Robust and Efficient Modulation Recognition Based on Local Sequential IQ Features

Wei Xiong, Petko Bogdanov, Mariya Zheleva
Department of Computer Science, University at Albany – SUNY
{*wxiong, pbogdanov, mzheleva*}*@albany.edu*

***Abstract*—Modulation recognition plays a key role in emerging spectrum applications including spectrum enforcement, resource allocation, privacy and security. While critical for the practical progress of spectrum sharing, modulation recognition has so far been investigated under unrealistic assumptions: (i) a transmitter's bandwidth must be scanned alone and in full, (ii) prior knowledge of the technology must be available and (iii) a transmitter must be trustworthy. In reality these assumptions cannot be readily met, as a transmitter's bandwidth may only be scanned intermittently, partially, or alongside other transmitters, and modulation obfuscation may be introduced by short-lived scans or malicious activity.**

This paper bridges the gap between real-world spectrum sensing and the growing body of methods for modulation recognition designed under simplifying assumptions. We propose to use local features, besides global statistics, extracted from raw IQ data, which collectively enable a robust framework for modulation recognition that outperforms baselines from the state-of-the-art. Specifically, we exploit the discriminative power of *local* patterns from consecutive IQ samples extracted based on a Fisher Kernel framework that captures non-linearity in the underlying data. With these domain-informed features, we employ lightweight linear support vector machine classification for modulation detection. Our framework is robust to noise, partial transmitter scans and data biases without utilizing prior knowledge of the underlying transmitter technology. The recognition accuracy of our approach consistently outperforms baselines in both simulated and real-world traces. We demonstrate up to a 98% accuracy and a 30% improvement over several counterparts from the literature with partial scans in a USRP testbed.

I. Introduction

Modulation recognition (modrec) has been of interest since the dawn of wireless communications with critical importance to civil and defence applications [23]. Existing work, however, tackles modrec as a signal decoding task, and requires targeted spectrum scans that excursively focus on a transmitter of interest. These scan requirements are increasingly challenged by emerging spectrum sensing systems, which perform sweep-based wideband spectrum scans, introducing intermittency, partiality and biases in a given transmitter's scan. *This divide between existing modrec approaches and the reality of next generation spectrum measurement demands a rethink of existing moderec methodology.*

Modulation recognition in practice consists of a two-step process: data collection (i.e. spectrum sensing) and data analysis (i.e. recognition). While the quality and quantity of collected data inevitably affects the recognition accuracy, existing modrec approaches are largely disconnected from the underlying spectrum sensing techniques that generate the data necessary for analysis. This disconnect will further widen with the advent of autonomous spectrum sensing and agile transmitter technology. ***Future spectrum sensing infrastructures*** will have to leverage multiple dedicated [2], [21], [30] or crowdsourced [7], [9], [24] spectrum sensors, collecting traces in a wide frequency band. To support this heterogeneous environment, the sensor infrastructures will have to sequentially "step through" the spectrum, while collecting data from contiguous sub-bands [8], [24]. As a result, individual transmitter's activity will be scanned intermittently, with partial coverage of their occupied frequency band and alongside other transmitters or unoccupied spectrum sub-bands. ***Modrec data analysis*** is in essence a classification problem approached via various machine learning techniques from lightweight support vector machines (SVM) [14] to artificial neural networks [23]. Of key importance to the detection speed and accuracy are the features employed for classification which are extracted from raw measured IQ samples. The state-of-the-art feature-based methods employ two families of features: *order statistics (OS)* [16] and *high order cumulants (HOC)* [12], [19], [32]. The former family employs sorted IQ sample components for classification, while the latter extracts high-order statistics from the distribution of samples. *Both families of global features disregard the sequential order of IQ samples, effectively treating them as a "bag" of independent samples. We demonstrate that the information encoded in this local sequential order is crucial for robust modrec in practice.*

In addition, many prior modrec work imposes stringent requirements on the data collection including (i) 100% scan coverage of a transmitter's bandwidth, (ii) 0° rotation of the modulation constellation and (iii) balanced representation of constellation symbols in a spectrum scan. In our evaluation, we observe that relaxing these requirements leads to severe deterioration in modrec performance. In §III-C, we demonstrate a significant sensitivity of HOC and OS features to scan partiality, data bias, and constellation rotation. In particular, individual HOC and OS features converge across modulation types, while their standard deviations increase. These trends cause a dramatic reduction of their discriminative power, which, in turn, leads to poor modrec performance regardless of the utilized classifier. *The mismatch between future spectrum sensing requirements and the assumptions of existing modrec methodology calls for novel data-driven approaches for robust modulation recognition in the face of partial, intermittent, biased or noisy scans.*

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

To bridge this gap, we design a framework that leverages novel features from local patterns in IQ samples for robust modulation recognition with partial, biased and noisy scans. Specifically, we use the phase and amplitude of IQ samples to create ordered subsequences of values, dubbed *shingles*. We represent an IQ sample sequence in terms of its shingles based on a Fisher Kernel generative framework [27], where we quantify gradient statistics for shingles being generated by a Gaussian Mixture Model (GMM) dictionary of prototypical shingles. We train and employ SVM [10] classifier for runtime detection of a transmitter's modulation without prior knowledge of the scan's partiality, transmitter technology, data bias or the channel signal to noise ratio (SNR). We demonstrate robust performance of our method in both a realistic MATLAB simulation and a USRP testbed.

Our paper makes the following key contributions:

- We are the first to propose and demonstrate the potential of local IQ patterns for modulation recognition in future spectrum sensing platforms.
- We design an adaptive Fisher Kernel framework employing lightweight SVM classifiers for robust modrec based on local sequential patterns.
- Our proposed approach exhibits a significant improvement of modrec accuracy over baselines in both realistic simulation and real-world spectrum measurement within a USRP testbed.

