# Fusing Sorted Random Projections for Robust Texture and Material Classification

Li Liu, Paul W. Fieguth, *Member, IEEE*, Dewen Hu, Yingmei Wei, and Gangyao Kuang

***Abstract*— This paper presents a conceptually simple, and robust, yet highly effective, approach to both texture classification and material categorization. The proposed system is composed of three components: 1) local, highly discriminative, and robust features based on sorted random projections (RPs), built on the universal and information-preserving properties of RPs; 2) an effective bag-of-words global model; and 3) a novel approach for combining multiple features in a support vector machine classifier. The proposed approach encompasses the simplicity, broad applicability, and efficiency of the three methods. We have tested the proposed approach on eight popular texture databases, including Flickr Materials Database, a highly challenging materials database. We compare our method with 13 recent state-of-the-art methods, and the experimental results show that our texture classification system yields the best classification rates of which we are aware of 99.37% for Columbia–Utrecht, 97.16% for Brodatz, 99.30% for University of Maryland Database, and 99.29% for Kungliga Tekniska högskolan-textures under varying illumination, pose, and scale. Moreover, the proposed approach significantly outperforms the current state-of-the-art approach in materials categorization, with an improvement to classification accuracy of 67%.**

***Index Terms*— Data fusion, kernel methods, materials textures, random projection (RP), rotation invariance, support vector machines (SVMs), texture classification.**

## I. Introduction

TEXTURE is a fundamental characteristic of the appearance of virtually all natural surfaces and constitutes a powerful visual cue. Texture classification plays an important role in the computer vision and pattern recognition, and has a wide range of applications including content-based image retrieval, medical image analysis, remote sensing, industrial inspection, document segmentation, and terrain classification for mobile robot navigation [1]–[3].

Manuscript received October 30, 2013; revised April 27, 2014 and July 24, 2014; accepted September 15, 2014. Date of publication September 19, 2014; date of current version March 3, 2015. This work was supported in part by the National Natural Science Foundation of China under Contract 61202336 and in part by the Doctoral Fund, Ministry of Education, China, under Contract 20124307120025. This paper was recommended by Associate Editor S. Sull.

L. Liu and Y. Wei are with the School of Information System and Management, National University of Defense Technology, Changsha 410073, China (e-mail: dreamliu2010@gmail.com; weiyingmei126@126.com).

P. W. Fieguth is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: pfieguth@uwaterloo.ca).

D. Hu is with the School of Mechatronics and Automation, National University of Defense Technology, Changsha 410073, China (e-mail: dwhu@nudt.edu.cn).

G. Kuang is with the School of Electronic Science and Engineering, National University of Defense Technology, Changsha 410073, China (e-mail: feiyunlyi@hotmail.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCSVT.2014.2359098

The field of texture classification has moved toward the challenging problem of classifying real world textures with 3-D variations, with the variations due to imaging environment changes, including both geometrical and photometric. Consequently, more challenging texture databases, such as the Columbia–Utrecht (CUReT) database [4], became popular in which a texture class became a set of images corresponding to a single physical sample, captured under a variety of imaging conditions in the laboratory.

More recently, researchers have been working to push texture classification into everyday materials imaged in real-world contexts. Specifically, efforts are being made to extend texture classification to material recognition, the accurate classification of entire material categories with each category having many *different* physical material samples [6], [7].

Frequently, the problem of material categorization has been treated as a texture classification problem. However, the material–texture distinction is tricky and the two tasks are somewhat different: the notion of a category is the key. The texture classification task used in much of the literature is essentially to match an unseen image to the other (training) images obtained from the same physical sample. In material classification, categories group similar items, but similarity does not imply a same sample, rather to broad high-level categories, such as paper, plastic, fabric, and wood. Moreover, often it is not even clear where boundaries between categories should necessarily be drawn. There exist relatively few image databases that focus on materials, with recent examples, including Kungliga Tekniska högskolan-textures under varying illumination, pose, and scale (KTH-TIPS2) [8], [9] and the Flickr Materials Database (FMD) [6]. KTH-TIPS2 has four physical samples photographed in various pose, illumination, and scale conditions, however, it contains rather few physical samples in each material category. Instead, the recently created FMD was designed for the material categorization problem, possesses a significant number of categories and images, and presents a substantial challenge to classification.

In this paper, we propose to investigate these two problems of texture classification and material categorization under a common kernel-based discriminative framework. Recently, the bag-of-words (BoWs) approach has emerged to become the dominant paradigm for texture classification [8], [11]–[16], [21]–[23]. The key components in building a BoW-based classification system includes:

1) local texture features;
2) nonlocal statistical representation of local features;
3) an effective classifier;
4) a suitable similarity measure used within the classifier.

It is generally agreed that local highly discriminative yet robust texture features play the more important role [11]–[13], [15], [23] and many kinds of local texture descriptors have been proposed, such as Leung and Malik filter bank [10], 8 maximum responses (MR8) [12], Gabor, local binary pattern (LBP) [15], Patch [13], scale-invariant feature transform (SIFT) [23], rotation-invariant feature transform (RIFT) [11], spin (SPIN) [11], random projection (RP) [16], and sorted RP (SRP) [17]. With no single feature necessarily capturing all of the relevant local information, some sort of feature combining seems relevant.

Of the possible features to combine, the strongest are the random features proposed in [16] and [17]. In [16], RPs were developed, a universal, information-preserving, dimensionality-reduction technique to project from the image patch domain to a compressed space without loss of salient information. The method is very simple, and yet was shown to outperform patch features, LBP, and various filter bank-based methods. The rotation sensitivity of the RP approach led to further work in which Liu *et al.* [17] proposed SRP features, for rotation-invariant texture classification, which demonstrated striking classification performance.

This paper investigates the problem of fusing multiple SRP feature channels for the purposes of efficient image level texture classification and material categorization, problems possessing significant intra-class variability. The proposed approach represents images as distributions (signatures[1] or histograms) of SRP features and learns a support vector machine (SVM) classifier with kernels based on selecting effective measures for comparing distributions, and using multiple kernel learning (MKL) as an effective method for combining the base kernels.

In this paper, we show how texture classification and material categorization can be simultaneously addressed within a robust combined SRP framework, and reach the following conclusions.

1) Fused SRP features do lead to an increase in classification accuracy, so clearly the different SRP features possess complementary information.
2) In contrast to the excellent comparative study of [23], the histograms-$\chi^2$-SVM framework is more efficient and effective than the signatures-earth mover's distance (EMD)-SVM framework.
3) The large intra-class variation in the FMD material database has a devastating effect on the performance of other texture classification techniques, however, the proposed approach gives excellent material categorization performance on the FMD database, significantly better than the state of the art. Experiments clearly show the effectiveness of the proposed model selection strategy.
4) The extensive experimental results demonstrate that the proposed approach can effectively classify texture images under a variety of conditions, producing consistently good classification results, including what we believe to be the best reported results for the CUReT, Brodatz, KTH-TIPS, University of Maryland Database (UMD), and FMD databases.

This paper is organized as follows. In Section II, we give a brief review of background and related work. In Section III, we present the proposed combined SRP approach in detail. In Section IV, methods are developed which maximize the material classification performance and minimize the number of models used to characterize the material categories. In Section V, we evaluate the proposed approach against various state-of-the-art alternatives on commonly used texture and material databases. Our preliminary work published in [24] and [25] acted as a basis for this research. Our work in [24] and [25] only deals with the texture classification problem, however, the current paper also addresses the much more challenging problem of material recognition. Moreover, in this paper, methods are developed which minimize the number of models used to characterize the various material classes.

## II. Background and Related Work

### A. Texture Classification

Texture classification methods can loosely be divided into two categories: the one is a sparse approach that first detects certain salient regions in a given image and then applies local descriptors, such as SIFT, RIFT, and SPIN to describe the selected regions [11], [23], [26]; the other is a dense approach that extracts local features pixel-by-pixel over the input image [10], [12]–[16]. Relative to dense methods, sparse approaches are complex, produce relatively high-dimensional features, and lack stability. Therefore, the dense approach is more common and widely studied [10], [12]–[20].

As more challenging texture classification problems have been studied, researchers have built richer representations by combining multiple types of complementary texture descriptors. Ojala *et al.* [15] proposed to combine two complementary operators, LBP and variance, as a more powerful approach for rotation-invariant texture classification. Liao *et al.* [27] proposed an improved texture method by combining dominant local binary pattern and Gabor filters. Lazebnik *et al.* [11] developed a sparse affine-invariant texture representation which first applies two complementary local region detectors to detect salient local texture regions and then uses two different local descriptors, SPIN and RIFT, both of which are high dimensional, to extract local texture features. Following the work of Lazebnik *et al.* [11], Zhang *et al.* [23] presented an approach by combining three local descriptors, SIFT, RIFT, and SPIN, claiming better performance over single descriptor. Xu *et al.* [28] designed a multifractal spectrum (MFS) approach by combining local features of intensity and first- and second-order Gaussian derivatives. Finally, a state-of-the-art material categorization method [6] combines multiple features including texture, color, shape, and edge features in a Bayesian framework.

### B. Material Categorization

Material recognition is closely related to, but different from, texture recognition. Unlike the more developed field of texture

[1]A signature summarizes a feature distribution in the form of a set of cluster centers together with the fractional cluster weights indicating the relative size of each cluster.

classification, there exists relatively little work in the field of material categorization, surprising given how important material understanding is for humans and how important it must become for intelligent robot vision [29].

Material perception has been studied by modeling the underlying reflectance properties of materials using the bidirectional reflectance distribution function and bidirectional texture function [4], [30]. Material recognition and object recognition are often perceived as closely related problems and their solutions share many properties in common [23]. Other interesting examples, in which material and object recognition interact, include the work of [31] and [32].

Unlike some other visual recognition tasks in computer vision, it is difficult to find good, reliable features for material recognition, therefore, the state of the art in material recognition [6] combines a rich set of seven low- and mid-level features that capture various aspects of material appearance with their proposed augmented latent Dirichlet allocation (aLDA) method. Nevertheless, this method achieves a classification rate of only 45% on the newly published challenging material benchmark database FDM, clearly demonstrating the limitation of current methods. Sharan [7] have established that human being can recognize material categories reliably, quickly in challenging conditions, and that categorization performance cannot be explained by simple, low-level cues like color or texture. It is therefore of significant interest to explore questions regarding the best features for material recognition.

### C. Random Projection

RP has become a widely used method for dimensionality reduction, and has been shown to have promising theoretical properties: it is a general data reduction technique, such that the choice of RP matrix does not depend upon the data in any way, and RPs have been shown to have special promise for high-dimensional data clustering.

The key idea of RP arises from the Johnson–Lindenstrauss (JL) lemma [33], [34], which states that a set $\mathcal{D}$ of $d$ points in $\mathbb{R}^{n\times 1}$ with $n$ typically large can be linearly projected into a lower dimensional Euclidean space $\mathbb{R}^m$ using a random orthonormal matrix, while approximately preserving the relative distances between any two of these points. Subsequent research [35], [40] simplified the proof of the above result by showing that such a projection can be generated using an $m\times n$ matrix $\mathbf{\Phi}$, whose entries are randomly drawn from certain probability distributions [35], specifically including the Gaussian distribution [35], [40]. In recent years, the JL lemma has found numerous applications that include the compressed sensing (CS) problem [36], searching for approximate nearest neighbors in high-dimensional Euclidean space and dimension reduction in databases [35], and learning mixtures of Gaussians [34].

