COMBINING SORTED RANDOM FEATURES FOR TEXTURE CLASSIFICATION

1Li Liu, 2Paul Fieguth, 3Gangyao Kuang

1School of Electronic Science and Engineering, National University of Defense Technology, Changsha, China 410073
2Department of System Design Engineering, University of Waterloo, Waterloo, Canada N2L 3G1
Email: dreamliu2010@gmail.com, pfieguth@uwaterloo.ca, kuangyates@vip.sina.com

ABSTRACT

This paper explores the combining of powerful local texture descriptors and the advantages over single descriptors for texture classification. The proposed system is composed of three components: (i) highly discriminative and robust sorted random projections (SRP) features; (ii) a global Bag-of-Words (BoW) model; and (iii) the use of multiple kernel Support Vector Machines (SVMs) combining multiple features. The proposed system is also very simple, stemming from (1) the effortless extraction of the SRP features, (2) the simple orderless histogramming in the BoW model, (3) a strategy with low computational complexity for multiple kernel SVMs.

We have tested our texture classification system on three popular and challenging texture databases and find that the SVMs combining of SRP features produces outstanding classification results, outperforming the state-of-the-art for CUReT (99.37%) and KTH-TIPS (99.29%), and with highly competitive results for UIUC (98.56%).

Index Terms— Texture classification, random projection, compressed sensing, rotation invariance, support vector machines, kernel methods.

1. INTRODUCTION

Texture is a fundamental characteristic of the appearance of virtually all natural surfaces and is a powerful visual cue. The classification of textures is a fundamental human ability and an important, yet elusive, goal for computer vision research. The basic building components in the design of robust texture classification systems are (i) local highly discriminative and robust texture features, (ii) non-local statistical representations of local features, (iii) the design of a distance/similarity measure, and (iv) the choice of classifier.

Recent years have seen significant interest in the paradigm of a Bag-of-Words approach which enjoys the advantage of powerful local texture descriptors, but representing textures non-locally by the distribution of local textons [1, 2, 3, 4, 5, 6]. Undoubtedly, discriminative and robust texture features are a crucial factor in superior texture classification; a variety of local texture descriptors have been proposed recently [1, 2, 3, 4, 5, 6]. However, no method significantly outperforms the others, so some sort of feature combining seems relevant.

Of the possible features to combine, the Random Projections (RP) [2, 3] and SRP of Liu et al. are attractive – universal, information-preserving, dimensionality-reducing. They claim that the performance achieved by these random features, despite the use of a relatively simple nearest-neighbor classifier, can outperform the state-of-the-art in patch features, LBP and various filter bank-based methods.

Therefore motivated by the excellent classification results reported in [2, 3], this paper seeks to build on those results by coupling the random features with a more substantial classification scheme:

1. The use of SVMs rather than nearest neighbor, and
2. The combining of multiple features.

Combining descriptors has been explored in [1, 4] in texture classification and texture material categorization. The works in [1, 4] are sparse approaches, and a fixed combination of different region detectors and region descriptors is tried. The method of Varma and Ray [10] is based on multiple kernel learning (MKL), where they attempted to learn optimal combinations of local texture features. They demonstrated better classification performance can be obtained, however their approach increases the classifier complexity significantly.

2. BACKGROUND

A BoW approach represents an image as a collection of regions described by some local descriptors, spatially possibly sparse [1, 4] or dense [3, 5, 6, 7, 8]. An interesting alternative, the so-called MFS-based approach, was proposed by Xu et al. [17, 18] where, as opposed to sparse and dense approaches, the MFS approach characterizes the marginal histogram bins of the extracted features using fractal geometry, and this characterization encodes the spatial distribution of the image pixels in the bin.

Rather than a specialized feature extractor, tuned to a particular texture database, random projection (RP) [14] refers to the technique of projecting a set of points from a high-dimensional space to a randomly chosen low-dimensional subspace. The technique has been used for combinatorial optimization, information retrieval, face recognition and machine learning. Random features represent a computationally simple and efficient means of preserving texture structure without introducing significant distortion.