II. Related work

Modulation recognition has been an active area of research with two main streams of methodology: likelihood-based (LB) [26] and feature-based (FB) [11]. While optimal, LB approaches suffer high computational complexity and are not resilient to RF chain imperfections (e.g. timing and frequency offset), and wireless channel effects (e.g. non-Gaussian noise) [32]. In addition, LB approaches explicitly rely on a model modulation constellation, which is not always readily available or may be significantly distorted due to small scan overlap with the transmitter, missing or unbalanced constellation symbols and high noise regimes. FB approaches offer a lower complexity alternative and have been heavily utilized in recent modrec literature [3], [12], [14], [16], [19], [32]. FB modrec extracts features from measured IQ data and performs modulation classification based on these features. The state-of-the-art techniques adopt order statistics (OS) [16], high order cumulants (HOC) [12], [14], [32] and kernel density functions [3] as features and employ various classification techniques including support vector machines [14] and artificial neural networks [23]. All the above approaches pose unrealistic requirements to spectrum sensing including 100% transmitter scan overlap with the transmitter, side band exclusion, and are sensitive to the signal's noise level. All methods except [19] assume no bias in symbol representation. These requirements are in direct disagreement with future spectrum sensing infrastructures, which will use dedicated [2], [21], [30] or crowdsourced [7], [9], [24] sensors for wideband intermittent sensing in support of spectrum sharing technology, policy and enforcement. Lu et al. [19] consider modulation recognition from incomplete and biased scans, however, the method employs HOC features which, as we demonstrate in §III (i) are highly-sensitive to scan imperfections, (ii) but encode complementary information to our proposed local features, and thus can be successfully combined in order to boost modrec performance (see §V). Recently, deep neural networks (DNN) have been employed for modrec with promising performance outcomes [25], [29]. Such approaches are orthogonal to our work, as they use simple input data comprised of raw IQ samples while employing complex classifiers. In contrast, we focus on domain-informed feature design, and in this paper, employ lightweight SVM classification, however, our features can be employed in a DNN framework as well.

The discriminative power of local patterns has been demonstrated in various signal domains, including images [15], [31], [34]–[36], video [22], audio [18] and text [6]. The-state-of-the-art techniques employ dictionary learning for feature extraction and various classifiers for classification [31], [36]. A key benefit of local patterns is their resilience to global changes in the underlying signal. As demonstrated in §III, these benefits carry over in the modulation recognition domain, where the misrepresentation of a constellation symbol, higher noise level or partial transmitter overlap inevitably affect global features, but preserve inherent signatures in local IQ patterns.

III. Preliminaries and Limitations of Existing Features for Modulation Recognition

A. Problem formulation and notation

The input to modulation recognition is a set of IQ samples represented as complex numbers of the form $I + iQ$, collected by a sensor at a specified center frequency and bandwidth. We transform each sample into (amplitude, phase) pairs $x = (A, \phi) = (\sqrt{I^2 + Q^2}, \arctan \frac{Q}{I})$. Let $x = ((A_1, \phi_1), (A_2, \phi_2) \ldots, (A_n, \phi_n))$ denote an ordered sequence (series) of samples to which we will also refer as an instance. Given a set of such instances $X = [x^{(1)}, x^{(2)}, \ldots x^{(m)}]$ and the corresponding modulation types employed by the underlying sampled transmitters $y = [y^{(1)}, y^{(2)}, \ldots, y^{(m)}]$, the objective in supervised modulation recognition is to learn a classifier $f(x) = y$ which can predict the modulation type of newly observed instances. All existing feature-based techniques (including ours) do not work directly with the samples x to learn a classifier, but instead extract features from them which are then employed for classification.

B. Global features: order statistics and cumulants

All existing methods treat samples within an instance x as independent and extract features that summarize their statistical properties. There are two main classes of such features, both aiming to summarize the overall distribution of all instance samples, thus, we refer to them as global features.

Higher order cumulants (HOCs) [4]. This approach seeks to summarize the statistical properties of the IQ samples using high order complex cumulants [13]. Within this framework the instance observations are modeled as samples from a complex-valued stationary random process $x(n)$ and high-order cumulants associated with the empirical distribution are

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

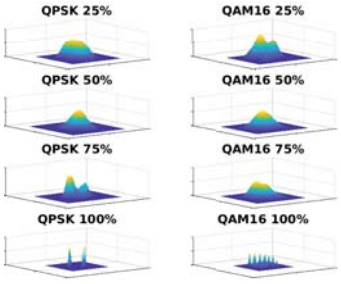


Fig. 1: Effects of partial scan on constellation shape for QPSK (left) and 16-QAM (right).

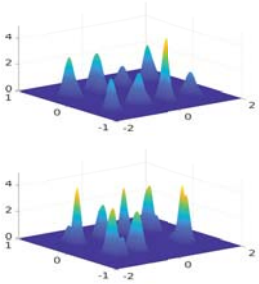
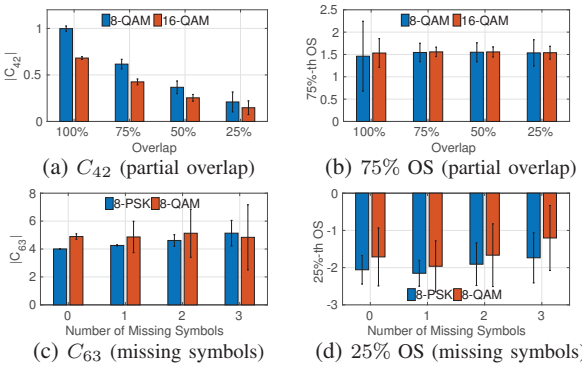


Fig. 2: Bias due to imbalanced (top) or missing symbols (bottom) leads to poor modrec.



(a) C_{42} (partial overlap) (b) 75% OS (partial overlap)
(c) C_{63} (missing symbols) (d) 25% OS (missing symbols)

Fig. 3: Mean and standard deviations of (a) C_{42} with varying scan overlap; (c) C_{63} with missing symbols; (b) the 75% OS with partial overlap; and (d) the 25% OS with missing symbols (all averaged over 100 instances).

estimated and used as predictive features [32]. Subsets of the fourth-order $\{C_{40}, C_{41}, C_{42}\}$ and sixth-order cumulants $\{C_{60}, C_{61}, C_{62}, C_{63}\}$ have received most attention in the modrec literature [4], [12], [14], [32]. These quantities are defined in terms of estimates of moments associated with the empirical IQ sample observations. For example, C_{42} is defined as:

$$C_{42} = M_{42} - |M_{20}|^2 - 2M_{21}^2, \quad (1)$$

where $M_{kv} = E[x(n)^{k-v}x^*(n)^v]$ are the empirical estimates of the moments associated with the stationary process from which the IQ samples are drawn, and $x^*(n)$ denotes the complex conjugation of $x(n)$. We omit the exhaustive definition of all the above cumulants due to space limitations and refer the reader to [4] for details. To remove the effect of the signal scale on cumulants, they are typically normalized by C_{21} [32]: $\hat{C}_{kv} = C_{kv}/(C_{21})^{k/2}$.In addition, since some cumulants are complex numbers, their L_2 is adopted as a real feature in classification [14].