The information-preserving and dimensionality-reduction power of RP is firmly demonstrated by the theory of CS [37], [38], which has grown out of the surprising realization that for sparse and compressible signals, a small number of linear nonadaptive measurements in the form of RPs can capture most of the salient information in the high-dimensional signal and allow for accurately reconstruction. The basic problem in CS is to design a stable measurement matrix to obtain the minimal number of linear nonadaptive measurements that allows for stable reconstruction of the original signal [37]–[39]. A sufficient condition for stable reconstruction of the original signal [39] is that the measurement matrix obeys a condition known as the *restricted isometry property* (RIP).[2] Baraniuk *et al.* [40] give a simple technique for verifying the RIP property for random matrices that underlies CS, clearly illustrate that the RIP can be thought of as a consequence of the JL lemma, and that any distribution that yields a satisfactory JL-embedding will also generate matrices satisfying the RIP. As a consequence, random Gaussian projections approximately preserve pairwise distances in the dataset.

### D. Kernel SVM Classification

Discriminative and robust texture features appear as the most important fact contributing to superior texture classification performance [1], [2], however, given a good texture feature the key is to find a suitable classifier. SVM are currently the most popular classifier in BoW due to its robustness against large feature vectors [41].

The provable success of SVM for image classification and object recognition motivates the study of its potential in texture classification. Although SVM have previously been used in texture classification [8], [42], they have been limited to a single kernel, and the success of SVM in these cases is often dependent on the choice of a good kernel.

Because a great many texture features have been proposed, with no method obviously outperforming the others, some aspect of combining features seems relevant. Recent advances in MKL have positioned it as an attractive tool for tackling many learning tasks. The idea of combining descriptors has been explored [9], [11], [23], [43] in texture classification and material categorization. The method of Varma and Garg [43] is based on MKL, where they attempted to learn optimal combinations of local texture features and where they demonstrated better classification performance. However, there is little evidence that the multiple features used by Varma and Garg represent complementary information about texture.

## III. Multiple SRP Classifiers

### A. Local Texture Feature Extraction: SRPs

The local SRP features, as shown in the left panel of Fig. 1, were first proposed in [17] for rotation-invariant texture classification. The SRP takes the sorted raw pixel intensities [Fig. 1(a)] or intensity differences [SRP Radial-Diff and SRP Angular-Diff in Fig. 1(b) and (c)] in a circular neighborhood to form a feature vector $\underline{\boldsymbol{x}}$ (i.e., $\underline{\boldsymbol{x}}^{\mathrm{Circ}}$, $\underline{\Delta}^{\mathrm{Rad}}$, or $\underline{\Delta}^{\mathrm{Ang}}$), which is then transformed to a lower dimensional vector $\underline{\boldsymbol{y}} = \mathbf{\Phi}\underline{\boldsymbol{x}}$ by

[2]For each integer $k$, define the isometry constant $\delta_k$ of a matrix $\mathbf{\Phi}$ as the smallest number such that $(1-\delta_k)\|\underline{\boldsymbol{x}}\|_{l_2}^2 \le \|\mathbf{\Phi}\underline{\boldsymbol{x}}\|_{l_2}^2 \le (1+\delta_k)\|\underline{\boldsymbol{x}}\|_{l_2}^2$ holds for all $k$-sparse vectors. If $0 < \delta_k < 1$, matrix $\mathbf{\Phi}$ is said to satisfy the RIP condition of order $k$. A vector $\boldsymbol{x}$ is said to be $k$-sparse if it has at most $k$ nonzero entries [39].

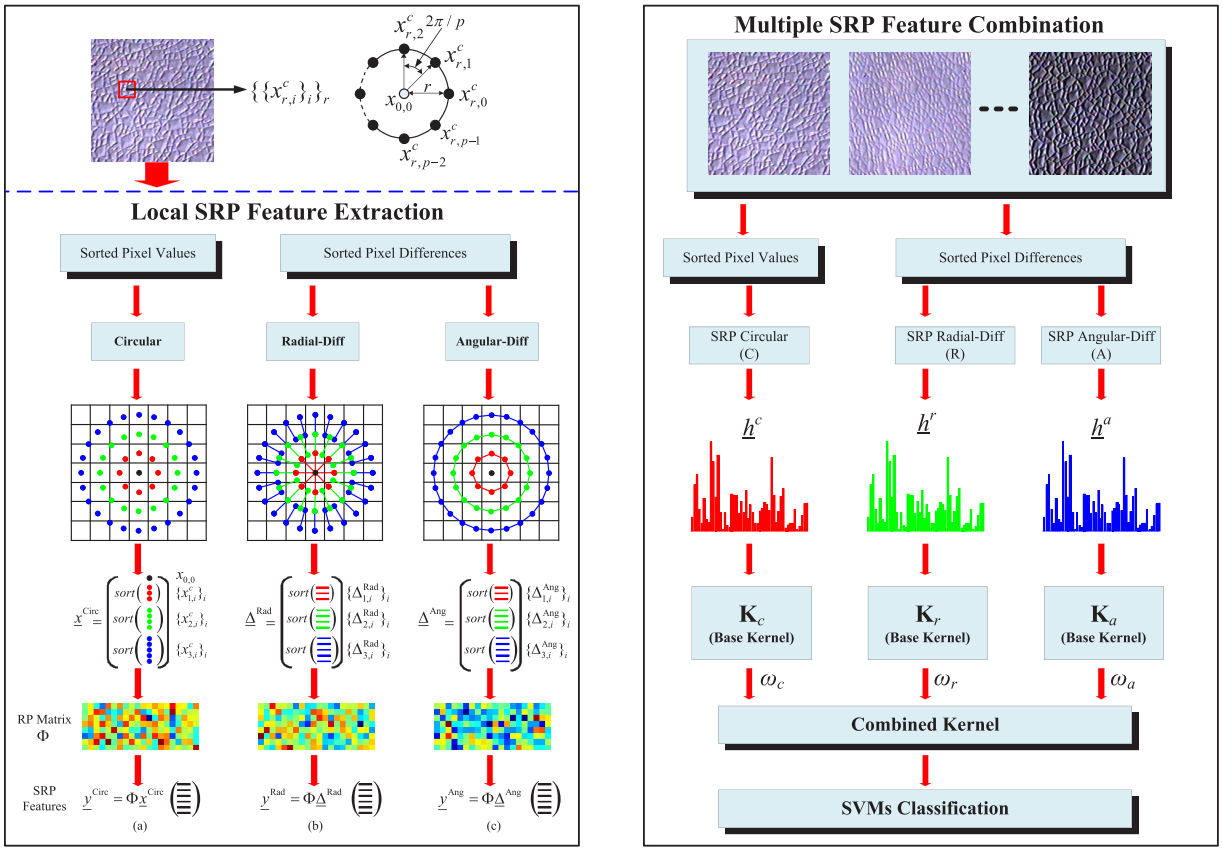


Fig. 1. Overall framework for the proposed approach. Left panel: extracting SRP features on an example local image patch of size $7 \times 7$. (a) Sorting pixels. (b) and (c) Sorting pixel differences. Right panel: architecture of combining multiple SRP features: we start with three base descriptors and compute the corresponding BoW features $\underline{\boldsymbol{h}}^c$, $\underline{\boldsymbol{h}}^r$, $\underline{\boldsymbol{h}}^a$, and associated distance functions (e.g., $\chi^2$). The BoW histogram features and distance functions are then kernelized to yield base kernels matrices $\mathbf{K}^c$, $\mathbf{K}^r$, and $\mathbf{K}^a$. Given the base kernels, the combined descriptor's kernel is approximated such as $K^* = \omega_c K^c + \omega_r K^r + \omega_a K^a$, where $\omega$ denotes the weights corresponding to each base kernel. The optimization is carried out in an SVM framework so as to achieve the best classification on the training set.

a RP matrix $\boldsymbol{\Phi}$ whose entries are independently sampled from a zero-mean, unit-variance normal distribution.

Specifically, as shown in Fig. 1(a), the $\underline{\boldsymbol{x}}^{\text{Circ}}$ feature extracted from a local neighborhood of size $(2a+1) \times (2a+1)$ centered at pixel $x_{0,0}$ is defined as

$$\underline{\boldsymbol{x}}^{\text{Circ}} = \left[x_{0,0}, sort\left(\left[x^c_{1,0}, \ldots, x^c_{1,p_1-1}\right]\right), \ldots, sort\left(\left[x^c_{a,0}, \ldots, x^c_{a,p_a-1}\right]\right)\right]^T \quad (1)$$

where $\{x^c_{r,i}\}_{i=0}^{p_r-1}$ denotes the values of the $p_r$ neighbors of $x_{0,0}$ on a circle of radius $r$, as shown at the top of Fig. 1. Those locations which do not fall exactly in the center of a pixel are estimated by interpolation. The function $sort(\cdot)$ sorts the inputs in nondescending order. Similarly, the sorted radial and angular difference features, $\underline{\Delta}^{\text{Rad}}$ and $\underline{\Delta}^{\text{Ang}}$, as shown in Fig. 1(b) and (c), are computed as

$$\underline{\Delta}^{\text{Rad}} = \left[sort\left(\Delta^{\text{Rad}}_{1,0}, \ldots, \Delta^{\text{Rad}}_{1,p_1-1}\right), \ldots, sort\left(\Delta^{\text{Rad}}_{a,0}, \ldots, \Delta^{\text{Rad}}_{a,p_a-1}\right)\right]^T \quad (2)$$

$$\underline{\Delta}^{\text{Ang}} = \left[sort\left(\Delta^{\text{Ang}}_{1,0}, \ldots, \Delta^{\text{Ang}}_{1,p_1-1}\right), \ldots, sort\left(\Delta^{\text{Ang}}_{a,0}, \ldots, \Delta^{\text{Ang}}_{a,p_a-1}\right)\right]^T \quad (3)$$

where $\Delta^{\text{Rad}}_{r,i} = x^c_{r,i} - x^c_{r-1,i}$, $\Delta^{\text{Ang}}_{r,i} = x^c_{r,i} - x^c_{r,i-1}$, $x^c_{r,i}$, and $x^c_{r-1,i}$ correspond to the gray values of pairs of pixels of the same radial direction.

Ojala *et al.* [15] proposed the LBP descriptor, which is based signed differences, but which contains less textural information than the SRP. The motivations in using intensity difference descriptors are as follows.

1) Its computational simplicity, in contrast to SIFT, RIFT, and SPIN descriptors [11], [23].
2) Its gray scale invariance, very attractive in situations where the gray scale is subject to changes due to varying illumination conditions, but maintaining contrast sensitivity.
3) The signed difference space is even sparser than the patch space, and thus allows the use of RP.
4) Pairwise pixel interactions carry important structural information, and both short- and long-range interactions are relevant.

Since the intensity-based SRP feature and difference-based SRP features are somewhat similar to the SPIN and RIFT descriptors, motivated by the work of Lazebnik *et al.* [11] and Zhang *et al.* [23] who proposed combining those descriptors which capture complementary information, we propose to combine the three SRP descriptors for texture classification, with the expectation that combined SRP features would be richer and more robust than a single one.

### B. Global Representation: Histograms Versus Signatures

Given some number of local random features extracted from local texture patches, the key to effective texture classification is to represent the feature distributions to learn the nonlocal behavior of the texture, for which there are two basic methods: *histograms* and *signatures*. We wish to compare these approaches.