The information-preserving and dimensionality reduction power of RP is firmly demonstrated by the theory of compressed sensing (CS) [11, 12], which states that for sparse and compressible signals, a small number of nonadaptive linear measurements in the form of random projections can capture most of the salient information in the signal. Moreover, RP also provides a feasible solution to the well-known Johnson-Lindenstrauss (JL) lemma [14], which states that a point set in a high-dimensional Euclidean space can be mapped down onto a space of dimension logarithmic in the number of points with the distances between the points approximately preserved. RP plays an important role in both JL embedding and CS.

3. PROPOSED TEXTURE CLASSIFICATION

3.1. A Review of Sorted Random Features

The simple and efficient SRP features were proposed by Liu et al. [3] for rotation-invariant texture classification. In this paper we use
three different SRP features, illustrated in Fig. 1. The SRP takes
the sorted raw pixel intensities or intensity differences in a circular
neighborhood to form a feature vector \( y \) which is then transformed
to a lower-dimensional vector by a random projection matrix \( \Phi \), i.e.
\[
y_{\text{Circ}} = \Phi y_{\text{Circ}}.
\]

According to the results reported in [3], of the three proposed
SRP methods the Radial-Diff approach consistently performs the
best, most likely because differences capture more meaningful im-
ages.\footnote{In [1], Lazebnik et al. proposed combining descriptors capturing complementary in-
formation, we intend to combine the three SRP descriptors for tex-
ture classification, with the expectation that combined SRP features
would be richer and more robust than a single one.}

Since the intensity-based SRP feature and difference-based SRP
features are somewhat similar to the SPIN and RIFT descriptors,
resembling the work of Lazebnik et al. [1] and Zhang et al. [4],
who proposed combining descriptors capturing complementary in-
formation, we intend to combine the three SRP descriptors for tex-
ture classification, with the expectation that combined SRP features
would be richer and more robust than a single one.

3.2. Single SRP Feature

Textures are modeled by the joint distribution of a given SRP fea-
ture. This distribution is then represented by texton frequencies, and
textons and texture models are learned from training images (details
in [2, 3]). Classification of a novel image proceeds by mapping the
image to a texton distribution and comparing this distribution to the
learnt models. More specifically, the texture classification frame-
work includes the following steps:

1. Universal texton dictionary learning stage, in which a univer-
sal texton dictionary is learned by clustering one of the SRP
features aggregated over training images from the same tex-
ture class.

2. Histogram of textons learning stage, in which a histogram \( h \)
of textons is learnt for each particular training sample by la-
bling each of its pixels with the closest texton. Each texture
class then is represented by a set of models \( \{ h \} \) corresponding
to the training samples of that class.

3. The classification stage, where the process to compute the
normalized histogram of textons \( h_{\text{new}} \) for a novel image is the
same as for each training sample. The calculated model \( h_{\text{new}} \)
is classified into one of the known classes, based on a his-
togram distance metric, such as the \( \chi^2 \) statistic:
\[
\chi^2(h_{\text{new}}, h) = \frac{1}{2} \sum_{k=1}^{n} \left( \frac{h_{k}(h_{\text{new}}) - h_{k}(h)}{h_{k}(h) + h_{k}(h_{\text{new}})} \right)^2.
\]

3.3. Combining SRP Features

The benefits of SVMs for histogram-based classification is clearly
demonstrated in [4, 8]. Although SVMs were originally designed
for binary classification, texture classification is multi-class, so we
use the one-against-one technique, which trains a classifier for each
possible pair of classes.

Recent approaches to texture classification [1, 4, 10] have
demonstrated that combining several types of descriptors in a single
classifier can significantly boost the classification performance. Fur-
thermore, [1, 4] suggest the use of multiple complementary features,
features providing orthogonal information. In [10], Varma and Ray
combine many local descriptors in a kernel SVMs framework, and
showed that the learned kernel yields superior classification results.