Order statistics (OS) [16]. The k-th order statistic of a random real sample is its k-th smallest value. This simple notion gives rise to a modrec approach proposed in [16] which employs the ordered values of the amplitude A, phase ϕ and the baseband I and Q components derived from an observed sample sequence x. OS features offer an alternative global summary of the distribution of the IQ samples. Note that in this representation the order of IQ samples is lost, however, as we demonstrate, this order contains information that can be used to discriminate modulations in realistic scenarios.

C. Limitations of global feature approaches

While the two families of global features discussed above have been successfully employed by many recent modrec approaches, they inherently rely on assumptions about (i) the overlap of the sensing range with the underlying transmitter's frequency range; (ii) the balance of observed symbols in a sample; and (iii) the phase offset (or constellation rotation). When these assumptions are relaxed in practical modrec "in the wild", the discriminative power of the global features deteriorates. In what follows, we analyze the robustness of HOC and OS with respect to the above assumptions in order to quantify and understand their limitations.

1) Effects of partial scan overlap with the transmitter. In sweep-based spectrum sensing, a transmitter may only be scanned partially as the exact frequency range may not be available a priori. *What is the effect on the shape of the modulation constellation, and thus, the global feature modrec performance when a transmitter is only partially scanned?* Fig. 1 illustrates qualitatively this effect for QPSK and 16-QAM. The constellations in both cases transition from sets of well-pronounced symbol clusters at 100% overlap to a single (or multi-modal) at lower overlap, from which the original constellation is hard to recover. This visual deterioration is also reflected in a decreased stability of both HOC and OS features (Fig. 3). The averages of C_{42} for 8-QAM and 16-QAM converge, while their standard deviations increase (Fig. 3a). The same trend is observed for the 75%-th order statistic (Fig. 3b) and for other HOC and OS features (omitted due to space limitations). As a result the discriminative power of this feature decreases with the overlap, thus, reducing its utility regardless of the adopted classifier.

2) Effects of bias in instance samples. Bias in the instance symbols may arise due to scan intermittency (i.e. insufficient IQ samples to obtain uniform symbol representation) or malicious transmitters which purposefully obfuscate the constellation to deceive modrec [28]. Biases both due to small number of samples (Fig. 2, top) or missing symbols (Fig. 2 bottom) affect the overall constellation, and similar to partial scans, have a negative impact on the discriminative power of global features. Fig. 3c and 3d show the behavior of C_{63} and the 25% OS for 8-PSK and 8-QAM with increasing number of randomly missing symbols. The respective feature values converge between modulation types, while their variance increases drastically with increasing number of missing symbols. Once again, this behavior suggests a deteriorating discriminative power of global features with missing symbols, further evaluated in §V.

3) Effects of constellation rotation. Global OS features assume 0° rotation of the modulation's constellation [16], i.e. prior knowledge of the transmitter's technology. This may not be available when sensing arbitrary agile transmitters in the wild, once again negatively affecting the performance of global feature-base modrec (details in §V).

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

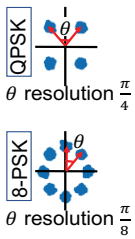


Fig. 4: Example phase transitions. θ=$\pi/8$ does not occur in QPSK.

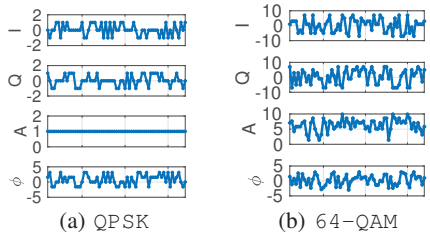


Fig. 5: LP of Polar and Cartesian representation of IQ samples over time (x-axis). Both magnitude and shape of the LPs vary across modulations.

IV. Methodology: Discriminative Local Features

Sample bias, partial overlap of the scan with the transmitter, constellation rotation and increasing noise levels all distort the global statistical properties of instances, thus deteriorating the quality of corresponding global feature modrec approaches. At the same time, these challenges are ubiquitous when the problem is considered in realistic settings. To improve the robustness of modrec techniques, we propose to capture information contained in the local ordering of IQ samples. The resulting local features, when combined with global HOC features, enable a robust and efficient modrec, exhibiting superior accuracy in both simulated and real-world scans. In what follows, we summarize the intuition behind our proposed local features and specify the methodology for their extraction.

A. Intuition: Local features encoding the order of IQ samples

While the order and relationships of individual IQ samples within an instance x has not been considered in the modrec literature, we postulate that it carries important information, which is better preserved in realistic settings and can be used to improve modrec accuracy. This intuition is inspired by the tremendous success of local features extracted from images in computer vision and particularly employed for natural image classification [31], [35]. In our case, we treat an instance x as a 1-dimensional signal as opposed to the typical 2D setting arising in computer vision.

Consider Fig. 4 which presents the constellations of QPSK and 8−PSK modulations. Intuitively, an instance x is a trajectory of transitions between constellation clusters. The distribution of angular distances (angle changes) between consecutive transitions arising from different modulations varies due to the varying inter-cluster distances in their constellations. For example, transitions in 8−PSK will be centered around multiples of $\pi/8$, while those in QPSK around multiples of $\pi/4$. Fig. 5 depicts a segment of the IQ sequences arising from QPSK and 64−QAM. We plot separately the I and Q components as well as the corresponding amplitude A and phase ϕ sequences. Qualitatively, it is evident that high-order modulations (i.e. 64−QAM) exhibit bigger variation compared to their low-order counterparts. Furthermore, the phase transitions of low-order modulations are sharp from sample to sample, while these transitions are smoother with higher-order modulations.

To capture the sequential information encoded in IQ sample subsequences, we focus on intervals of their time domain and propose to extract modulation-specific transition signatures. We demonstrate that, such signatures are more robust to noise, sample bias, constellation rotation and partial transmitter overlap than global features alone, and thus can be employed to improve modrec accuracy.

B. Local sequential features based on the Fisher Kernel

Let $x_A, x_\phi \in \mathbb{R}^n$ denote the real-valued sequences of observed amplitude and phase values in an instance x. We employ the same framework to extract local features from each of those sequences separately, as they might be advantageous for different modulation types. For example, the amplitude sequence x_A will be discriminative for amplitude-related modulation properties (e.g. PSK v.s. QAM families), while the phase sequence x_ϕ will be useful to differentiate the order of the modulation (e.g. QPSK v.s. 8−PSK). The same framework can be applied to the sequences of I and Q components, however, in our experimental evaluation they did not offer additional discriminative power. In what follows, we will simplify the notation by denoting $x = [x_1, x_2, \ldots, x_n]$ as either of the real-valued sequences x_A or x_ϕ.