Let $\mathcal{D} = \{\{\mathbf{I}_{c,t}\}_{t=1}^{T}\}_{c=1}^{C}$ denotes the whole texture dataset, with $C$ distinct texture classes and each class having $T$ texture samples. Let $\mathcal{Y}_{c,t} = \{\underline{\boldsymbol{y}}_{c,t,i}\}_i$ denotes the random feature vector set extracted from the corresponding texture sample $\mathbf{I}_{c,t}$, and let $\mathcal{Y}_c = \{\{\underline{\boldsymbol{y}}_{c,t,i}\}_i\}_{t=1}^{T_1}$ denotes the random feature vector set extracted from all the training samples available for class $c$, where $T_1$ is the number of training samples per class.

*1) Histogram Framework:* A global texton dictionary learning process is needed, so a set of random feature vectors $\mathcal{Y}_c$, aggregated over training samples from texture class $c$, are clustered by $k$-means, resulting in $K$ cluster centers $\mathcal{W}_c = \{\underline{\boldsymbol{w}}_{c,k}\}_{k=1}^{K}$ which are the so-called textons. The textons learned from the $C$ classes form the global texton dictionary $\mathcal{W} = \{\{\underline{\boldsymbol{w}}_{c,k}\}_{k=1}^{K}\}_{c=1}^{C}$. A histogram $\underline{\boldsymbol{h}}_{c,t}$ of compressed textons is learned for each particular training sample $\mathbf{I}_{c,t}$ by labeling each feature vector extracted from its pixels with the closest texton. The texture class then is represented by a set of normalized histogram models $\mathcal{H}_c = \{\underline{\boldsymbol{h}}_{c,t}\}_t$, corresponding to the training samples of that class. A classifier needs only to be able to assess the degree of dissimilarity between two histograms, for example, using a $\chi^2$ statistic

$$D(\underline{\boldsymbol{h}}_1, \underline{\boldsymbol{h}}_2) = \frac{1}{2}\sum_k \frac{[\underline{\boldsymbol{h}}_1(k) - \underline{\boldsymbol{h}}_2(k)]^2}{\underline{\boldsymbol{h}}_1(k) + \underline{\boldsymbol{h}}_2(k)}. \tag{4}$$

*2) Signature Framework:* In contrast to the histogram approach, the signature framework has no global texton dictionary learning stage. Instead, each image is represented as a signature $S_{c,t} = \{(p_{c,t,i}, \underline{\boldsymbol{u}}_{c,t,i})\}_{i=1}^{K}$, which is learned for each training sample $\mathbf{I}_{c,t}$ by clustering only $\mathcal{Y}_{c,t} = \{\underline{\boldsymbol{y}}_{c,t,i}\}_i$, where $K$ is the number of clusters, $\underline{\boldsymbol{u}}_{c,t,i}$ is the center of the $i$th cluster, and $p_{c,t,i}$ is the cluster frequency vector by counting how many of the pixels belong to cluster $\underline{\boldsymbol{u}}_{c,t,i}$. The EMD is used to measure the dissimilarity between signatures that are compact representations of distributions. The EMD between two signatures $S_1 = \{(p_i, \underline{\boldsymbol{u}}_i)\}_{i=1}^{K_1}$ and $S_2 = \{(q_j, \underline{\boldsymbol{v}}_j)\}_{j=1}^{K_2}$ is defined as

$$D(S_1, S_2) = \frac{\sum_{i=1}^{K_1}\sum_{j=1}^{K_2} \hat{f}_{ij}\, d(\underline{\boldsymbol{u}}_i, \underline{\boldsymbol{v}}_j)}{\sum_{i=1}^{K_1}\sum_{j=1}^{K_2} \hat{f}_{ij}} \tag{5}$$

where $d(\underline{\boldsymbol{u}}_i, \underline{\boldsymbol{v}}_j)$ is the so-called *ground distance* between cluster centers $\underline{\boldsymbol{u}}_i$ and $\underline{\boldsymbol{v}}_j$, for which we use the Euclidean distance, and $\hat{f}_{ij}$ is the optimal *flow*, which can be determined by solving a linear programming problem. While the EMD works very well on signatures it should not, in general, be applied to histograms. Small histogram invalidate the ground distance as the bin centers are rather far, while computing the EMD on large histograms can be very slow.

To compare the performance of the histogram and signature frameworks, Table I shows the classification results for the SRP Radial-Diff feature, results which are representative of other tested features. In contrast to the previous findings in [23], we can see that the histogram framework performs consistently and significantly better than the signature framework. Furthermore, the signatures framework with the EMD measure is time consuming, becoming computationally prohibitive when the number of texture classes and the number of codewords increases [44]. Motivated by these results, the rest of this paper employs the histogram framework.

TABLE I

CLASSIFICATION ACCURACY (%) OF DIFFERENT KERNELS FOR THE SRP RADIAL-DIFF FEATURE ON FIVE TEXTURE DATASETS. FOR THE HISTOGRAM-BASED CLASSIFIERS, THE NUMBER OF TEXTONS $K$ USED PER CLASS ARE 10, 40, 40, 10, AND 40, RESPECTIVELY, WHILE FOR THE SIGNATURE-BASED CLASSIFIERS $K = 40$. THE NUMBER OF TRAINING SAMPLES PER CLASS ARE 46, 20, 20, 3, AND 41, RESPECTIVELY. THE TESTED PATCH SIZE IS $13 \times 13$ FOR ALL DATABASES

| Paradigm | Histograms | | | Signatures | |
|---|---|---|---|---|---|
| Classifier | $\chi^2$+NNC | RBF+SVM | $\chi^2$+SVM | EMD + NNC | EMD+SVM |
| $\mathcal{D}^{C}$ | 96.52 | 97.72 | **99.05** | N/A | N/A |
| $\mathcal{D}^{\text{UIUC}}$ | 95.43 | 97.18 | **98.08** | 88.14 | 93.28 |
| $\mathcal{D}^{\text{UMD}}$ | 98.26 | 98.53 | **98.67** | 92.97 | 96.08 |
| $\mathcal{D}^{\text{B}}$ | 94.73 | 94.24 | **96.04** | 91.38 | 92.67 |
| $\mathcal{D}^{\text{KT}}$ | 97.35 | 98.72 | **99.02** | 95.28 | 95.78 |

### C. SVM Classification

The benefits of SVM for histogram-based classification has clearly been demonstrated in [8], [9], [23], and [45]. SVM was originally designed for binary classification, however, the texture classification problem is multiclass, for which there are two basic strategies: *one-against-one* and *one-against-other*. Hsu and Lin [46] showed *one-against-one*, which trains a classifier for each possible pair of classes, to be more suitable for practical use, therefore, the *one-against-one* technique is used in this paper.

Based on two-class SVM, our multiclass texture classification problem is constructed by cascading multiple standard two-class SVM classifiers. Given a texture classification problem involving $C$ textures, each two-class SVM classifier is trained via two distinct types of texture images, meaning that there are $C(C-1)/2$ two-class SVM classifiers.

Our features for classification are bag-of-visual-words histograms $\{\underline{\boldsymbol{h}}_i\}_i$ of the texture images. In a two-class case, say classes $k$ and $l$, let $\underline{\boldsymbol{h}}_i^{kl}$ be a histogram feature vector and $z_i^{kl} \in \{-1, +1\}$ its class label. After a training phase a new sample, $\underline{\boldsymbol{h}}$, is assigned to one of the two classes by the following decision function:

$$f^{kl}(\underline{\boldsymbol{h}}) = \operatorname{sgn}\left(\sum_{k\in\Omega^{kl}} \alpha_i^{kl} z_i^{kl} \mathbf{K}(\underline{\boldsymbol{h}}, \underline{\boldsymbol{h}}_i^{kl}) + b^{kl}\right)$$

where $\Omega^{kl}$ is the set of support vector indices, $\mathbf{K}(\underline{\boldsymbol{h}}, \underline{\boldsymbol{h}}_i^{kl})$ the kernel, $\alpha_i^{kl}$, $i \in \Omega^{kl}$ the learned weight of $\underline{\boldsymbol{h}}_i^{kl}$, and $b^{kl}$

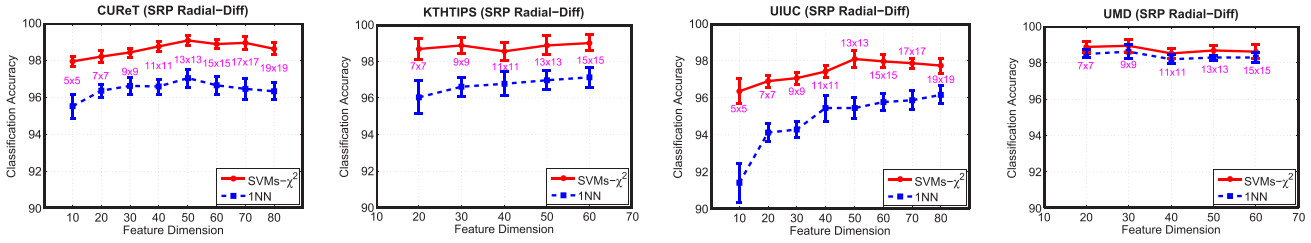


Fig. 2. Comparison of SVM and NNC across different databases. The classification accuracy is plotted as a function patch size for the SRP Radial-Diff feature on four texture databases: (a) CUReT, (b) KTH-TIPS, (c) UIUC, and (d) UMD. The number of training images per class is 46, 41, 20, and 20, respectively. The number of textons K used per class is 10, 20, 40, and 40, respectively. The results convincingly argue that SVM consistently outperforms NNC.

a learned threshold parameter. After all $C(C-1)/2$ classifiers are constructed, a plurality voting strategy is used to assign the feature histogram to a class.

For the histogram framework we apply two types of kernels: Gaussian radial basis function (RBF) kernel $\mathbf{K}(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j) = \exp(-\gamma \|\underline{\boldsymbol{h}}_i - \underline{\boldsymbol{h}}_j\|^2)$ and the $\chi^2$ kernel $\mathbf{K}(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j) = \exp(-\gamma D(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j))$, where $D(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j)$ is defined in (4).

Fig. 2 shows the experimental results of the SRP Radial-Diff feature on four popular and challenging texture databases, with the purpose of comparing SVM with a nearest neighbor classifier (NNC). From our experiments, we observe that the SVM approach significantly and consistently outperforms NNC across all tested datasets, suggesting that SVM are superior to NNC in texture classification, consistent with previous findings [8], [9], [23] that favor the use of SVM for texture classification.

Since a key ingredient for the success of SVM is the kernel function, which is often chosen by the user and represents a heuristic element in our approach, to test the robustness of SVM with respect to the kernel function we ran an extensive set of experiments benchmarking three kernels, plus a baseline comparison with $\chi^2$-NNC and EMD-NNC, as shown in Table I. The tabulated results clearly show the histogram-$\chi^2$-SVM to be the strongest method, therefore the rest of this paper will focus on this context.

### *D. Combining Multiple SRP Features*

We have made a case for SRP features over other texture features, for histograms over signatures, and for SVM over NNC, thus we are now in place to develop a combined multiple SRP feature/SVM classifier, as shown in Fig. 1. To be sure, there is an established literature on MKL [47], [48], however, given that training a single-kernel SVM is already computationally expensive, we have concerns regarding the computational complexity of MKL methods. Instead, since simple kernel combination methods are capable of reaching the same classification accuracy as MKL [47], [48], we are highly motivated to use such methods for texture classification, therefore, in this paper, we propose a new method for combining multiple random features.