Since the descriptors in this paper (especially SRP Rad-Diff) are,
on their own, already very discriminative, there may be limitations to
applying MKL; furthermore, simple kernel combination methods are
capable of reaching the same classification accuracy as MKL. There-
fore, we propose to combine kernels in a pre-defined deterministic
way and subsequently use the resulting kernel for SVMs training.

To incorporate the \( \chi^2 \) distance into the SVMs framework, we
use the kernel \( K(h_{\text{new}}, h) = \exp(-\gamma h^2(h_{\text{new}}, h)) \). In our case, when
multiple descriptor types are used, we represent each texture sample
using \( F \) Bag-of-Words histograms derived from \( F \) feature descrip-
tors. The multiple kernel method we consider is to combine several
kernels by multiplication. Richer representations can be achieved
in such case, since taking products of kernels corresponds to taking
a tensor product of their feature spaces, leading to a much higher
three different SRP features are illustrated in Fig. 1. The SRP takes
the sorted raw pixel intensities or intensity differences in a circular
neighborhood to form a feature vector \( y \) which is then transformed
to a lower-dimensional vector by a random projection matrix \( \Phi \), i.e.

\[
y_{\text{Circ}} = \Phi y_{\text{Circ}}.
\]

According to the results reported in [3], of the three proposed
SRP methods the Radial-Diff approach consistently performs the
best, most likely because differences capture more meaningful im-
ages.\footnote{In [1], Lazebnik et al. proposed combining descriptors capturing complementary in-
formation, we intend to combine the three SRP descriptors for tex-
ture classification, with the expectation that combined SRP features
would be richer and more robust than a single one.}

Since the intensity-based SRP feature and difference-based SRP
features are somewhat similar to the SPIN and RIFT descriptors,
resembling the work of Lazebnik et al. [1] and Zhang et al. [4],
who proposed combining descriptors capturing complementary in-
formation, we intend to combine the three SRP descriptors for tex-
ture classification, with the expectation that combined SRP features
would be richer and more robust than a single one.

3.2. Single SRP Feature

Textures are modeled by the joint distribution of a given SRP fea-
ture. This distribution is then represented by texton frequencies, and
textons and texture models are learned from training images (details
in [2, 3]). Classification of a novel image proceeds by mapping the
image to a texton distribution and comparing this distribution to the
learnt models. More specifically, the texture classification frame-
work includes the following steps:

1. Universal texton dictionary learning stage, in which a univer-
sal texton dictionary is learned by clustering one of the SRP
features aggregated over training images from the same tex-
ture class.

2. Histogram of textons learning stage, in which a histogram \( h \)
of textons is learnt for each particular training sample by la-
bling each of its pixels with the closest texton. Each texture
class then is represented by a set of models \( \{ h \} \) corresponding
to the training samples of that class.

3. The classification stage, where the process to compute the
normalized histogram of textons \( h_{\text{new}} \) for a novel image is the
same as for each training sample. The calculated model \( h_{\text{new}} \)
is classified into one of the known classes, based on a his-
togram distance metric, such as the \( \chi^2 \) statistic:

\[
\chi^2(h_{\text{new}}, h) = \frac{1}{2} \sum_{k=1}^{n} \left( \frac{h_{k}(h_{\text{new}}) - h_{k}(h)}{h_{k}(h) + h_{k}(h_{\text{new}})} \right)^2.
\]

3.3. Combining SRP Features

The benefits of SVMs for histogram-based classification is clearly
demonstrated in [4, 8]. Although SVMs were originally designed
for binary classification, texture classification is multi-class, so we
use the one-against-one technique, which trains a classifier for each
possible pair of classes.

Recent approaches to texture classification [1, 4, 10] have
demonstrated that combining several types of descriptors in a single
classifier can significantly boost the classification performance. Fur-
thermore, [1, 4] suggest the use of multiple complementary features,
features providing orthogonal information. In [10], Varma and Ray
combine many local descriptors in a kernel SVMs framework, and
showed that the learned kernel yields superior classification results.