We adopt a generative framework to model a sequence x in terms of all of its subsequences of length l, to which we will refer as *shingles*. Specifically, let x_i^l denote a shingle starting at position i of length l. An instance x of length n has a total of $n - l + 1$ such shingles. Our key assumption is that observed instance shingles are generated from some parametric generating distribution p_λ parametrized by a set of parameters λ. This representation is similar to n-gram based models for text [6] and patch-based representations for images [31].

We adopt a *Gaussian Mixture Model (GMM)* as the generating distribution p_λ, which is a typical choice in patch-based representation of images [31]. A K-component GMM is fully specified by $\lambda = \{w_k, \mu_k, \Sigma_k\}, k = 1 \ldots K$, where $w_k \geq 0$ is the non-negative mixing weight of the k-th component and μ_k and Σ_k are its mean vector and covariance matrix respectively. We further disregard mixed covariance terms for shingles and instead work with a variance vector σ_k^2 (i.e. we assume a diagonal covariance matrix). This assumption is justified in our case as consecutive constellation symbols within shingles are determined by the encoded data and we do not place any assumptions on their sequence. Note that the shingle size l determines the dimensions of μ_k and σ_k^2.

We adopt the *Fisher Kernel (FK)* representation which defines similarities between shingles in terms of dot products of their *Fisher Vectors (FVs)* [17]. Formally, a FV $f_\lambda(x_i^l)$ representing shingle x_i^l is defined as:

$$f_\lambda(x_i^l) = L_\lambda \nabla_\lambda \log p_\lambda(x_i^l), \tag{2}$$

where $\nabla_\lambda \log p_\lambda(x_i^l|\lambda)$ is the gradient of the log-likelihood of the observed shingle x_i^l being generated by p_λ, where the gradient is evaluated at x_i^l; and L_λ is the square root of the inverse of the *Fisher Information Matrix (FIM)*. L_λ normalizes the dynamic range of gradient vectors similar to its use in [17].

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

We obtain the local feature representation $f_\lambda(x)$ of the whole instance x, given a GMM model p_λ, as the average Fisher Vector of all observed shingles within the instance:

$$f_\lambda(x) = \frac{1}{n-l+1} \sum_{i=1}^{n-l+1} f_\lambda(x_i^l). \tag{3}$$

In other words, the instance FV is the average of the normalized gradient statistics of all involved shingles, where L_λ is treated as a normalization factor. We apply the same transformation to both the amplitude x_A and phase x_ϕ sequences and concatenate the resulting FVs. In what follows, we discuss how to derive the normalization L_λ and gradient statistics $\nabla_\lambda \log p_\lambda(x_i^l|\lambda)$ for individual shingle FVs.

The likelihood $p_\lambda(x_i^l)$ in GMM is defined as the average weighted likelihood of the shingle x_i^l arising from the individual Gaussian components:

$$p_\lambda(x_i^l) = \sum_{k=1}^{K} w_k p_k(x_i^l), \tag{4}$$

where $p_k(x_i^l)$ is the pdf of the k-th l-variate Gaussian component in the GMM. To ensure that $p_\lambda(x_i^l)$ is a valid probability distribution the weights need to be all non-negative and sum to 1, i.e. $\sum_{k=1}^{K} w_k = 1$. We use the gradient statistics with respect to the mean μ_k and variance σ_k vectors of each component resulting in the following component-wise L_λ-normalized gradients:

$$f_{\mu_k}(x) = \frac{\nabla_{\mu_k} \log p_\lambda(x)}{\sqrt{w_k}} = \frac{\gamma_k(x)}{\sqrt{w_k}} \left[\frac{x-\mu_k}{\sigma_k^2}\right] \tag{5}$$

$$f_{\sigma_k}(x) = \frac{\nabla_{\sigma_k} \log p_\lambda(x)}{\sqrt{w_k}} = \frac{\gamma_k(x)}{\sqrt{w_k}} \left[\frac{(x-\mu_k)^2}{\sigma_k^3} - \frac{1}{\sigma_k}\right], \tag{6}$$

where the $\gamma_k(x)$ is the soft assignment (posterior probability) of the shingle to component k defined as:

$$\gamma_k(x) = \frac{w_k p_k(x)}{\sum_{i=1}^{K} w_i p_i(x)}, \tag{7}$$

and where exponentiation and division operations involving vectors x, μ_k and σ_k in Eqs. 5, 6 are element-wise operations (recall that they are l-dimensional vectors). Note that we do not consider the gradient statistic with respect to w_k in our FV representation, arriving at a $(4lK)$-dimensional vector, representing 2 series (amplitude and phase), maintaining shingle-length (i.e. l-dimensional) gradient statistics (both mean and variance) for each of the K components of the GMM. We omit a gradient statistic with respect to the component weights w_k which could be interpreted as prior component probabilities, as they require more data to robustly estimate (GMM model estimation is discussed next), than their variance and mean vectors. Investigation of whether these additional statistics boost the performance might be a fruitful further direction.

An illustrative example of the gradient evaluation for a shingle x in a two-component mixture model with unit variance vectors is presented in Fig. 6. The well-agreeing GMM components result in close-to optimal corresponding gradient statistics. The final FV is composed of concatenating $[\nabla_{\mu_1} \nabla_{\mu_2} \nabla_{\sigma_1} \nabla_{\sigma_2}]$ (normalization by $1/\sqrt{1/2}$ omitted).

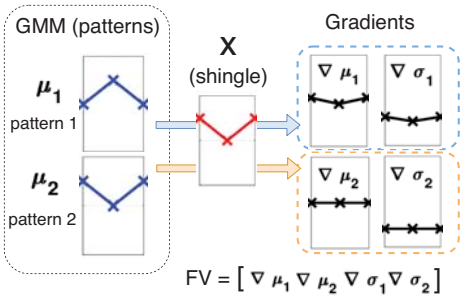


Fig. 6: Example of local features computed for $l = 3$-dimensional shingle x and 2-component GMM with μ_2 agreeing "better" with the shingle than μ_1 ($\sigma_{1,2} = \mathbb{1}$, $w_{1,2} = 0.5$). The gradient statistics for each of the components are shown on the right. The mean gradient statistics of the "better"-agreeing component ∇_{μ_1} is closer to $\mathbb{0}$ (expected as per Eq. 5) and its ∇_{σ_1} is closer to $-\mathbb{1}$ (as per Eq. 6).

It is worth noting that, while resorting to a kernel method for representation of our local features as opposed to working directly with component likelihoods $p_k(x)$, results in higher dimensional representation, it comes with the usual advantages. Namely, when the kernel is appropriately selected, it allows modelling non-linear data using simple linear classifiers. In addition, the specific FV kernel has been shown to perform very well in the natural images domain and typically requires small number of Gaussian components for good discriminative power, thus allowing good scalability [31]. We experimented with non-kernel local feature representations and did not find similar improvements over state-of-the-art global feature methods as the ones exhibited by the FV kernel representation.