Given $F$ feature descriptors, we represent each texture sample using $F$ BoWs histograms. In this paper, we propose to combine the normalized histograms $\underline{\boldsymbol{h}}^{\mathrm{Rad}}$, $\underline{\boldsymbol{h}}^{\mathrm{Circ}}$, and $\underline{\boldsymbol{h}}^{\mathrm{Ang}}$ for the corresponding SRP Radial-Diff, Circular, and Angular-Diff features into a combined histogram $\underline{\boldsymbol{h}}^{\mathrm{Comb}}$ for each texture image. We do have some freedom here because we do not need to give equal weight to each histogram feature. Similar to the strategy adopted in [49], we introduce a weight vector $\omega > 0$, so that the combined histogram is $\underline{\boldsymbol{h}}^{\mathrm{Comb}} = [(\omega_r \underline{\boldsymbol{h}}^{\mathrm{Rad}})^T (\omega_c \underline{\boldsymbol{h}}^{\mathrm{Circ}})^T (\omega_a \underline{\boldsymbol{h}}^{\mathrm{Ang}})^T]^T$. Instead of giving a fixed weight to each histogram feature, we learn the weights $\omega$ which give the best classification performance. Since the distance between two histograms (single or combined) is measured using the $\chi^2$ distance (4), the distance between two combined histograms $\underline{\boldsymbol{h}}_i^{\mathrm{Comb}}$ and $\underline{\boldsymbol{h}}_j^{\mathrm{Comb}}$ is computed as

$$\begin{aligned}
&\chi^2\left(\underline{\boldsymbol{h}}_i^{\mathrm{Comb}}, \underline{\boldsymbol{h}}_j^{\mathrm{Comb}}\right) \\
&= \frac{1}{2}\sum_{k=1}^{CK}\left\{\frac{\left[\omega_r \underline{\boldsymbol{h}}_i^{\mathrm{Rad}}(k) - \omega_r \underline{\boldsymbol{h}}_j^{\mathrm{Rad}}(k)\right]^2}{\omega_r \underline{\boldsymbol{h}}_k^{\mathrm{Rad}}(k) + \omega_r \underline{\boldsymbol{h}}_j^{\mathrm{Rad}}(k)} + \frac{\left[\omega_c \underline{\boldsymbol{h}}_i^{\mathrm{Circ}}(k) - \omega_c \underline{\boldsymbol{h}}_j^{\mathrm{Circ}}(k)\right]^2}{\omega_c \underline{\boldsymbol{h}}_k^{\mathrm{Circ}}(k) + \omega_c \underline{\boldsymbol{h}}_j^{\mathrm{Circ}}(k)}\right. \\
&\qquad \left. + \frac{\left[\omega_a \underline{\boldsymbol{h}}_i^{\mathrm{Ang}}(k) - \omega_a \underline{\boldsymbol{h}}_j^{\mathrm{Ang}}(k)\right]^2}{\omega_a \underline{\boldsymbol{h}}_k^{\mathrm{Ang}}(k) + \omega_a \underline{\boldsymbol{h}}_j^{\mathrm{Ang}}(k)}\right\} \\
&= \omega_r \chi^2\left(\underline{\boldsymbol{h}}_i^{\mathrm{Rad}}, \underline{\boldsymbol{h}}_j^{\mathrm{Rad}}\right) + \omega_c \chi^2\left(\underline{\boldsymbol{h}}_i^{\mathrm{Circ}}, \underline{\boldsymbol{h}}_j^{\mathrm{Circ}}\right) + \omega_a \chi^2\left(\underline{\boldsymbol{h}}_i^{\mathrm{Ang}}, \underline{\boldsymbol{h}}_j^{\mathrm{Ang}}\right)
\end{aligned} \tag{6}$$

where $C$ is the number of texture classes and $K$ is the number of textons per class corresponding to some feature. Therefore, by combining kernels calculated on each individual feature we have

$$\begin{aligned}
&\mathbf{K}^*\left(\underline{\boldsymbol{h}}_i^{\mathrm{Comb}}, \underline{\boldsymbol{h}}_j^{\mathrm{Comb}}\right) \\
&= \left[\mathbf{K}_r\left(\underline{\boldsymbol{h}}_i^{\mathrm{Rad}}, \underline{\boldsymbol{h}}_j^{\mathrm{Rad}}\right)\right]^{\omega_r} \left[\mathbf{K}_c\left(\underline{\boldsymbol{h}}_i^{\mathrm{Circ}}, \underline{\boldsymbol{h}}_j^{\mathrm{Circ}}\right)\right]^{\omega_c} \left[\mathbf{K}_a\left(\underline{\boldsymbol{h}}_i^{\mathrm{Ang}}, \underline{\boldsymbol{h}}_j^{\mathrm{Ang}}\right)\right]^{\omega_a}.
\end{aligned}$$

Equation (6) indicates that the weighting gives a linear combination of the similarity between histograms computed from different features, offering the capacity to give higher weights to the more discriminative features during learning. Moreover, it is a Mercer kernel and also has the capability to ignore features which do not perform well.

We will first compare the performance of combinations of two or three different features with that of a single feature in the *histogram* framework, focusing on individual or combined performance of the SRP Circular, Radial-Diff, and Angular-Diff features. The kernel combination weights are learned by cross validation, to find appropriate weights for fusing the

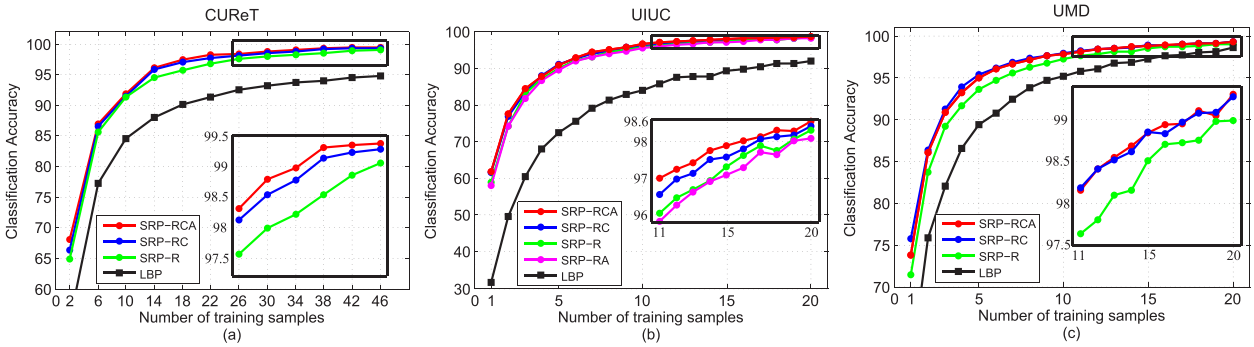


Fig. 3. Classification rate versus number of training samples, comparing single and combined SRP results on three datasets: (a) CUReT, (b) UIUC, and (c) and UMD. The corresponding classification accuracies achieved by the LBP [15] method are also included for comparison. Implementation parameters involved are the same as those in Table II. In general, the combined approach (RCA) offers superior performance. All our SRP approaches outperform the LBP method significantly and consistently.

SPR features. We begin with an initial weight distribution, and then separately and sequentially adjust each individual weight, and repeat the process until no further improvement can be attained. Such a scheme is to ensure that the resulting fused kernel achieves the best performances. However, based on our investigation, we find that the weights in the combining scheme defined in (6) and (7) can be kept fixed in this section, since the same weights can give very good results. However, varying the weights in (6) and (7) will be used later, in the more challenging material categorization task.

Fig. 3 and Table II show results for various datasets and as a function of the number of training examples, comparing the combined descriptors with the best single descriptor SRP Radial-Diff. The results achieved by the LBP method proposed in [15] are also included for comparison. What is clear from both the table and the figure is that, uniformly across all datasets and across all degrees of training data, the combined classifiers outperform the single one, implying that there is some degree of complementarity in the SRP features. The combination of all three descriptors, the SRP-RCA classifier, is our proposed choice. Moreover, all our SRP approaches outperforms the LBP method significantly and consistently.

## IV. Model Selection

Many samples may be redundant or irrelevant because some of the training samples might not be responsible for the classification of the testing images. It is essential and indispensable to select a subset of training images that is most relevant to classification. Model selection methods select a subset of examples from the original training data, removing noisy, redundant, and both kinds of samples. We wish to obtain a representative training set with a reduced size, when compared with the original, and with a similar or even higher classification accuracy for new incoming data. In our case, an image is represented as a BoW histogram model, which may be thought of as a point in the feature space, therefore, the models for a particular texture (material) class simply consist of a set of points in the feature space. In our tests, so far the number of models was the same as the number of training images, however, not all training models are responsible for good classification performance, therefore, it may be convenient to discard irrelevant training models. The key issue, then, is how to find a reliable number of models appropriate for each class. No prior information is available to help answering this question.

TABLE II

Comparing Single and Combined SRP Features: The Subtables Show a Comparison of Single and Combined SRP Results, With Combinations of SRP Radial-Diff (R), SRP Circular (C), and SRP Angular-Diff (A); the Patch Sizes Used Are $13 \times 13$, $13 \times 13$, $9 \times 9$, $9 \times 9$, and $11 \times 11$ for $\mathcal{D}^{C}$, $\mathcal{D}^{\mathrm{KT}}$, $\mathcal{D}^{\mathrm{UIUC}}$, $\mathcal{D}^{\mathrm{UMD}}$, and $\mathcal{D}^{\mathrm{CRot}}$, Respectively. The Number of Textons $K$ Used Per Class Are 10, 40, 100, 40, and 10, Respectively. The Weights Used Are Set as $\omega_r = \omega_c = \omega_a = 0.5$. The Results Given by the LBP [15] Are Also Included as a Baseline. (a) CUReT (92 Samples Per Class in Total). (b) KTHTIPS (81 Samples Per Class in Total). (c) (UIUC) (40 Samples Per Class in Total). (d) UMD (40 Samples Per Class in Total). (e) CUReTRot (92 Samples Per Class in Total)

(a)

| Methods | | | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| R | C | A | 2 | 10 | 18 | 26 | 34 | 38 | 46 |
| √ | √ | √ | **68.07** | **91.77** | **97.45** | **98.31** | **98.98** | **99.31** | **99.37** |
| √ | √ | | 66.33 | 91.44 | 96.97 | 98.13 | 98.78 | 99.13 | 99.28 |
| √ | | | 64.88 | 91.30 | 95.71 | 97.57 | 98.22 | 98.53 | 99.05 |
| LBP [15] | | | 54.87 | 84.55 | 90.11 | 92.56 | 93.65 | 94.01 | 94.76 |

(b)

| Methods | | | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| R | C | A | 5 | 10 | 20 | 25 | 30 | 35 | 40 |
| √ | √ | √ | **81.18** | 88.99 | 96.10 | **97.74** | 98.38 | 98.71 | 99.06 |
| √ | √ | | 80.90 | **89.49** | **96.40** | 97.32 | **98.40** | **99.07** | **99.29** |
| √ | | | 79.72 | 88.93 | 95.81 | 97.45 | 98.22 | 98.62 | 99.01 |
| LBP [15] | | | 74.24 | 83.00 | 90.71 | 92.36 | 93.59 | 94.67 | 95.04 |

(c)

| Methods | | | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| R | C | A | 1 | 5 | 10 | 13 | 15 | 18 | 20 |
| √ | √ | √ | **61.82** | 90.84 | **96.61** | **97.42** | **97.89** | **98.30** | **98.56** |
| √ | √ | | 61.62 | **90.96** | 96.00 | 97.14 | 97.59 | 98.13 | 98.42 |
| √ | | | 59.00 | 89.84 | 95.67 | 96.69 | 97.31 | 97.75 | 98.30 |
| LBP [15] | | | 31.63 | 72.34 | 84.02 | 87.79 | 89.20 | 91.15 | 91.94 |

