = \text{bias}^2(\hat{f}(x)) + \text{var}(\hat{f}(x))$. Then, by transforming and expanding with Talyor series, we get

$$\text{MSE}(\hat{f}(x)) \approx \frac{1}{4} h^4 k_2^2 f''(x)^2 + \frac{1}{nh} j_2$$

where $k_2 = \int z^2 K(z) dz$ and $j_2 = \int K(z)^2 dz$. Therefore, the global estimation accuracy can be expressed in terms of the mean integrated square error (MISE) as

$$\text{MISE} \approx \frac{1}{4} h^4 k_2^2 \beta(f) + \frac{1}{nh} j_2 , \quad \text{where } \ \beta(f) = \int f''(x)^2 dx .$$

As MISE is a function of bandwidth h, a simple way to obtain the optimal bandwidth is to take gradient of MISE and set it to zero, which results in

$$h_{\text{opt}} = \left[\frac{1}{n} \frac{\gamma(K)}{\beta(f)}\right]^{\frac{1}{5}} , \quad \text{where } \ \gamma(K) = j_2 \, k_2^{-2} .$$

However, as h_{opt} depends on the pdf which is unknown, its practical computation is not possible.

To select the kernel bandwidth h, one commonly applied method is to assume a reference distribution, mostly the Gaussian, and then to compute the optimal bandwidth as $h_{\text{opt}} = (\frac{4\sigma_{\text{s}}^5}{3n})^{\frac{1}{5}}$, where σ_{s} is the sample standard deviation $\sigma_{\text{s}} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})}$. Based on the Gaussian assumption, a better expression of bandwidth is given as $h_{\text{opt}} = \frac{0.9\tilde{\sigma}}{n^5}$, where $\tilde{\sigma} = \min\left(\sigma_{\text{s}}, \frac{\text{IQR}}{1.34}\right)$ and IQR is the inter-quartile range, i.e., the difference between the 75th and 25th percentile points. Another way to estimate the bandwidth is to test a set of bandwidth values, and select the one with highest accuracy. A more data-driven method is the 'plug-in' estimation, by using a separate smooth trick for $f''(x)$ estimation and calculating the gradient based bandwidth.[27–29]

2.3 Clutter Identification

For the clutter identification purpose, we formulate the dictionary based on a data-driven approach using the KDE paradigm as

$$\boldsymbol{D} = [\, f_1(\boldsymbol{s})\ f_2(\boldsymbol{s})\ \cdots\ f_N(\boldsymbol{s})\,]\ ,$$

where each column is created for a pre-defined clutter distribution estimated via KDE. Specifically, to create each dictionary column, S samples are used to calculate an estimated clutter pdf, $f_n(\boldsymbol{s})$, which is then normalized on a set support of W points. Thus, the final dictionary has dimension $W \times N$. In a similar manner, we create the test signal as an estimated pdf $g(\boldsymbol{s})$ by first collecting N_t target-free radar measurements and then applying KDE with the same support W which is used to build the dictionary. Once $g(\boldsymbol{s})$ has been estimated, the BOMP method is applied to select the column(s) from the dictionary $\boldsymbol{D}$ that is(are) the best match to the estimated pdf $g(\boldsymbol{s})$ of the measured clutter data.[15]

3. NUMERICAL RESULTS

In this section, we demonstrate some numerical examples to compare different kernel types as well as the bandwidth at a few specific test sample sizes and dictionary sizes while applying the BOMP method. By randomly choosing the test parameters from a dictionary and creating test samples with KDE, the change of accuracy rate in different scenarios reveals the impact of the kernel types and bandwidth. The dictionary is predefined with the following four distributions:[30]

1. K-distribution: $s_{\text{K}} = |\sqrt{\tau}n|$, where s_{K} follows a K-distribution[30] when $\tau \sim \text{Gamma}(k, \theta)$ [k is the shape parameter and θ is the scale parameter] and $n \sim \mathcal{CN}(0, \sigma_n^2)$.
2. Weibull distribution: $s_{\text{Wbl}} \sim \text{Wbl}(\alpha, \beta)$, where s_{Wbl} follows a Weibull distribution with the shape parameter α and the scale parameter β.
3. Log-normal distribution: $s_{\text{LN}} \sim \text{LogN}(\mu_{\text{LN}}, \sigma_{\text{LN}}^2)$, where s_{LN} follows a log-normal distribution, implying that $(\ln y_{\text{LN}} - \mu_{\text{LN}})/\sigma_{\text{LN}} \sim \mathcal{N}(0, 1)$.
4. Student-t distribution: $s_{\text{St}} = \sqrt{\tau}w$, where s_{St} follows a non-standardized Student-t distribution when $1/\tau \sim \text{Gamma}(v, 1/v)$ and $w \sim \mathcal{N}(0, \sigma_w^2)$.