C. Model learning: GMM dictionary and classification

To enable modrec employing our local features scheme, we need to first estimate a GMM model from shingle observations in actual instances and then train a modulation classifier based on the feature encoding of instances.

GMM dictionary learning. The FV representation outlined in §IV-B depends on a GMM generating distribution for shingles. Intuitively, we need to learn a "dictionary" of prototypical shingles, observed in instances across modulations and learn their component-wise mean μ_k, variances σ_k and relative weights w_k. Given a fixed dictionary size K and a shingle length l, we learn a GMM based on a training data set X containing instances of all modulation classes we aim to predict. Note, that since we do not use the class information y associated with instances in X, our dictionary GMM learning is unsupervised. Supervised alternatives may allow even sparser discriminative representations for classification [20], however, we leave this direction for future exploration. To learn the GMM model from a training set X we first extract shingles from the instances and use the seminal Expectation Maximization (EM) approach [5]. Details about selecting the dictionary size K and shingle length l are discussed in §V.

Classification. As we discuss earlier, the advantage of the FK methodology is that it captures non-linear information in its representation, and hence, simple classification techniques

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

are expected to perform well. Thus, we adopt a simple linear SVM classifier with soft margin for our modrec task [10]. We expect that other classification schemes may further improve the classification performance, but resort to a simple SVM in this work as our goal is to evaluate the utility of our local features and also employ a classifier which is typically employed by baseline global feature methods.

Our local feature scheme captures local transition information, however, we expect that the global sample distribution statistics may encode additional non-redundant information and thus consider classification schemes in which we concatenate the fisher vector f_λ with the 7 HOC features widely adopted in prior work. This combination is expected to "lift" the modrec performance of local features alone, particularly when the dictionary is learned on a rotated constellation w.r.t. that used in testing instances. We confirm this expectation empirically in §V.

D. Algorithmic complexity

Both the dictionary learning process and classifier training do not need to be repeated during actual modulation recognition, as long as they are performed on a training set that features instances from all target modulations. Thus, both processes can be thought of as "offline", i.e. they do not occur during actual modrec at work.

The complexity of modrec with our employed local features is the cost of encoding shingles from an instance x. Asymptotically, it depends on the dictionary and shingle sizes and and the number of samples instances $O(nlK)$, as there are $O(n)$ shingles in an instance and their gradient statistics of size l need to be evaluated with respect to each of the K GMM dictionary components. In practice we resort to short $l = 3$ shingles and small dictionary size $K = 50$ as they show optimal performance. Thus, assuming that K and l are constants relative to the number of samples n, the complexity of local patterns is linear $O(n)$ similar to that for computing HOC [32] and asymptotically better than OS (when using all samples) [16] which require sorting the samples in $O(n \log n)$.

V. Evaluation

We evaluate the robustness of our methodology with partial, biased and noisy scans in over-the-air and simulated settings. We begin by describing our implementation and data sets. In §V-B-§V-D we evaluate the modrec performance of our method compared to state of the art HOC [32] and OS [16]. For these results, we vary the classifier training, while using a universally-trained dictionary, as described in §V-A. Unless otherwise noted, all accuracy results were obtained as an average from a 10-fold validation.

A. Implementation, data and parameters

Implementation. Our method is implemented in MATLAB with all experiments executed on Ubuntu 14 machines. The Fisher Vector dictionary learning module is implemented using [33]. For classification, we adopt the SVM classifier model from MATLAB. We use one-versus-rest label coding to transform multi-classification to binary classification. We use the same classification approach across all compared features (i.e. HOC, OS, LP and LP+HOC).

Data. We use two datasets for our evaluation: one generated in a MATLAB simulation and one from a USRP testbed. For our simulation we use MATLAB Communications System Toolbox to implement a transmitter and receiver connected by a AWGN channel. The transmitter is configured to use `QPSK`, `8-PSK`, `8-QAM`, `16-QAM` and `64-QAM`. We tune various blocks of our transmitter-receiver chain to generate the necessary datasets as follows. For *partial scans*, we tune the low-pass filter at the receiver side by setting its cut-off to a fraction of the transmitter's bandwidth. For *biased scans*, we purposely modify the input signal at the transmitter side to reduce or remove the occurrence of a given symbol. To control the *noisiness* of the collected scan we tune the SNR level of the AWGN channel. Finally, to control the *constellation rotation*, we modify the modulation block at the transmitter side. Our simulated datasets are used for results presented in §V-B-§V-D. We also present results from partial scans from USRP-based transmissions, as detailed in §V-F.

Default parameters. All performance results presented in §V-B-§V-F were obtained with a single universal dictionary of patches trained at SNR 10dB, with no data bias, at 100% transmitter overlap with mixed constellation rotation. The patch size, is set to 3 and the dictionary size to 50. A natural question is whether the dictionary learning parameterization (i.e. how we set the patch and dictionary size) and training data play role in our algorithm's performance. We explore this question in §V-G and show that the above universally-trained dictionary is feasible across all real-world settings.

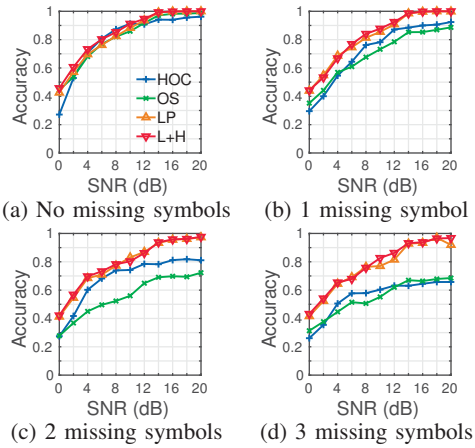
B. Robustness to data bias

We begin by evaluating our method with data bias. All scans were collected at 100% coverage of the transmitter bandwidth. We train the classifier on data with equal representation of all constellation symbols (i.e. unbiased data). A separate classifier was trained for each SNR level (i.e. classification is SNR-aware). We then test using data with purposely removed 1, 2 or 3 symbols. Fig. 7 presents our results. For unbiased data (Fig. 7a) all methods perform similarly. As bias is introduced, methods using global features deteriorate immediately even with SNR of 20dB. At 3 missing symbols global features can achieve a maximum of 69% accuracy at 20dB, whereas our method maintains high accuracy of 98%. Table I (left) shows a breakdown of performance of LP+HOC across modulations at SNR=10dB. For low-order modulations the accuracy is maximal and decreases as modulation order increases. These results demonstrate the potential of LP+HOC to successfully detect a transmitter's modulation in the face of data bias.