(d)

| Methods | | | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| R | C | A | 1 | 5 | 10 | 13 | 15 | 18 | 20 |
| √ | √ | √ | 73.78 | 94.95 | 97.85 | **98.54** | **98.86** | **99.11** | **99.30** |
| √ | √ | | **75.78** | **95.33** | **97.91** | 98.51 | 98.85 | 99.08 | 99.27 |
| √ | | | 71.45 | 93.56 | 97.26 | 98.09 | 98.51 | 98.75 | 98.99 |
| LBP [15] | | | 60.26 | 89.35 | 95.19 | 96.70 | 97.23 | 98.00 | 98.63 |

(e)

| Methods | | | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| R | C | A | 2 | 10 | 18 | 26 | 34 | 38 | 46 |
| √ | √ | √ | **66.64** | **89.43** | **95.75** | **96.75** | **97.93** | **98.28** | **98.62** |
| √ | √ | | 65.05 | 88.87 | 95.36 | 96.61 | 97.67 | 98.13 | 98.49 |
| √ | | | 62.28 | 87.78 | 92.99 | 94.98 | 95.95 | 96.57 | 96.87 |
| LBP [15] | | | 49.89 | 76.72 | 82.53 | 85.44 | 86.77 | 87.24 | 88.20 |

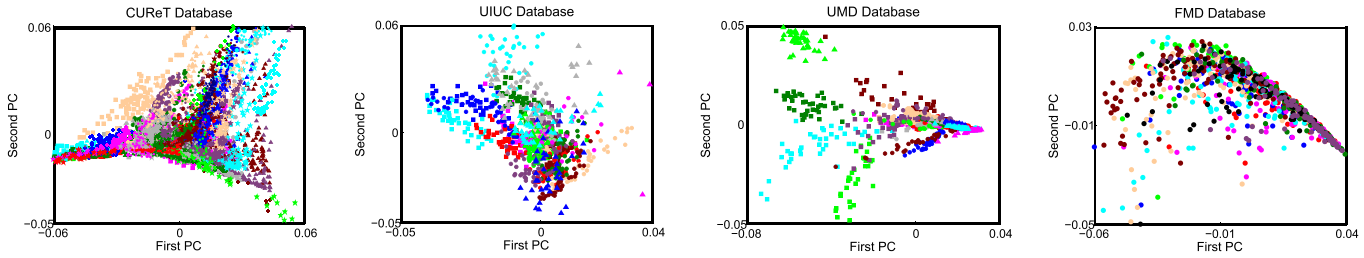


Fig. 4. Projection on the first two principal components (PCs) of the BoW histograms of the SRP Radial-Diff feature are shown for all images in (a) 61 classes in the CUReT database, (b) 25 classes in the UIUC database, (c) 25 classes in the UMD database, and (d) 10 classes in the FMD database. The colors and the marker types indicate the various texture/material categories. Clearly, the FMD material samples are the most difficult to separate.

TABLE III
SUMMARY OF TEXTURE DATASETS USED IN OUR EXPERIMENTS

| Texture Dataset | Dataset Notation | Image Rotation | Controlled Illumination | Scale Variation | Significant Viewpoint | Texture Classes | Sample Size | Samples per class | Samples in Total |
|---|---|---|---|---|---|---|---|---|---|
| CUReT | $\mathcal{D}^{\mathrm{C}}$ | √ | √ | | | 61 | 200 × 200 | 92 | 5612 |
| CUReTRot | $\mathcal{D}^{\mathrm{CRot}}$ | √ | √ | | | 61 | 140 × 140 | 92 | 5612 |
| Brodatz | $\mathcal{D}^{\mathrm{B}}$ | | | | | 111 | 215 × 215 | 9 | 999 |
| UIUC | $\mathcal{D}^{\mathrm{UIUC}}$ | √ | | √ | √ | 25 | 640 × 480 | 40 | 1000 |
| UMD | $\mathcal{D}^{\mathrm{UMD}}$ | √ | | √ | √ | 25 | 320 × 240 | 40 | 1000 |
| KTHTIPS | $\mathcal{D}^{\mathrm{KT}}$ | | √ | √ | | 10 | 200 × 200 | 81 | 810 |
| KTHTIPS2b | $\mathcal{D}^{\mathrm{KT2b}}$ | | √ | √ | | 11 | 200 × 200 | 432 | 4752 |
| FMD | $\mathcal{D}^{\mathrm{FMD}}$ | √ | | √ | √ | 10 | 512 × 384 | 100 | 1000 |

Many machine learning techniques have been developed [50] to reduce the number of models in a classification algorithm. Our model selection problem is closely related to the well-known problem [50] of prototype or instance selection. Much work has been done on prototype selection to reduce the training set, to reduce the effect of noise on accuracy, and to obtain the same or even better classification ability compared with using the whole training set. It is beyond the scope of this paper to comprehensively test model selection approaches; our goal is to investigate whether model selection can improve classification performance.

A formal specification of the model selection problem follows: let $\mathcal{H}_{\mathrm{Tr}}$ and $\mathcal{H}_{\mathrm{Te}}$ denote the training models set and the testing models set, respectively, with each set having $N_{\mathrm{Tr}}$ and $N_{\mathrm{Te}}$ models. Let $\mathcal{H}_S \subset \mathcal{H}_{\mathrm{Tr}}$ be the subset of selected models resulting from the execution of a model selection algorithm; then, we classify a new model $\underline{\boldsymbol{h}}_{\mathrm{new}}$ from $\mathcal{H}_{\mathrm{Te}}$ acting over $\mathcal{H}_S$ instead of $\mathcal{H}_{\mathrm{Tr}}$. Inspecting the plots in Fig. 4, it is apparent that the FMD is more challenging than the CUReT, University of Illinois at Urbana-Champaign (UIUC), and UMD texture databases, so for this reason we chose the FMD material database as the basis for developing and evaluating model selection.

Informative models can be derived from training samples in various ways. A method for finding prototypes can be categorized as cluster-based learning algorithms, in which prototypes are not samples per se, but can be derived as the weighted averages of samples. Whereas we propose to develop a greedy algorithm, designed to maximize the classification accuracy, while minimizing the number of models used, with decremental and incremental searches for a subset $\mathcal{H}_S$ of BoW models to keep from training set $\mathcal{H}_{\mathrm{Tr}}$.

A *decremental* search [50] begins with the training model set $\mathcal{H}_S = \mathcal{H}_{\mathrm{Tr}}$, and then searches iteratively for instances to remove from $\mathcal{H}_S$, on the basis of the one for which the classification accuracy decreases the least when it is dropped. We investigate model reduction in a SVM framework. Models are selected from the training set $\mathcal{H}_{\mathrm{Tr}}$ and classification results reported only on the testing set $\mathcal{H}_{\mathrm{Te}}$. This iteration is repeated until no more models are left. Note that the proposed model selection algorithm is constrained to select models only from the training set and classification performance is being reported on the test set. This emulates the setup of Cula and Dana [5] and Varma and Zisserman [12], where the model reduction algorithm has access to both training and test images for each texture class. However, it must be emphasized that in real world classification, the test set is not available for inspection to the training set. Nevertheless, as we will show later in the experiment section, there is a significant level of difference between the performance of the model selection algorithm on one hand and the proposed fused SRP approach on the other. An *incremental* search [50] can also be proposed, beginning with an empty subset $\mathcal{H}_S$, and iteratively adding instances in $\mathcal{H}_{\mathrm{Tr}}$ to $\mathcal{H}_S$ based on the criterion that the model is chosen to be the one for which the classification accuracy increases the most when it is added.

Although the decremental search implies a higher computational cost than incremental algorithms, the main disadvantage with the incremental approach is that the feature decisions are based on little information, and indeed our experimental tests in the context of texture classification showed the decremental approach to significantly outperform incremental, and so we will adopt the decremental approach in this paper.

## V. EXPERIMENTAL EVALUATION

In this section, we test the proposed approach against *eight* public benchmark databases, as summarized in Table III,


Fig. 5. Image samples from the KTHTIPS2b database: each row shows one example image from each of four samples of a category.

for texture classification and material categorization, comparing our results to the current state of the art.

### A. Datasets and Experimental Setup

For texture classification, we have performed extensive evaluation on *seven* popular texture databases. For material categorization, we have used the challenging FMD materials database from Massachusetts Institute of Technology. All the eight datasets are summarized in Table III and described in the following.

For CUReT, we use the same subset of images as [12], [13], [16], and [17]. The texture appearances vary significantly from one to the next due to being captured under different illuminations and viewing directions. The CUReTRot dataset is newly generated from CUReT by rotating each sample according to a randomly generated angle, helping to validate rotation invariance. For Brodatz [51], we use the same dataset as [11], [17], and [23]. The UIUC database [11] has been designed to require local invariance. Textures are acquired under significant scale and viewpoint changes, arbitrary rotations, and uncontrolled illumination conditions. The UMD database [22] has each image downsampled to 240 $\times$ 320 using bilinear interpolation. The KTH-TIPS database [8] extends CUReT by imaging new samples of ten of the CUReT textures at a subset of the viewing and lighting angles used in CUReT, but over a range of scales. Although KTH-TIPS is designed to be combined with CUReT in testing, we follow Zhang *et al.* [23] in treating it as a stand-alone dataset.

The KTHTIPS2b [8], [9] is a more challenging database dealing with classifying images from unseen physical samples of materials. It contains four samples of 11 different materials shown in Fig. 5, each sample imaged at nine different scales equally spaced logarithmically over two octaves, and 12 lighting and pose setups, totaling 4572 images. However, there is almost no rotation changes in this database. For the experiments on KTHTIPS2b, we follow the training and testing scheme used in [8] and [9]. We perform experiments training on one, two, or three samples; testing is always conducted only on unseen samples.

The FMD created in [6] selects photos from Flickr as samples for common material categories. This database contains 10 common material categories, as shown in Fig. 6. The images capture a wide range of appearances within each material category. Liu *et al.* [6] reported state-of-the-art results by exploring a large set of heterogeneous features and proposing an aLDA model to combine these features under a Bayesian generative framework.


Fig. 6. Ten material categories in the FMD [6]. The FMD database captures a range of appearances within each material category.

TABLE IV
RP PROJECTING DIMENSIONALITY USED FOR THE CORRESPONDING PATCH SIZES IN OUR EXPERIMENTS. THEORETICAL DETAILS FOR DECIDING THE DIMENSIONALITY CAN BE FOUND IN [16]

| Patch Size | $5 \times 5$ | $7 \times 7$ | $9 \times 9$ | $11 \times 11$ | $13 \times 13$ | $15 \times 15$ | $17 \times 17$ | $19 \times 19$ |
|---|---|---|---|---|---|---|---|---|
| RP Dim | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 |

We will be performing a comparative evaluation of our proposed combined SRP-SVM approach against 13 state-of-the-art texture classification methods, as shown in Table VI, with a detailed comparison with the following four state-of-the-art methods, which have been shown to perform better than other methods, based on the comparative study of Zhang *et al.* [23].

1) The method of Caputo *et al.* [8], an extension of the MR8 approach by Varma and Zisserman [12]. They use the MR8 descriptor and SVM with a $\chi^2$ kernel for classification.