Since the descriptors in this paper (especially SRP Rad-Diff) are,
on their own, already very discriminative, there may be limitations to
applying MKL; furthermore, simple kernel combination methods are
capable of reaching the same classification accuracy as MKL. There-
fore, we propose to combine kernels in a pre-defined deterministic
way and subsequently use the resulting kernel for SVMs training.

To incorporate the \( \chi^2 \) distance into the SVMs framework, we
use the kernel \( K(h_{\text{new}}, h) = \exp(-\gamma h^2(h_{\text{new}}, h)) \). In our case, when
multiple descriptor types are used, we represent each texture sample
using \( F \) Bag-of-Words histograms derived from \( F \) feature descrip-
tors. The multiple kernel method we consider is to combine several
kernels by multiplication. Richer representations can be achieved
in such case, since taking products of kernels corresponds to taking
a tensor product of their feature spaces, leading to a much higher
dimensional feature representation and corresponding SVMs kernel

\[
K'(h_{\text{new}}, h) = \prod_{i=1}^{F} K_i(h_{\text{new}}, h).
\]

4. EXPERIMENTAL RESULTS

To make the comparisons as meaningful as possible, we use the same
experimental settings as in [2, 3]. Each sample is normalized to be
zero mean and unit standard deviation, and the extracted SRP vec-
tor is normalized via Weber’s law. All results are reported over 50
random partitions of training and testing. The kernel parameters are
found by cross-validation within the training set. The values of the
parameters and of SVMs are specified using a grid search scheme.
In this work, the publicly available LibSVM library is employed.
The parameters \( C \) and \( \gamma \) are searched exponentially in the ranges
of \([2^{-5}, 2^{5}]\) and \([2^{-15}, 2^{5}]\), respectively, with a step size of \(2^1\)
to probe the highest classification rate.

To compare the performance single features with that of combi-
inations of features, we consider the three SRP descriptors. A first
test examined by overall performance of the product and average
kernels, with the product kernel performing slightly better, thus we
have decided against showing results for the average kernel in this
paper.

Fig. 2(a,b) and Table 2 show results for four datasets, comparing
the combined descriptors with the best single one (SRP Radial-Diff).

\[
\chi^2(h_{\text{new}}, h) = \frac{1}{2} \sum_{k=1}^{n} \left( \frac{h_{k}(h_{\text{new}}) - h_{k}(h)}{h_{k}(h) + h_{k}(h_{\text{new}})} \right)^2.
\]
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.

Fig. 2 (c,d) compares our approach with the state-of-the-art of Zhang et al. [4] and Lazebnik et al. [1], who have attempted to combine local RIFT, SIFT and SPIN descriptors. Our method improves on the state-of-the-art on D^\text{UIUC} when sufficient training data is available. For D^\text{KT} our approach significantly outperforms competing methods.

Table 3 gives a comprehensive summary of the results for our proposed approach against 12 recent state-of-the-art results. We can observe that our approach scores very well across all three commonly used datasets, producing what we believe to be the best reported result on the CUREt and KTH-TIPS databases, and very nearly meeting the best reported result for UIUC. It needs to be emphasized that our method is universal and achieved this state-of-the-art performance without any database-specific parameter tuning.

5. CONCLUSION AND FUTURE WORK

This paper explored the combination of SRP features using multiple kernel SVMs for texture classification. Combining SRP features is found to produce consistently better classification performance than a single SRP feature. We have tested our texture classification system on three popular and challenging texture databases, and the experimental results yield the best classification rates of which we are aware of 99.37% for CUREt and 99.29% for KTH-TIPS.

6. REFERENCES

Fig. 2. Classification rate vs. number of training samples on datasets $D_{KT}$ and $D_{UIUC}$: the left image compares single and combined classifiers, and the right image compares our proposed classifier with two state-of-the-art approaches from [4] and [1].