C. Robustness to scan partiality

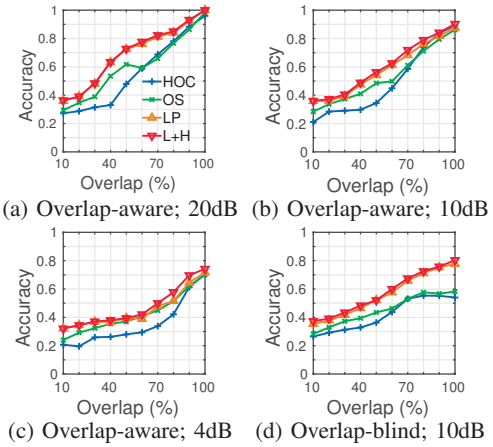
We evaluate the performance of our method with scan partiality. We first focus on performance, where the training and testing of the classifier are overlap-aware, meaning that a different classifier is trained at each partial overlap. Fig. 8a-8c present modrec accuracy for 20, 10 and 4dB, respectively. Our method (LP+HOC) persistently outperforms

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.



(a) No missing symbols (b) 1 missing symbol
(c) 2 missing symbols (d) 3 missing symbols

Fig. 7: Performance with missing symbols. Classifier is SNR-aware and trained on unbiased data. Scans cover 100% of the transmitter's bandwidth.



(a) Overlap-aware; 20dB (b) Overlap-aware; 10dB
(c) Overlap-aware; 4dB (d) Overlap-blind; 10dB

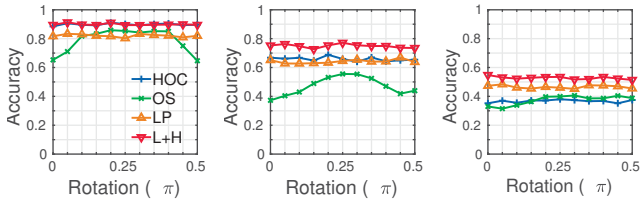
Fig. 8: Performance on partial scans with overlap-aware (a-c) and overlap-blind (d) classifier training.

existing counterparts across all SNR regimes. For a 100% scan, our method performs on par with the literature for high SNR regimes (20 and 10dB) and outperforms the state in low SNR regimes (4dB). Table I (right) presents a breakdown of LP+HOC accuracy across modulations at SNR=20dB. Our method maintains high accuracy for low-order modulations even when a transmitter's bandwidth is scanned only at 50%. The accuracy with high order modulations deteriorates as the overlap decreases. These results demonstrate that LP+HOC (i) leads to better modrec performance across all partial scans and (ii) is robust in noisy channel conditions.

Prior knowledge of the scan partiality in the classifier training phase poses a practical challenge to the real-world applicability of our method, as one needs to determine the fraction at which a transmitter is scanned before recognizing its modulation. Thus, we consider overlap-blind classification in which the classifier is trained on a mix of all possible partial scans as opposed to at every overlap individually. Fig. 8d shows our results for overlap-blind modrec at a challenging SNR of 10dB. Global features (HOC and OS) alone fail in classification even at a 100% coverage. LP+HOC and LP, on the other hand, outperform their global counterparts across

TABLE I: LP+HOC accuracy with scan bias (L) and partial overlap (R).

	Bias at 10dB, #missing symbols				Overlap at 20dB, %			
	0	1	2	3	100	90	80	70
4-PSK	1.00	1.00	0.87	1.00	1.00	1.00	1.00	0.99
8-PSK	1.00	1.00	0.98	0.94	1.00	1.00	1.00	0.97
8-QAM	1.00	0.97	0.83	0.84	1.00	1.00	1.00	1.00
16-QAM	0.77	0.66	0.65	0.61	1.00	0.81	0.64	0.63
64-QAM	0.74	0.74	0.71	0.70	1.00	0.86	0.64	0.56



(a) 100% overlap (b) 75% overlap (c) 50% overlap

Fig. 9: Phase-blind classification of non-biased scans at SNR 10dB.

all partial overlap. Furthermore, in comparison with the the overlap-aware modrec performance at 10dB (Fig. 8b), our method only suffers a marginal performance deterioration when overlap-blind classifier training is employed. These results demonstrate our method's applicability in the wild without the need of prior knowledge of the partiality of a transmitter's scan.

D. Robustness to constellation rotation

We evaluate the performance of our approach with unknown constellation rotation. We use spectrum scans at a challenging SNR regime of 10dB. We vary the constellation rotation from 0 to $\pi/2$ while scanning a transmitter bandwidth at 100%, 75% and 50%. We train our classifier on a mix of constellation rotations. Fig. 9 presents our results. For 100% scan HOC and LP+HOC perform equally well, whereas OS and LP have lower accuracy. As the transmitter overlap decreases, our method maintains maximal performance, whereas all other counterparts suffer dramatic deterioration in modrec accuracy due to their susceptibility to constellation rotation.

E. Robustness to noise

All results so far were obtained with a SNR-aware classifier, meaning that a separate classifier was trained for each SNR level. This approach requires prior knowledge of the channel, which while feasible, adds steps and computational overhead to the modrec procedure. To address this issue, we explore SNR-blind modrec, where the classifier is trained on a mix of instances at different SNR levels. Specifically, we consider SNR levels from 0 to 20dB in increments of 2. At each SNR level we generate 1000 instances and train the classifier on the mix of these instances. We then test at each SNR level.

Fig. 10 shows the classification accuracy across SNR. We compare our proposed method LP+HOC with HOC [32] and OS [19]. For the HOC features, we use all fourth order and sixth order cumulants. For the OS features, we use the amplitude and the phase order statistics. The same linear SVM classifier is used across all of HOC, OS and LP+HOC. The results show that our method is persistently able to achieve maximal performance across all SNR regimes. HOC alone suffers severe performance deterioration across all SNR levels,

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

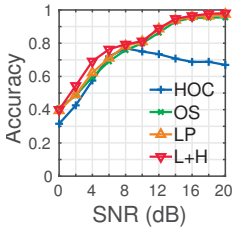


Fig. 10: SNR-blind classification of 100% overlap, no bias.

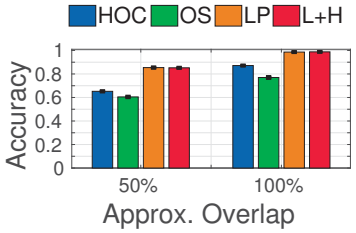
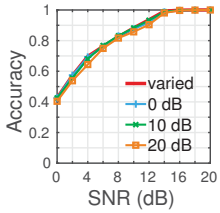


Fig. 11: Modrec from partial scans in a USRP testbed.