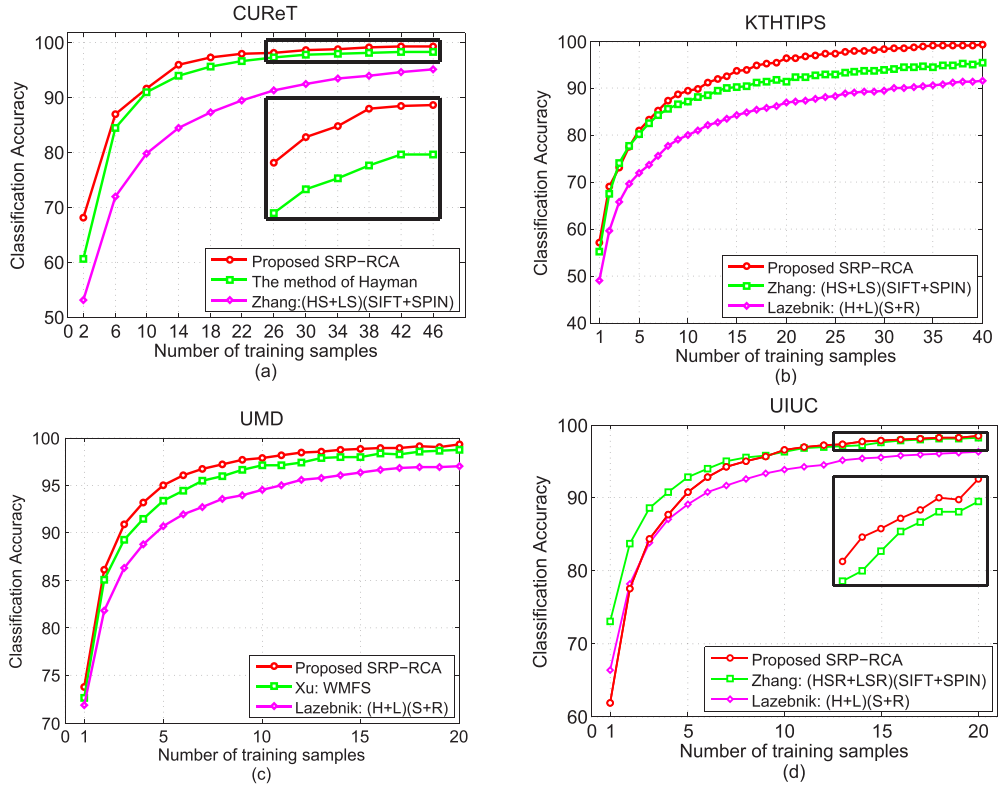


Fig. 7. Classification rate versus the number of training samples, comparing classification scores of the proposed combined SRP approach with those of the state of the art methods as a function of the number of training samples on four texture datasets: (a) CUReT, (b) KTH-TIPS, (c) UMD, and (d) UIUC.

TABLE V

COMPARING THE PROPOSED APPROACH WITH STATE OF THE ART: THE IMPLEMENTATION PARAMETERS ARE SAME AS THOSE IN TABLE II. (a) CUReT (92 SAMPLES PER CLASS IN TOTAL). (b) KTHTIPS (81 SAMPLES PER CLASS IN TOTAL). (c) UIUC (40 SAMPLES PER CLASS IN TOTAL). (d) UMD (40 SAMPLES PER CLASS IN TOTAL)

(a)

| Methods | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|
| | 2 | 10 | 18 | 26 | 34 | 38 | 46 |
| R+C+A | **68.07** | **91.77** | **97.45** | **98.31** | **98.98** | **99.31** | **99.37** |
| Hayman [8] | 60.64 | 91.09 | 97.39 | 98.26 | 97.59 | 98.13 | 98.46 |
| Zhang [24] | 53.04 | 79.78 | 87.39 | 91.30 | 93.48 | 94.13 | 95.22 |

(b)

| Methods | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|
| | 5 | 10 | 20 | 25 | 30 | 35 | 40 |
| R+C+A | **81.18** | **88.99** | **96.10** | **97.74** | **98.38** | **98.71** | **99.06** |
| Zhang [24] | 80.23 | 87.21 | 91.40 | 93.02 | 93.95 | 94.42 | 95.45 |
| Lazebnik [11] | 71.86 | 80.08 | 86.98 | 88.37 | 89.54 | 90.70 | 91.63 |

(c)

| Methods | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 5 | 10 | 13 | 15 | 18 | 20 |
| R+C+A | 61.82 | 90.84 | **96.61** | **97.42** | **97.89** | **98.30** | **98.56** |
| Zhang [24] | **73.12** | **92.84** | 96.35 | 97.16 | 97.57 | 98.11 | 98.25 |
| Lazebnik [11] | 66.35 | 89.19 | 93.92 | 95.14 | 95.54 | 96.08 | 96.35 |

(d)

| Methods | Classification Accuracy % Number of training samples per class | | | | | | |
|---|---|---|---|---|---|---|---|
| | 1 | 5 | 10 | 13 | 15 | 18 | 20 |
| R+C+A | **73.78** | **94.95** | **97.85** | **98.54** | **98.84** | **99.11** | **99.30** |
| Xu [23] | 72.58 | 93.40 | 97.05 | 97.87 | 97.92 | 98.51 | 98.68 |
| Lazebnik [11] | 71.90 | 90.71 | 94.54 | 95.79 | 96.29 | 96.87 | 96.95 |

2) The method of Lazebnik *et al.* [11], first characterizing a texture using Harris-affine corners and Laplacian-affine blobs, with two descriptors (SPIN and RIFT) used for feature extraction. The NNC classifier with EMD is used.

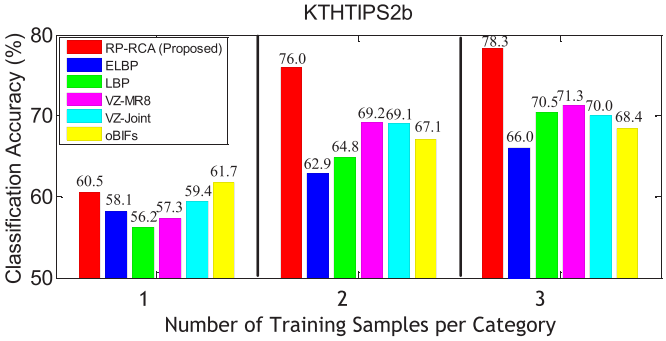


Fig. 8. Comparison with state-of-the-art methods on the KTHTIPS2b database. Number of textons per class used is 100, and the patch size $11 \times 11$ is used. Results are original, except those of VZ-MR8, VZ-Joint, and LBP, which are quoted from [8]. All results are obtained with SVM classification, except those of ELBP [54] and oBIFs [56], which are obtained with NNC.

3) The method of Zhang *et al.* [23], based on the method of Lazebnik *et al.* [11], combining three types of local descriptors (SPIN, RIFT, and SIFT) and a kernel SVM classifier with the EMD.
4) *wavelet multifractal spectrum* [22], based on a combination of wavelet transform and multifractal analysis. It first performs scale normalization using the scale estimated from local affine invariant Laplacian regions, then computes the multiorientation wavelet pyramid and wavelet leaders, finally computes the MFS vectors as the global texture feature. An SVM classifier with RBF kernel is used.

TABLE VI

COMPARING THE CLASSIFICATION SCORES ACHIEVED BY THE PROPOSED SRP-RCA APPROACH WITH THOSE ACHIEVED BY 13 STATE-OF-THE-ART METHODS ON FIVE DATASETS. SCORES ARE AS ORIGINALLY REPORTED, EXCEPT AS MARKED (∗), WHICH ARE TAKEN FROM IN [24]. THE BRACKETED NUMBERS REPORT THE NUMBER OF TRAINING SAMPLES PER CLASS

| Methods | $\mathcal{D}^{C}$ (46) 46 training samples per class | $\mathcal{D}^{B}$ (3) 3 training samples per class | $\mathcal{D}^{KT}$ (41) 41 training samples per class | $\mathcal{D}^{UIUC}$ (20) 20 training samples per class | $\mathcal{D}^{UMD}$ (20) 20 training samples per class |
|---|---|---|---|---|---|
| Proposed SRP-RCA | **99.37**% ± **0.24**% | **96.78**% ± **0.36**% | **99.06**% ± **0.44**% | **98.56**% ± %**0.39** | **99.30**% ± **0.29**% |
| 2. VZ-MR8 [12] | 97.43% | | | | |
| 3. VZ-Patch [13] | 98.03% | 92.9% ± 0.8%(∗) | 92.4% ± 2.1%(∗) | 97.83% ± 0.66% | |
| 4. Hayman *et al.* [8] | 98.46% ± 0.09% | 95.0% ± 0.8%(∗) | 94.8% ± 1.2%(∗) | 92.0% ± 0.1.3%(∗) | |
| 5. Lazebnik *et al.* [11] | 72.5% ± 0.7%(∗) | 88.15% | 91.3% ± 1.4%(∗) | 96.03% | |
| 6. Mellor *et al.* [55] | | 89.71% | | | |
| 7. Zhang *et al.* [24] | 95.3% ± 0.4% | 95.4% ± 0.7% | 95.5% ± 1.3% | **98.7**% ± **0.4**% | |
| 8. Varma and Ray [44] | | | | **98.76**% ± **0.64**% | |
| 9. Crosier and Griffin [14] | 98.6% ± 0.2% | | 98.5% ± 0.7% | **98.8**% ± **0.5**% | |
| 10. Xu-MFS *et al.* [29] | | | | 92.74% | 93.93% |
| 11. Xu-OTF *et al.* [22] | | | | 97.40% | 98.49% |
| 12. Xu-WMFS *et al.* [23] | | | | **98.60**% | 98.68% |
| 13. Liu *et al.* [18] | 98.52% ± 0.19% | **96.34**% ± **0.38**% | 97.71% ± 0.49% | 96.27% ± 0.44% | **99.13**% ± **0.16**% |

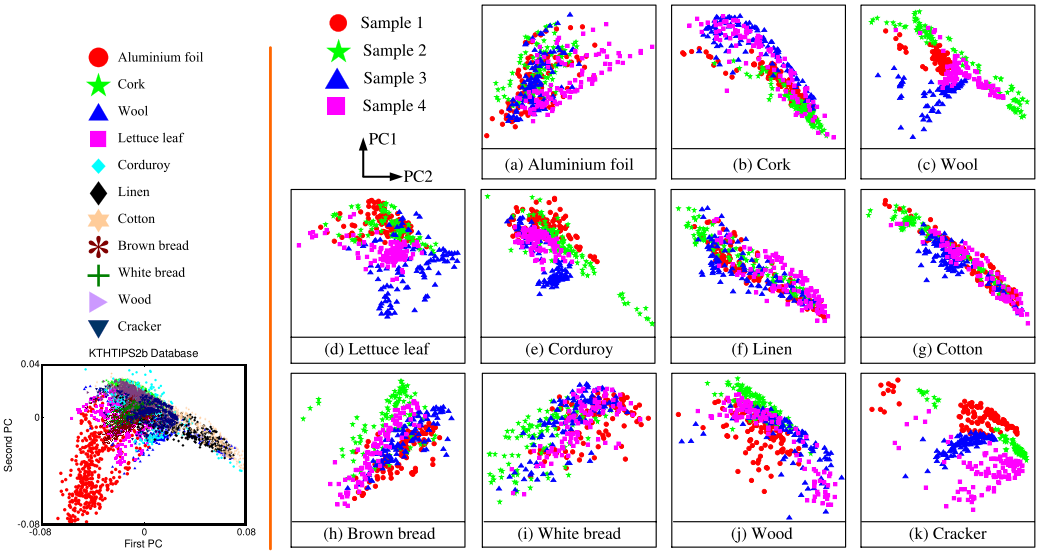


Fig. 9. In the left panel, the projection of the first two PCs of the BoW histograms of the SRP Radial-Diff feature are shown for all images in the 11 categories of KTHTIPS2b. The colors indicate the various texture categories. Correspondingly, in the right panel, figures (a)–(k) provide more detail plots for each category. The colors indicate the various samples in each category. (a) Aluminum foil. (b) Cork. (c) Wool. (d) Lettuce leaf. (e) Corduroy. (f) Linen. (g) Cotton. (h) Brown bread. (i) White bread. (j) Wood. (k) Cracker.