3.1 Effects of Kernel Types

In order to study the impact of different kernels, we compared the accuracy of clutter distribution identification under four different kernel types and various sample sizes, while keeping a fixed kernel bandwidth calculated by the rule-of-thumb. Specifically, the details of the four kernels are shown in Table 1; the dictionary sample sizes are chosen as 500, 1000, and 2500, and the test sample sizes are varied from 300 to 2800. For each sample size, the accuracy test is computed by $10,000$ Monte Carlo trials and the clutter distribution parameters are chosen as follows:

1. K-distributions with fixed $\sigma_n = 1$, fixed $\theta = 1$, and $k \in \{0.1 : 0.2 : 3.9\} \cup \{4 : 2 : 24\} \cup \{50 : 25 : 200\}$,
2. Weibull distributions with fixed scale $\beta = 1$, and shape parameters $\alpha \in \{0.1 : 0.1 : 3.9\} \cup \{4 : 2 : 20\}$,
3. Log-normal distributions with $\mu_{\text{LN}} = 0$, and $\sigma_{\text{LN}} \in \{0.05 : 0.05 : 1\} \cup \{1.1 : 0.1 : 3\}$,

Table 1. Kernel functions

Kernel Name	Function
Normal	$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}$
Triangle	$K(u) = 1 - \lvert u\rvert, \lvert u\rvert \leq 1$
Box	$K(u) = \frac{1}{2}, \lvert u\rvert \leq 1$
Epanechnikov	$K(u) = \frac{3}{4}(1 - u^2), \lvert u\rvert \leq 1$

4. Student-t distribution with $\sigma_w = 1$, and $v \in \{0.1 : 0.2 : 4.9\} \cup \{5 : 5 : 25\} \cup \{50 : 25 : 200\}$.

The accuracy of clutter identification results are presented in Figure 1(a) to Figure 1(d) respectively for the four chosen kernels.

3.2 Effects of Kernel Bandwidths

To study the impacts of kernel bandwidth, we compared the accuracy of clutter distribution identification under two different bandwidth selection methods:

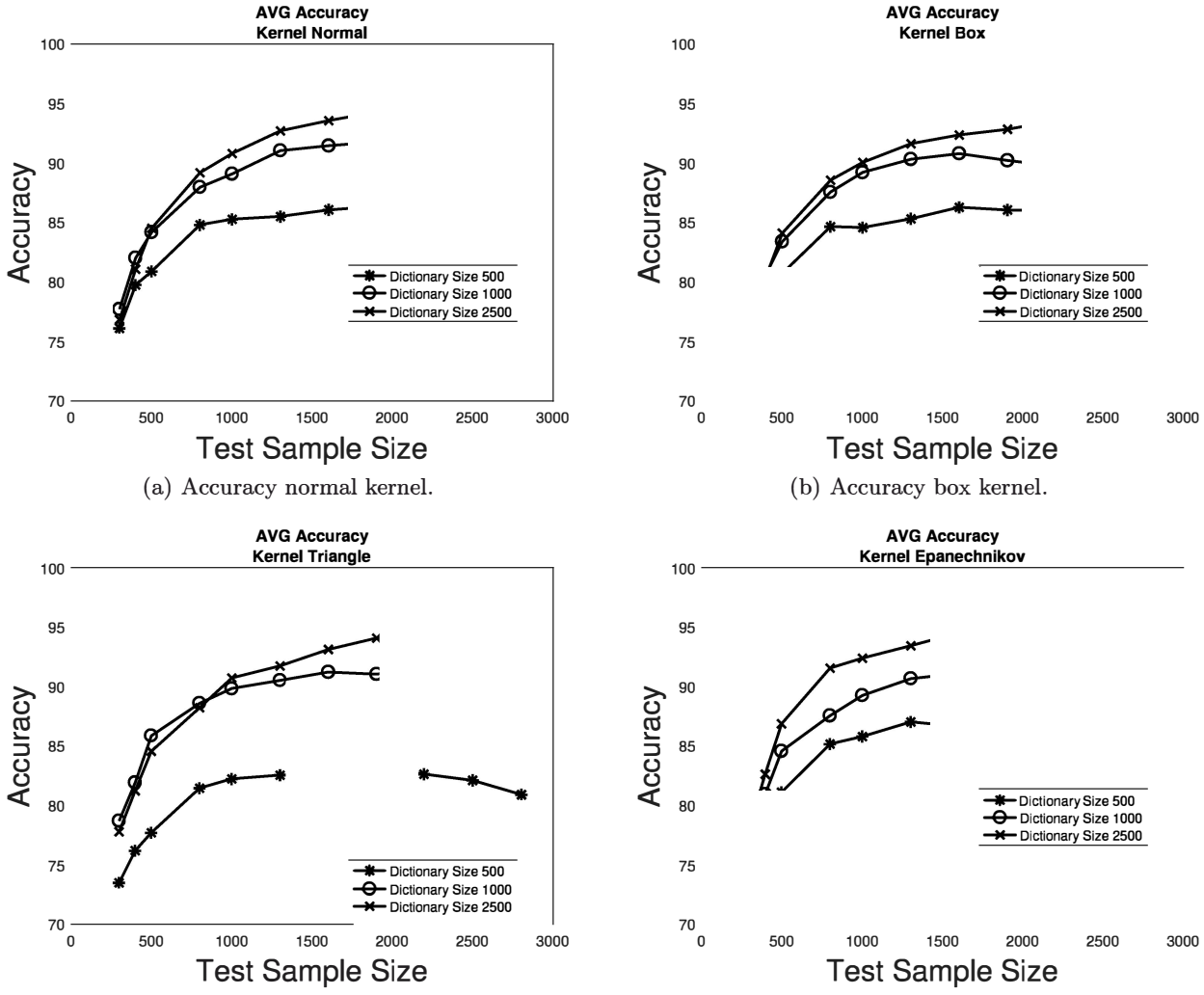
1. Rule-of-thumb: $\hat{h}_{\text{opt}} = \frac{0.9\sigma}{n^5}$.
2. Subjective bandwidth set: creating a set as $\{0.1 : 0.1 : 1.1\} * \hat{h}_{\text{opt}}$, while considering the rule-of-thumb bandwidth as reference.

The other numerical parameters remain the same; for example, the dictionary sample sizes of 500, 1000, and 2500, and the test sample sizes were varied from 300 to 2800. Figure 2(a) to Figure 2(d) show the average identification accuracy computed for the four chosen kernel functions having bandwidth values varying from 10% to 110% of the default value.

4. DISCUSSION

In the first group of figures, Figure 1(a) to Figure 1(d), we notice that the average clutter identification accuracy is higher when dictionary size is larger. With larger dictionary size, each column in the dictionary has more support points, which helps solving the BOMP optimization problem more accurately. Also, the test sample size is roughly correlated with the accuracy, while a fall back is detected when test sample are larger than 1500. This situation is obviously shown in low dictionary size, for which the normal kernel and Epanechnikov kernel are more robust than the others. When applying larger dictionary size, all kernels perform well and with the increase of the test sample size, the accuracy rate roughly rises from 78% to 95% at same pace. When the dictionary has a size of 1000, the accuracy of box kernel is inferior to the others. When dictionary size is doubled from 500 to 1000, the accuracy rate of kernel box increases much slower than the others. Comparing Figure 1(a) and Figure 1(d), the kernel normal seems to be more robust to discrepancies between the dictionary and test sample sizes than the Epanechnikov kernel for large dictionary sizes.