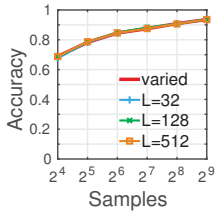


Fig. 12: Dictionary learning performance over SNR (left) and instance size (right) for four training mechanisms. DL does not require explicit training for each SNR or instance size.

while OS performs on par with our method at high SNR and slightly worse in low SNR.

F. Modrec from partial scans in the wild

Finally, we evaluate the performance of our method in real over-the-air transmissions from a USRP-based testbed. We use a transmitter comprised of a USRP B210 attached to an Intel i7-5600U CPU host, and a receiver comprised of a USRP B210 with an Intel i7-6700 CPU host. Both hosts are running on low-latency Linux kernel. Using `GNURadio` [1], the transmitter generates a signal modulated with `BPSK`, `QPSK`, `8PSK`, `QAM16` and `QAM64`. The two USRP devices are located in line of sight at a distance of roughly 8 inches. We set the transmitter gain as 65 and receiver gain as 10. At the transmitter side, we use a sample rate of 320 kHz. Based on our observation we use the following bandwidth at the receiver side as roughly 50% coverage overlap: 40 kHz(`BPSK`), 80 kHz(`QPSK`), 120 kHz(`8PSK`), 120 kHz(`QAM-16`) and 160 kHz(`QAM-64`). Scans collected at the receiver were stored in a file and later analyzed. Fig. 11 presents our results. At 100% transmitter overlap, the proposed LP+HOC method achieves 98% accuracy, outperforming the HOC based classifier by 12%. At 50% overlap, both features suffer performance deterioration. Proposed feature classifiers fall to 85% accuracy but still outperform the HOC classifier, which is at 65% accuracy. These results show our method's applicability and robustness to real-world modulation recognition with partial scans.

G. Effects of dictionary learning on modrec accuracy

So far, all experimental results were obtained with a single universally-trained dictionary as detailed in §V-A. In this section we evaluate the feasibility of such a universal dictionary across various realistic scenarios. Of key interest is whether the training data (i.e. SNR level, scan overlap and bias) and parameter setting (i.e. dictionary and patch size) play a role in modrec performance. For this analysis, we generate a synthetic data set using our simulator (§V-A) with 100% transmitter overlap, no data bias and mixed constellation rotation. We consider ten SNR levels (0-20dB in increments of 2) and five modulations: `QPSK`, `8-PSK`, `8-QAM`, `16-QAM` and `64-QAM`. For each modulation and SNR level we generate $m = 1000$ training instances of $n = 512$ each and another 500 testing instances of the same size. The SVM classifier was SNR-aware and trained on 100% scans with no data bias.

1) Should dictionary learning be SNR-, overlap-, and bias-aware? We adopt four training approaches in the dictionary learning phase: (1) *Varied*: we train and test the dictionary for each SNR level, (2) *0dB*: we train the dictionary at SNR=0dB and test at each level, (3) *10dB*: we train the dictionary at SNR=10dB and test at each level and (4) *20dB*: we train the dictionary at 20dB and test at each level. Fig. 12 (left) presents modrec accuracy for the four training strategies across a range of SNR values from 0 to 20dB. The classification performance remains the same across training approaches, which indicates that we can only train the dictionary once, at any SNR level and the learned patterns will be applicable across any SNR level. Similar conclusions can be drawn from our results with biased and partially-overlapping scans (results omitted in interest of space). This is particularly important to the real world applicability of our approach, as it demonstrates its robustness to various real-world conditions.

2) Does the instance length affect the dictionary learning performance? The instance length is defined in terms of the number of `IQ` samples that appear in a measured sequence. We adopt four dictionary training approaches: (1) *Varied*: we train and test the dictionary for each instance length, (2) *L=32*: we train the dictionary at instance length of 32 and test at all instance lengths, (3) *L=128*: we train the dictionary at instance length of 128 and test at all instance lengths and (4) *L=512*: we train the dictionary at instance length 512 and test on all instance lengths. Fig. 12 (right) presents modrec accuracy over increasing instance length for the four training schemes. We see that the classification performance remains the same across the four training approaches, indicating that we do not need to retrain the dictionary as our instance length changes.

3) Do dictionary learning parameters affect the modrec accuracy? Two key parameters of the dictionary learning step are the shingle size l and the number of components K for the GMM instantiation. We explore the performance with various (l, K) combinations ($l = 2, 3, 5$, $K = 20, 50, 100$) with SNR varying from 0 to 10dB in increments of 2. In interest of space we omit a figure and summarize our results as follows. We observe maximal performance across all SNRs for shingle size of 2 or 3, which deteriorates at $l = 5$. Similarly, a GMM instantiation with 20 or 50 components leads to good accuracy, however, the accuracy deteriorates with $K = 100$. Thus, we choose to use a shingle size of 3 and a dictionary size of 50.

H. Discussion

Our evaluation shows the merit of combining local and global features for robust modrec in the wild. Our hierarchical modrec approach LP+HOC outperforms methods based on

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.

global features by a large margin across all the explored realistic scenarios. A counterpart based on LP-only closely follows the performance of LP+HOC, however, it is not as robust in the face of arbitrary constellation rotation.

VI. Conclusion

We designed a novel modulation classification framework which is robust to imperfect spectrum scan data due to encoding local sequential patterns within `IQ` samples. We employed a Fisher Kernel representation which flexibly handles non-linearity in the underlying data and enables high-quality modulation recognition even with simple linear classification models such as linear soft-margin Support Vector Machines. We demonstrated our framework's applicability on real-world, partial, intermittent, biased and noisy scans. Our method consistently outperformed state-of-the-art approaches, and in addition our local features were demonstrated to encode complementary information to global alternatives. Thus, our framework can be effectively combined with existing features to further boost its individual recognition accuracy.

Our work addresses a critical disconnect between modrec requirements and spectrum sensing capabilities. Particularly, it addresses critical challenges posed by emerging spectrum-sharing technologies which will employ heterogeneous dedicated or crowdsourced sensors, scanning a wide frequency band sequentially, and thus producing intermittent, partial and noisy scans. The superior performance of our methodology on over-the-air partial scans indicates its potential for improved modrec in the wild. Our proposed framework and its inter-operability with previous approaches constitutes a solid foundation for future work on spectrum analytics with practical importance to future spectrum sharing technology, enforcement and security.

VII. Acknowledgements

This work was supported through NSF CISE Research Initiation Initiative (CRII) grant CNS-1657476 and NSF Smart and Connected Communities (SC&C) grant CMMI-1831547.