*1) Implementation Details:* To make the comparisons as meaningful as possible, for texture classification, we use the same experimental settings as in [13], [16], and [17] and the reader is referred to those papers for additional details. For material categorization, the training and testing strategy is kept the same as in [6]. The RP dimensions used in our experiments are summarized in Table IV, and the rationale behind the RP dimension selection can be found in [16].

Each sample is normalized to be zero mean, unit standard deviation, and the extracted SRP vector is normalized via Weber's law. All results are reported over 50 random partitions of the training and testing sets. Half of the samples per class are randomly selected for training and the remaining half for testing, except for $\mathcal{D}^{B}$, where three samples are randomly selected as training and the remaining six as testing.

We use the publicly available *LibSVM* library [52]. The values of the parameters $C$ and $\gamma$ of SVM are specified using a grid search scheme. The parameters $C$ and $\gamma$ are searched exponentially in the ranges of $[2^{-5}, 2^{18}]$ and $[2^{-15}, 2^{8}]$, respectively, with a step size of $2^{1}$ to probe the highest classification rate.

## B. Tests on Texture Databases

All results are taken directly from the original publications, except that the results of Hayman on $\mathcal{D}^{C}$ and of Lazebnik on $\mathcal{D}^{KT}$ are quoted from the recent comparative study of Zhang *et al.* [23], and the results of Lazebnik with SVM on $\mathcal{D}^{UMD}$ from the work of Xu *et al.* [22].

Fig. 7 and Table V compare the classification performance of our approach with that of state-of-the-art methods. Our method outperforms competing approaches on $\mathcal{D}^{C}$, $\mathcal{D}^{KT}$, and $\mathcal{D}^{UMD}$. On $\mathcal{D}^{UIUC}$ our method outperforms when sufficient training data are available.

Table VI presents a broader set of comparisons, comparing our proposed approach with 13 others. Our approach scores produces what we believe to be the best reported results on the CUReT, Brodatz, KTH-TIPS, and UMD datasets, and

TABLE VII
COMPARING SINGLE AND COMBINED SRP FEATURES ON THE FMD DATABASE, FOR A $9 \times 9$ PATCH SIZE, NUMBER OF TEXTONS PER CLASS $K = 100$

| Methods | | | Number of training samples per class | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| R | A | C | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 |
| √ | √ | √ | **36.26** | **42.60** | 46.54 | 49.73 | 52.41 | 53.22 | 54.18 | 56.60 | 57.40 |
| √ | √ | | 35.81 | 42.23 | **46.75** | **50.16** | **52.67** | **53.93** | **55.50** | **56.92** | **59.03** |
| √ | | | 35.66 | 42.36 | 46.54 | 49.42 | 51.80 | 53.15 | 54.46 | 56.07 | 57.20 |
| | √ | | 31.59 | 37.61 | 41.50 | 44.41 | 46.34 | 48.62 | 50.43 | 52.05 | 53.78 |

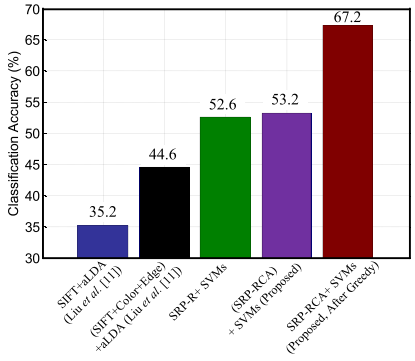


Fig. 10. Comparison of the best classification performance on the FMD. All scores are as originally reported.

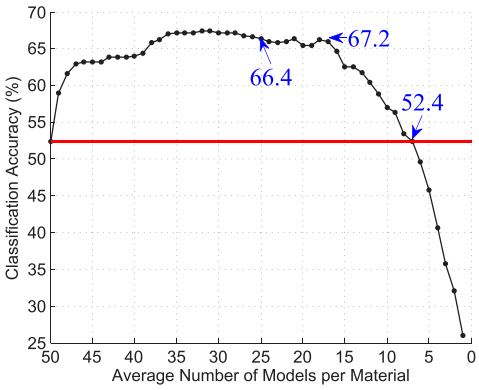


Fig. 11. Classification rate as a function of the number of training images selected by the greedy algorithm (decremental search) for the FMD. We can observe that the classification accuracy by the greedy algorithm improves significantly, while using, on average, fewer models. The patch size tested is $9 \times 9$; the number of textons learned per class is 100.

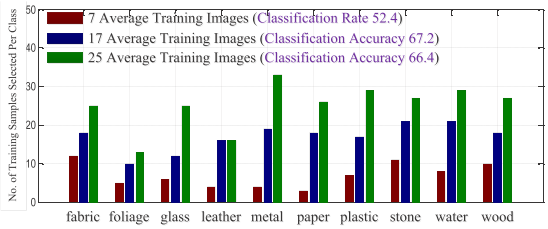


Fig. 12. Number of models selected for each category by the greedy algorithm, while classifying all 10 materials.

competitive performance on UIUC. We would like to point out that all of the state-of-the-art methods were tested on only a subset of the datasets, to some extent tuned to one or two. Our single approach matches or outperforms each competing method, even on its preferred dataset.

In terms of KTHTIPS2b, noting that there is only very small rotation variation, we choose to fuse three *unsorted* random features: RP, RP Radial-Diff, and RP Angular-Diff. Fig. 8 shows the performance comparison of our approach and several state-of-the-art methods.[3] It can be observed from Fig. 8 that our fused feature demonstrates superior performance, significantly improving the state-of-the-art performance for two training samples and three training samples. However, unlike the results on other tested texture datasets, the classification performance on the KTHTIPS2b database is still far from satisfying. The plots for the KTHTIPS2b similar to those in Fig. 4 are shown in Fig. 9. It can be observed from Fig. 9 (the subfigure in the left panel) that the 11 material categories form good clusters in the feature space. Since the samples (with all their 108 images) are distributed over either the training or the testing set, the classification would be easy if the BoW features of all the four samples overlap (meaning less intra-class variations) in the feature space. However, inspecting the subfigures in the right panel of Fig. 9, we can observe that some classes possess larger intra-class variations than others. This is consistent with the observation from Fig. 5 that the variation in appearance between the samples in each category is larger for some categories than others. It can be observed from Figs. 5 and 9 that cork, for instance, contains relatively little intra-category variation, whereas cracker and wool exhibit significant intra-category variation. These sources of intra-class variation provide a stern challenge indeed for recognition.

[3]Varma and Zisserman (VZ)-MR8 [12], VZ-Joint [13], LBP [15], extended local binary pattern [53], and basic image features [55].

### C. Tests on Material Database

Given the huge range of image samples in the FMD, as shown in Fig. 6, arguably any single feature, based on texture or color, may not be sufficient [7], [29]. Instead, broad material categorization may need multiple low-level cues in conjunction with high-level object knowledge, the sort of cues that the proposed BoWs method applied to combined random features might provide.

To automatically choose the value of the weighting vector $\omega$ in (6) that gives the best result in term of classification performance, we learn the weights $\omega_r$, $\omega_c$, and $\omega_a$ by cross validation. We have opted to use a varying $\omega$ and to replace one of its entries by an ascending series while keep other two entries fixed. Starting from a low value, it is possible to check every few iterations whether increasing this value provides a better classification rate. In this scheme, the parameters and are searched in the ranges of [0, 1], with a step size of 0.1 to probe the highest classification rate. In other words, the values of the parameters $\omega_a$, $\omega_c$, and $\omega_a$ can be learned using a grid search scheme.

Table VII presents the results on the FMD, comparing the single and combined SRP descriptors and their combination. In all cases, a combined approach outperforms individual features. More significantly, Fig. 10 presents the best classification rates achieved by the proposed approach competing

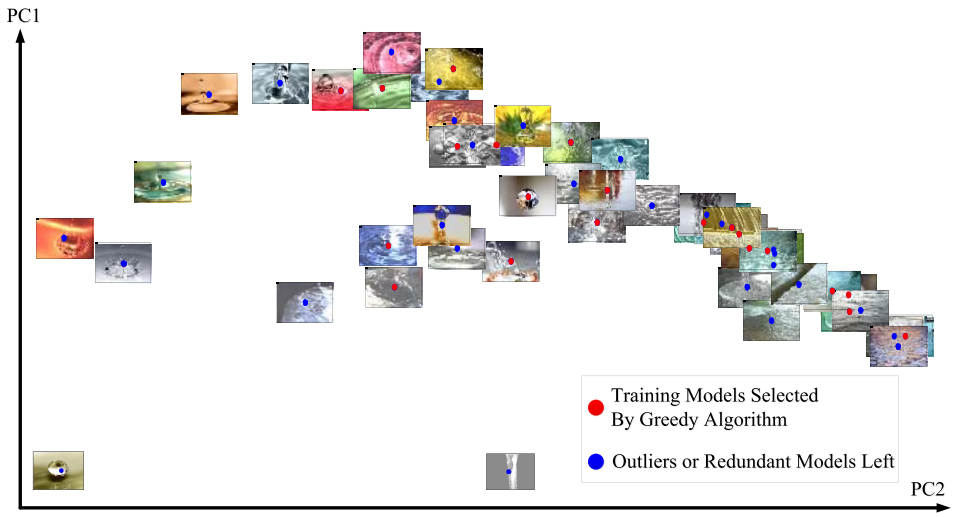


Fig. 13. PCA embedding of BoW histograms for the 50 training material images from the water class in two dimensions. The red dots represent the models selected by greedy algorithm. The results illustrate that using a fewer number of models can improve performance and that the greedy algorithm is good at removing outliers and redundant examples.

state-of-the-art methods. Both without combining (52.6% from SRP-R) and with combining (53.2% from SPR-RCA) significantly outperform the 35.2% performance of the best single feature (SIFT) reported in [6] with their aLDA approach, and the best reported score of 44.6% from [6] in combining color, texture, shape, and edge features with aLDA. According to our experimental results we find foliage the easiest to classify and metal the hardest, consistent with the findings from [6].

Although our proposed approach produces the highest accuracy of 53.2% on FMD, this accuracy it is still much lower than the rates which we saw for texture classification in Table VI. As a final step, we test the proposed decremental greedy search algorithm to boost classification performance by focusing on the most relevant subset of features. Fig. 11 shows the results of classifying the FMD using the proposed SRP-RCA approach. It is very interesting to note that the classification accuracy obtained using seven training samples per class can be the same as those obtained using all 50 training samples per class. The classification accuracies achieved by more than seven training samples are all higher than the classification score without the greedy approach; for example, at 17 training samples per class we obtain a classification accuracy of 67.2%, far higher than any other reported result. The increase in performance indicates that using fewer models can improve performance, meaning that the greedy algorithm is good at rejecting noisy and redundant examples.

To obtain some insight into the proposed approach, Fig. 12 shows the number of training models selected per material category for different average number of training samples selected per class. Fig. 13 visualizes the selected training samples for the water category (when on an average 17 training images are selected and used for classification), showing the rejection of outliers and redundant examples.

## VI. Conclusion

In this paper, a very simple, robust, yet highly effective approach has been proposed for texture classification and material categorization. The proposed approach employs the following:

1) the simplicity and universality of SRP random features in using RP for information-preserving dimensionality reduction;
2) the advantage of BoW in computational efficiency and global invariance to environment changes;
3) the advantages of a simple strategy for combining multiple SVM kernels compared to complex MKL.