In the second group of figures, Figure 2(a) to Figure 2(d), the accuracy rate of clutter distribution identification changes slightly when bandwidth varies from 10% to 110% of the bandwidth computed through the rule-of-thumb method. This implies that the BOMP method is robust for the application scenarios with different bandwidth. We do not observe a direct correlation between the changes in the bandwidth and accuracy of clutter distribution identification. The rule-of-thumb method, even though it is established for calculation of bandwidth for normal kernels, performs well for other kernels as well. Therefore, we suggest using the rule-of-thumb method for the calculation of the kernel bandwidths. Among the four chosen kernels, the triangle kernel still performs inferior to the others. The normal and Epanechnikov kernels are preferable, as both of them are robust in terms of bandwidth change. On the other hand, the box kernel requires a large dictionary size to achieve similar clutter distribution identification accuracies. Also, we note that the Epanechnikov kernel shows good results with small dictionary size, which is significant for scenarios in which not enough data are available to build the dictionary, especially when updating dictionary online. In such cases, we suggest to apply the Epanechnikov kernel function.



(a) Accuracy normal kernel.

(b) Accuracy box kernel.

(c) Accuracy triangle kernel.

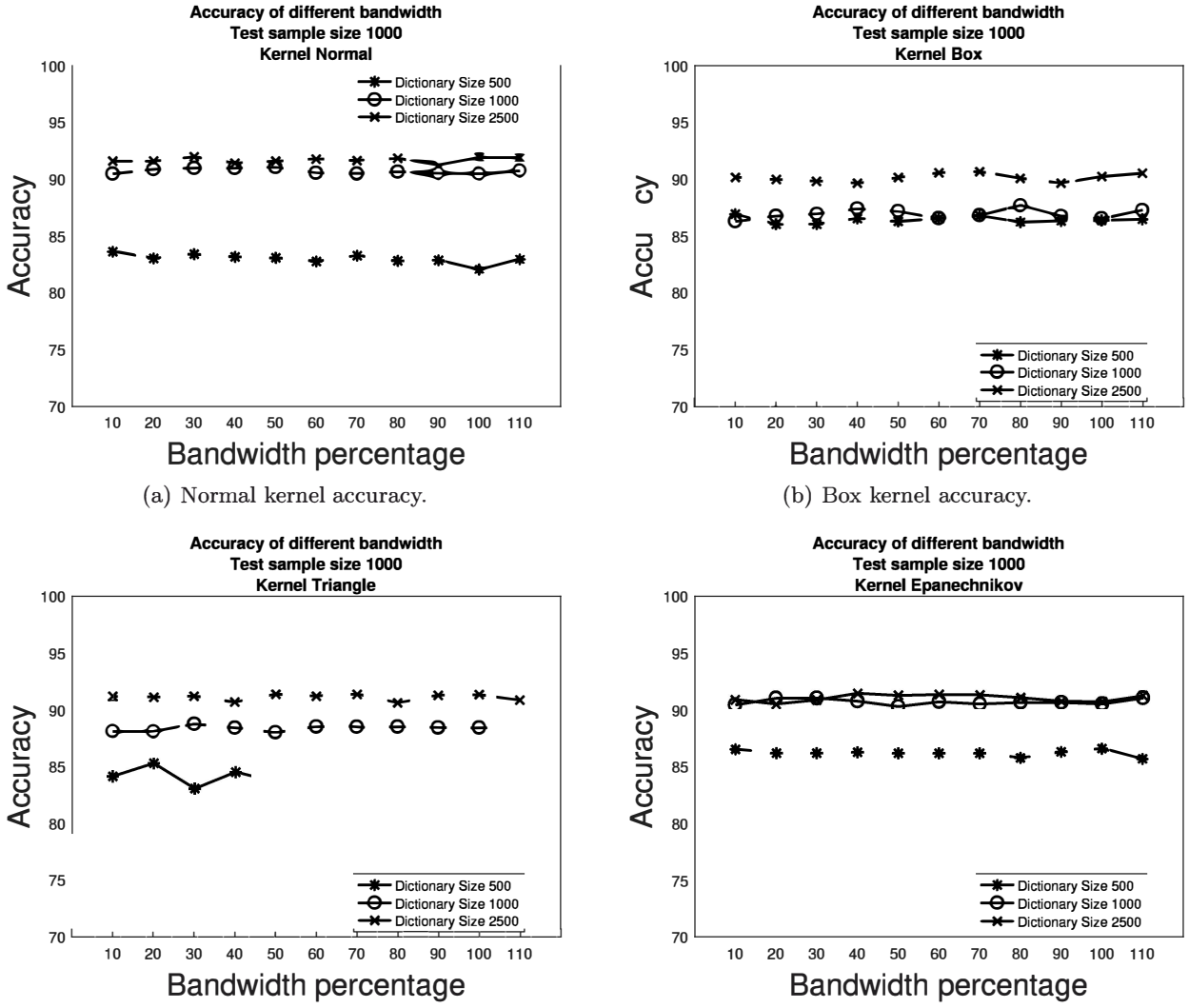
(d) Accuracy epanechnikov kernel.