References

[1] GNURadio. https://www.gnuradio.org/.

[2] Microsoft's Spectrum Observatory. https://observatory.microsoftspectrum.com/.

[3] H. Abuella and M. K. Ozdemir. Automatic modulation classification based on kernel density estimation. *Canadian Journal of Electrical and Computer Engineering*, 39(3):203–209, 2016.

[4] M. W. Aslam, Z. Zhu, and A. K. Nandi. Automatic modulation classification using combination of genetic programming and KNN. *IEEE Trans. on Wireless Communications*, 11(8):2742–2750, 2012.

[5] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg, 2006.

[6] W. B. Cavnar, J. M. Trenkle, et al. N-gram-based text categorization. *Ann Arbor, MI*, 48113(2):161–175.

[7] A. Chakraborty, A. Bhattacharya, S. Kamal, S. R. Das, H. Gupta, and P. M. Djuric. Spectrum patrolling with crowdsourced spectrum sensors. In *IEEE INFOCOM*, Honolulu, HI, 2018.

[8] A. Chakraborty, U. Gupta, and S. R. Das. Benchmarking resource usage for spectrum sensing on commodity mobile devices. In *ACM HotWireless*, New York, NY, 2016.

[9] A. Chakraborty, M. S. Rahman, H. Gupta, and S. R. Das. Specsense: Crowdsensing for efficient querying of spectrum occupancy. In *IEEE INFOCOM*, Atlanta, GA, 2017.

[10] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[11] O. A. Dobre, A. Abdi, Y. Bar-Ness, and W. Su. Survey of automatic modulation classification techniques: classical approaches and new trends. *IET Communications*, 1(2):137–156, 2007.

[12] O. A. Dobre, Y. Bar-Ness, and W. Su. Higher-order cyclic cumulants for high order modulation classification. In *IEEE MILCOM*, Boston, MA, 2003.

[13] J. Eriksson, E. Ollila, and V. Koivunen. Statistics for complex random variables revisited. In *IEEE ICASSP*, Taipei, Taiwan, 2009.

[14] H. Gang, L. Jiandong, and L. Donghua. Study of modulation recognition based on HOCs and SVM. In *IEEE VTC Spring*, Milan, Italy, 2004.

[15] S. Gao, I. W.-H. Tsang, L.-T. Chia, and P. Zhao. Local features are not lonely – Laplacian sparse coding for image classification. In *IEEE CVPR*, San Francisco, CA, 2010.

[16] L. Han, F. Gao, Z. Li, and O. A. Dobre. Low complexity automatic modulation classification based on order-statistics. *IEEE Trans. on Wireless Communications*, 16(1):400–411, 2017.

[17] T. Jaakkola and D. Haussler. Exploiting generative models in discriminative classifiers. In *NIPS*, Denver, CO, 1999.

[18] A. Kumar and B. Raj. Weakly supervised scalable audio content analysis. In *IEEE ICME*, Seattle, WA, 2016.

[19] G. Lu, K. Zhang, S. Huang, Y. Zhang, and Z. Feng. Modulation recognition for incomplete signals through dictionary learning. In *IEEE WCNC*, pages 1–6, San Francisco, CA, 2017.

[20] J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach. Supervised dictionary learning. In *NIPS*, Vancouver, B.C., Canada, 2009.

[21] M. A. McHenry, P. A. Tenhula, D. McCloskey, D. A. Roberson, and C. S. Hood. Chicago Spectrum Occupancy Measurements and Analysis and a Long-term Studies Proposal. In *ACM TAPAS*, Boston, MA, 2006.

[22] I. Mironica, B. Ionescu, J. Uijlings, and N. Sebe. Fisher kernel based relevance feedback for multimodal video retrieval. In *ACM ICMR*, Dallas, TX, 2013.

[23] A. K. Nandi and E. E. Azzouz. Modulation recognition using artificial neural networks. *Signal processing*, 56(2):165–175, 1997.

[24] A. Nika, Z. Li, Y. Zhu, Y. Zhu, B. Y. Zhao, X. Zhou, and H. Zheng. Empirical validation of commodity spectrum monitoring. In *ACM SenSys*, Stanford, CA, 2016.

[25] T. J. OShea, T. Roy, and T. C. Clancy. Over-the-air deep learning based radio signal classification. *IEEE Journal of Selected Topics in Signal Processing*, 12(1):168–179, 2018.

[26] P. Panagiotou, A. Anastasopoulos, and A. Polydoros. Likelihood ratio tests for modulation classification. In *IEEE MILCOM*, Los Angeles, CA, 2000.

[27] F. Perronnin, J. Sánchez, and T. Mensink. Improving the Fisher kernel for large-scale image classification. In *Proc. ECCV*, Crete, Greece, 2010.

[28] H. Rahbari and M. Krunz. Full frame encryption and modulation obfuscation using channel-independent preamble identifier. *IEEE Trans. on Information Forensics and Security*, 11(12):2732–2747, 2016.

[29] S. Rajendran, W. Meert, D. Giustiniano, V. Lenders, and S. Pollin. Deep learning models for wireless signal classification with distributed low-cost spectrum sensors. *IEEE Trans. on Cognitive Communications and Networking*, 4(3):433–445, 2018.

[30] S. Roy, K. Shin, A. Ashok, M. McHenry, G. Vigil, S. Kannam, and D. Aragon. Cityscape: A metro-area spectrum observatory. In *IEEE ICCCN*, Vancouver, B.C., Canada, 2017.

[31] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek. Image classification with the Fisher vector: Theory and practice. *International journal of computer vision*, 105(3):222–245, 2013.

[32] A. Swami and B. M. Sadler. Hierarchical digital modulation classification using cumulants. *IEEE Trans. on Communications*, 48(3):416–429, 2000.

[33] A. Vedaldi and B. Fulkerson. Vlfeat: An open and portable library of computer vision algorithms. In *Proc. ACM MM*, Firenze, Italy, 2010.

[34] K. Yu and T. Zhang. Improved local coordinate coding using local tangents. In *ICML*, Haifa, Israel, 2010.

[35] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. Local features and kernels for classification of texture and object categories: A comprehensive study. *International journal of computer vision*, 73(2):213–238, 2007.

[36] Q. Zhang and B. Li. Discriminative K-SVD for dictionary learning in face recognition. In *Proc. IEEE CVPR*, San Francisco, CA, 2010.

Authorized licensed use limited to: Technische Universitaet Muenchen. Downloaded on May 12,2021 at 09:21:15 UTC from IEEE Xplore. Restrictions apply.