We have tested our approach on six popular and challenging texture databases and one difficult material database. Based on extensive experimental results on a wide variety of texture datasets, our system yields the best classification rates of which we are aware for CUReT, Brodatz, UMD, KTH-TIPS, and FMD.

The proposed system is illumination and rotation invariant, however, one limitation of the proposed system is that it is not invariant to significant scale changes. Although work has been reported on scale invariance [11], [22], [23], no approach has been reported to handle scale changes of a large magnitude. Given the flexibility of our proposed method, we see research into scale-invariance as a promising direction for future research.

## References

[1] M. Tuceryan and A. K. Jain, "Texture analysis," in *Handbook of Pattern Recognition and Computer Vision*, C. H. Chen, L. F. Pau, and P. S. P. Wang, Eds. Singapore: World Scientific, 1993, pp. 235–276.

[2] T. Randen and J. H. Husøy, "Filtering for texture classification: A comparative study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 21, no. 4, pp. 291–310, Apr. 1999.

[3] T. N. Pappas, J. Zujovic, and D. L. Neuhoff, "Image analysis and compression: Renewed focus on texture," *Proc. SPIE, Vis. Inf. Process. Commun.*, vol. 7543, pp. 75430N-1–75430N-12, Jan. 2010.

[4] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, Jan. 1999.

[5] O. G. Cula and K. J. Dana, "3D texture recognition using bidirectional feature histograms," *Int. J. Comput. Vis.*, vol. 59, no. 1, pp. 33–36, 2004.

[6] C. Liu, L. Sharan, E. H. Adelson, and R. Rosenholtz, "Exploring features in a Bayesian framework for material recognition," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 239–246.

[7] L. Sharan, "The perception of material qualities in real-world images," Ph.D. dissertation, Dept. Elect. Eng. Comput. Sci., Massachusetts Institute Technology, Cambridge, MA, USA, 2009.

[8] B. Caputo, E. Hayman, M. Fritz, and J.-O. Eklundh, "Classifying materials in the real world," *Image Vis. Comput.*, vol. 28, no. 1, pp. 150–163, 2010.

[9] B. Caputo, E. Hayman, and P. Mallikarjuna, "Class-specific material categorisation," in *Proc. 10th IEEE Int. Conf. Comput. Vis.*, Oct. 2005, pp. 1597–1604.

[10] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vis.*, vol. 43, no. 1, pp. 29–44, Jun. 2001.

[11] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1265–1278, Aug. 2005.

[12] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *Int. J. Comput. Vis.*, vol. 62, nos. 1–2, pp. 61–81, 2005.

[13] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032–2047, Nov. 2009.

[14] M. Crosier and L. D. Griffin, "Using basic image features for texture classification," *Int. J. Comput. Vis.*, vol. 88, no. 3, pp. 447–460, 2010.

[15] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[16] L. Liu and P. W. Fieguth, "Texture classification from random features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 574–586, Mar. 2012.

[17] L. Liu, P. Fieguth, D. Clausi, and G. Kuang, "Sorted random projections for robust rotation-invariant texture classification," *Pattern Recognit.*, vol. 45, no. 6, pp. 2405–2418, Jun. 2012.

[18] S. ul Hussain, T. Napoleon, and F. Jurie, "Face recognition using local quantized patterns," in *Proc. Brit. Mach. Vis. Conf.*, 2012, pp. 1–11.

[19] G. Sharma, S. ul Hussain, and F. Jurie, "Local higher-order statistics (LHS) for texture categorization and facial analysis," in *Proc. 12th Eur. Conf. Comput. Vis.*, 2012, pp. 1–12.

[20] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Trans. Image Process.*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010.

[21] Y. Xu, S.-B. Huang, H. Ji, and C. Fermüller, "Combining powerful local and global statistics for texture description," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 573–580.

[22] Y. Xu, X. Yang, H. Ling, and H. Ji, "A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 161–168.

[23] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *Int. J. Comput. Vis.*, vol. 73, no. 2, pp. 213–238, 2007.

[24] L. Liu, P. Fieguth, G. Kuang, and H. Zha, "Sorted random projections for robust texture classification," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 391–398.

[25] L. Liu, P. Fieguth, and G. Kuang, "Combining sorted random features for texture classification," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 833–836.

[26] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1615–1630, Oct. 2005.

[27] S. Liao, M. W. K. Law, and A. C. S. Chung, "Dominant local binary patterns for texture classification," *IEEE Trans. Image Process.*, vol. 18, no. 5, pp. 1107–1118, May 2009.

[28] Y. Xu, H. Ji, and C. Fermüller, "Viewpoint invariant texture description using fractal analysis," *Int. J. Comput. Vis.*, vol. 83, no. 1, pp. 85–100, 2009.

[29] E. H. Adelson, "On seeing stuff: The perception of materials by humans and machines," *Proc. SPIE*, vol. 4299, pp. 1–12, Jun. 2001.

[30] S. C. Pont and J. J. Koenderink, "Bidirectional texture contrast function," *Int. J. Comput. Vis.*, vol. 62, nos. 1–2, pp. 17–34, 2005.

[31] S. G. Narasimhan, V. Ramesh, and S. K. Nayar, "A class of photometric invariants: Separating material from shape and illumination," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, Oct. 2003, pp. 1387–1394.

[32] S. K. Nayar and R. M. Bolle, "Reflectance based object recognition," *Int. J. Comput. Vis.*, vol. 17, no. 3, pp. 219–240, 1996.

[33] W. B. Johnson and J. Lindenstrauss, "Extensions of Lipschitz mappings into a Hilbert space," in *Proc. Conf. Modern Anal. Probab.*, 1984, pp. 189–206.

[34] S. Dasgupta, "Learning probability distributions," Ph.D. dissertation, Dept. Comput. Sci., Univ. California at Berkeley, Berkeley, CA, USA, 2000.

[35] D. Achlioptas, "Database-friendly random projections," in *Proc. 20th ACM Symp. Principles Database Syst.*, 2001, pp. 274–281.

[36] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–217, Feb. 2009.

[37] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[38] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[39] R. G. Baraniuk, "Compressive sensing [lecture notes]," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[40] R. G. Baraniuk, M. Davenport, R. A. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approx.*, vol. 28, no. 3, pp. 253–263, 2008.

[41] B. Schölkopf and A. J. Smola, *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, MA, USA: MIT Press, 2002.

[42] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 11, pp. 1542–1550, Nov. 2002.

[43] M. Varma and R. Garg, "Locally invariant fractal features for statistical texture classification," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[44] Y. Rubner, C. Tomasi, and L. Guibas, "The earth mover's distance as a metric for image retrieval," *Int. J. Comput. Vis.*, vol. 40, no. 2, pp. 99–121, 2000.

[45] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1055–1064, Sep. 1999.

[46] C.-W. Hsu and C.-J. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Netw.*, vol. 13, no. 2, pp. 415–425, Mar. 2002.

[47] M. Varma and D. Ray, "Learning the discriminative power-invariance trade-off," in *Proc. 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8.

[48] P. Gehler and S. Nowozin, "On feature combination for multiclass object classification," in *Proc. 12th IEEE Int. Conf. Comput. Vis.*, Sep./Oct. 2009, pp. 221–228.

[49] M.-E. Nilsback and A. Zisserman, "A visual vocabulary for flower classification," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Oct. 2006, pp. 1447–1454.

[50] S. García, J. Derrac, J. R. Cano, and F. Herrera, "Prototype selection for nearest neighbor classification: Taxonomy and empirical study," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 3, pp. 417–435, Mar. 2012.

[51] P. Brodatz, *Textures: A Photographic Album for Artists and Designers*. New York, NY, USA: Dover, 1966.

[52] C.-C. Chang and C.-J. Lin. (2001). *LIBSVM: A Library for Support Vector Machines*. [Online]. Available: http://www.csie.ntu.edu.tw/∼cjlin/libsvm

[53] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," *Image Vis. Comput.*, vol. 30, no. 2, pp. 86–99, 2012.

[54] M. Mellor, B.-W. Hong, and M. Brady, "Locally rotation, contrast, and scale invariant descriptors for texture analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 52–61, Jan. 2008.

[55] R. Timofte and L. Van Gool, "A training free classification framework for textures, writers, and materials," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2012, pp. 1–12.

**Li Liu** received the B.E. degree in communication engineering, the M.S. degree in photogrammetry and remote sensing, and the Ph.D. degree in information and communication engineering from National University of Defense Technology, Changsha, China, in 2003, 2005, and 2012, respectively.

She was a Visiting Student with University of Waterloo, Waterloo, ON, Canada, from 2008 to 2010. She has held a visiting appointment with Beijing University, Beijing, China. She joined the faculty of National University of Defense Technology in 2012, where she is currently a Lecturer with the School of Information System and Management.

Dr. Liu is the Co-Chair of the International Workshop on Robust local descriptors for computer vision at the 2014 Asian Conference on Computer Vision. Her research interests include computer vision, texture analysis, object recognition, pattern recognition, and video analysis and retrieval.

**Paul W. Fieguth** (S'87–M'96) received the B.A.Sc. degree in electrical engineering from University of Waterloo, Waterloo, ON, Canada, in 1991 and the Ph.D. degree in electrical engineering from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1995.

He joined the faculty of University of Waterloo in 1996, where he is currently a Professor of Systems Design Engineering. He has held visiting appointments with University of Heidelberg, Heidelberg, Germany; INRIA/Sophia, Valbonne, France; Cambridge Research Laboratory, Boston, MA, USA; Oxford University, Oxford, U.K.; and Rutherford Appleton Laboratory, Oxfordshire, U.K., and post-doctoral positions in computer science, and information and decision systems with University of Toronto, Toronto, ON, Canada, and MIT. His research interests include statistical signal and image processing, hierarchical algorithms, data fusion, and the interdisciplinary applications of such methods, in particular, remote sensing.

**Dewen Hu** received the B.Sc. and M.Sc. degrees from Xi'an Jiaotong University, Xi'an, China, in 1983 and 1986, respectively, and the Ph.D. degree from National University of Defense Technology, Changsha, China, in 1999.

He was a Visiting Scholar with University of Sheffield, Sheffield, U.K., from 1995 to 1996, where he became a Professor in 1996. He has been with National University of Defense Technology since 1986. His research interests include cognitive science, brain–computer interface, image processing, and neural networks.

Dr. Hu is an Action Editor of *Neural Networks*.

**Yingmei Wei** received the B.S., M.S., and Ph.D. degrees from National University of Defense Technology, Changsha, China, in 1993, 1996, and 2000, respectively, all in computer science and technology.

She joined the faculty of National University of Defense Technology in 2001, where she is currently a Professor with the School of Information System and Management. She was a Visiting Student with Zhejiang University, Hangzhou, China, from 1998 to 2000. She was a Visiting Scholar with University of Sheffield, Sheffield, U.K., from 2008 to 2009. Her research interests include multimedia, virtual reality, and computer graphics.

**Gangyao Kuang** received the B.E. and M.S. degrees from Central South University of Technology, Changsha, China, in 1988 and 1991, respectively, and the Ph.D. degree from National University of Defense Technology, Changsha, in 1995.

He was a Visiting Scholar with University of Waterloo, Waterloo, ON, Canada, from 2009 to 2010. He is currently a Professor with the School of Electronic Science and Engineering, National University of Defense Technology. His research interests include remote sensing, synthetic aperture radar (SAR) image processing, change detection, SAR ground moving target indication, and the classification of polarimetric SAR images.