Figure 1. Effects of kernel types.

5. CONCLUSIONS

In this work, we presented a sparse-recovery based clutter identification method and analyzed its performance with respect to two different kernel bandwidth selection methods and for different kernel functions. Based on the numerical examples, we demonstrated that the sparse-recovery based technique provided (i) robustness in terms of different kernel types and bandwidths; and (ii) high accuracy in identifying clutter measurements originating from different families of distributions. Our results further demonstrated that the Epanechnikov kernel performs the best compared to the other kernels.

In our future work, we plan to adaptively increase and decrease dictionary size of the sparse-recovery based method; such adaptive change in dictionary size will be crucial in order to characterize measured data that may not be well-represented by any specific distribution in the dictionary and to control the computational load. Furthermore, with real measured data, we will incorporate the sparse-recovery based clutter identification method into the design of a fully cognitive radar system, which will include the statistical tests for estimating change points in the clutter distribution, methods for identifying the new clutter distribution and adaptation techniques for detection/tracking algorithms to the newly learned clutter distribution.



(a) Normal kernel accuracy.

(b) Box kernel accuracy.

(c) Triangle kernel accuracy.

(d) Epanechnikov kernel accuracy.

Figure 2. Effects of kernel bandwidths.

ACKNOWLEDGMENTS

This material is based upon the work supported by the Air Force Office of Scientific Research (AFOSR), the DDDAS Program, under Grant No. FA9550-16-1-0386. The work of Sen was performed at the Oak Ridge National Laboratory, managed by UT-Battelle, LLC, for the U.S. Department of Energy, under Contract DE-AC05-00OR22725. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan).

REFERENCES

[1] Akcakaya, M., Sen, S., and Nehorai, A., "A novel data-driven learning method for radar target detection in nonstationary environments," *IEEE Signal Processing Letters* **23**, 762–766 (May 2016).

[2] Kay, S., [*Fundamentals of Statistical Signal Processing: Detection theory*], Prentice-Hall PTR (1998).

[3] Van Trees, H. L., [*Detection, Estimation, and Modulation Theory*], John Wiley & Sons (2004).

[4] Marier, L. J., "Correlated K-distributed clutter generation for radar detection and track," *IEEE Transactions on Aerospace and Electronic Systems* **31**, 568–580 (Apr. 1995).

[5] Palama, R., Greco, M. S., Stinco, P., and Gini, F., "Statistical analysis of bistatic and monostatic sea clutter," *IEEE Transactions on Aerospace and Electronic Systems* **51**, 3036–3054 (Oct. 2015).

[6] Akcakaya, M. and Nehorai, A., "Adaptive MIMO radar design and detection in compound-Gaussian clutter," *IEEE Transactions on Aerospace and Electronic Systems* **47**(3), 2200–2207 (2011).

[7] Sammartino, P. F., Baker, C. J., and Griffiths, H. D., "Adaptive MIMO radar system in clutter," in [*Proc. IEEE Radar Conference*], 276–281 (Apr. 2007).

[8] Balleri, A., Nehorai, A., and Wang, J., "Maximum likelihood estimation for compound-Gaussian clutter with inverse gamma texture," *IEEE Transactions on Aerospace and Electronic Systems* **43**, 775–779 (Apr. 2007).

[9] Wang, J., Dogandzic, A., and Nehorai, A., "Maximum likelihood estimation of compound-Gaussian clutter and target parameters," *IEEE Transactions on Signal Processing* **54**, 3884–3898 (Oct. 2006).

[10] Gini, F., Farina, A., and Lombardini, F., "Effects of foliage on the formation of K-distributed SAR imagery," *Signal Processing* **75**(2), 161–171 (1999).

[11] Sekine, M., Musha, T., Tomita, Y., Hagisawa, T., Irabu, T., and Kiuchi, E., "Weibull-distributed sea clutter," *IEE Proceedings F - Communications, Radar and Signal Processing* **130**, 476– (Aug. 1983).

[12] Chan, H. C., "Radar sea-clutter at low grazing angles," *IEE Proceedings F - Radar and Signal Processing* **137**, 102–112 (Apr. 1990).

[13] Bell, K. L., Baker, C. J., Smith, G. E., Johnson, J. T., and Rangaswamy, M., "Cognitive radar framework for target detection and tracking," *IEEE Journal of Selected Topics in Signal Processing* **9**, 1427–1439 (Dec. 2015).

[14] Haykin, S., "Cognitive radar: A way of the future," *IEEE Signal Processing Magazine* **23**, 30–40 (Jan. 2006).

[15] Kelsey, M., Sen, S., Xiang, Y., Nehorai, A., and Akcakaya, M., "Sparse recovery for clutter identification in radar measurements," in [*Proc. SPIE*], **1021106**, 1–10 (May 2017).

[16] Xiang, Y., Kelsey, M., Wang, H., Sen, S., Akcakaya, M., and Nehorai, A., "A comparison of cognitive approaches for clutter-distribution identification in nonstationary environments," in [*IEEE Radar Conference*], (Apr. 2018).

[17] Ozturk, A., "An application of a distribution identification algorithm to signal detection problems," in [*Proc. 27th Asilomar Conference on Signals, Systems and Computers*], **1**, 248–252 (1993).

[18] Rangaswamy, M., Weiner, D. D., and Ozturk, A., "Non-Gaussian random vector identification using spherically invariant random processes," *IEEE Transactions on Aerospace and Electronic Systems* **29**, 111–124 (Jan. 1993).

[19] Tropp, J. A. and Gilbert, A. C., "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on Information Theory* **53**, 4655–4666 (Dec. 2007).

[20] Tropp, J. A., "Greed is good: Algorithmic results for sparse approximation," *IEEE Transactions on Information Theory* **50**, 2231–2242 (Oct. 2004).

[21] Davis, G., Mallat, S., and Avellaneda, M., "Adaptive greedy approximations," *Constructive Approximation* **13**, 57–98 (Mar. 1997).

[22] Haupt, J., Castro, R., Nowak, R., Fudge, G., and Yeh, A., "Compressive sampling for signal classification," in [*Proc. 40th Asilomar Conference on Signals, Systems and Computers*], 1430–1434 (Oct. 2006).

[23] Wang, H., Vieira, J., Ferreira, P., Jesus, B., and Duarte, I., "Batch algorithms of matching pursuit and orthogonal matching pursuit with applications to compressed sensing," in [*Proc. International Conference on Information and Automation*], 824–829 (June 2009).

[24] Rubinstein, R., Zibulevsky, M., and Elad, M., "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," tech. rep., Computer Science Department, Technion Israel Institute of Technology (Aug. 2008).

[25] Elgammal, A., Duraiswami, R., Harwood, D., and Davis, L. S., "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proc. of the IEEE* **90**, 1151–1163 (July 2002).

[26] Comaniciu, D., "An algorithm for data-driven bandwidth selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 281–288 (Feb. 2003).

[27] Elgammal, A., Duraiswami, R., and Davis, L. S., "Efficient kernel density estimation using the fast Gauss transform with applications to color modeling and tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **25**, 1499–1504 (Nov. 2003).

[28] Jones, M. C., Marron, J. S., and Sheather, S. J., "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association* **91**(433), 401–407 (1996).

[29] Fan, J., Heckman, N. E., and Wand, M. P., "Local polynomial kernel regression for generalized linear models and quasi-likelihood functions," *Journal of the American Statistical Association* **90**(429), 141–150 (1995).

[30] Ward, K. D., Watts, S., and Tough, R. J., [*Sea Clutter: Scattering, the K Distribution and Radar Performance*], IET (2006).