COMBINING SORTED RANDOM FEATURES FOR TEXTURE CLASSIFICATION

[1]*Li Liu,* [2]*Paul Fieguth,* [1]*Gangyao Kuang*

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

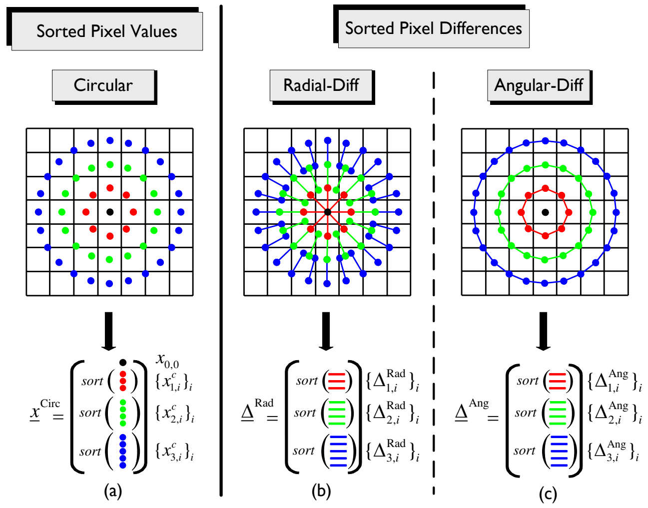

Fig. 1. Three sorting schemes on a local image patch of size 7×7: sorting pixels (a) or sorting pixel differences (b, c).

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 $\underline{\boldsymbol{x}}$, which is then transformed to a lower-dimensional vector by a random projection matrix $\mathbf{\Phi}$, *i.e.* $\underline{\boldsymbol{y}}^{\text{Circ}} = \mathbf{\Phi}\underline{\boldsymbol{x}}^{\text{Circ}}$.

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 image patterns than individual pixels, and that differences fall within a narrower range than pixel values, consequently providing a more compact description of texture.

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.* [1] and Zhang *et al.* [4] who proposed combining descriptors capturing complementary information, we intend 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.

3.2. Single SRP Feature

Textures are modeled by the joint distribution of a given SRP feature. 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 framework includes the following steps:

1. Universal texton dictionary learning stage, in which a universal texton dictionary is learned by clustering one of the SRP features aggregated over training images from the same texture class.

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

3. The classification stage, where the process to compute the normalized histogram of textons $\underline{\boldsymbol{h}}_{new}$ for a novel image is the same as for each training sample. The calculated model $\underline{\boldsymbol{h}}_{new}$ is classified into one of the known classes, based on a histogram distance metric, such as the χ^2 statistic: $\chi^2(\underline{\boldsymbol{h}}_1, \underline{\boldsymbol{h}}_2) = \frac{1}{2}\sum_{k=1} \frac{[\underline{\boldsymbol{h}}_1(k) - \underline{\boldsymbol{h}}_2(k)]^2}{\underline{\boldsymbol{h}}_1(k) + \underline{\boldsymbol{h}}_2(k)}$

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. Furthermore, [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. Therefore, we propose to combine kernels in a pre-defined deterministic way and subsequently use the resulting kernel for SVMs training.

To incorporate the χ^2 distance into the SVMs framework, we use the kernel $\mathbf{K}(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j) = \exp(-\gamma\chi^2(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j)$. In our case, when multiple descriptor types are used, we represent each texture sample using F Bag-of-Words histograms derived from F feature descriptors. 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 $\mathbf{K}^*(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j) = \prod_{l=1}^{F} \mathbf{K}_l(\underline{\boldsymbol{h}}_i, \underline{\boldsymbol{h}}_j)$.

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 vector 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 γ are searched exponentially in the ranges of $\left[2^{-5}, 2^{18}\right]$ and $\left[2^{-15}, 2^{8}\right]$, respectively, with a step size of 2^1 to probe the highest classification rate.

To compare the performance single features with that of combinations of features, we consider the three SPR 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). What is clear from both the table and the figure is that, uniformly

Table 1. Summary of texture datasets used in classification.

Texture Dataset	Dataset Notation	Image Rotation	Controlled Illumination	Scale Variation	Significant Viewpoint	Texture Classes	Sample Size	Samples per class	Samples in Total
CUReT	$\mathcal{D}^{\text{C}}$	$\surd$	$\surd$			61	200×200	92	5612
CUReTRot	$\mathcal{D}^{\text{CRot}}$	$\surd$	$\surd$			61	140×140	92	5612
UIUC	$\mathcal{D}^{\text{UIUC}}$	$\surd$		$\surd$	$\surd$	25	640×480	40	1000
KTH-TIPS	$\mathcal{D}^{\text{KT}}$		$\surd$	$\surd$		10	200×200	81	810

Table 2. A comparison of single and combined SRP results, with combinations of Radial-Diff (R), Circular (C) and Angular-Diff (A), applied to four datasets: $\mathcal{D}^{\text{C}}$, $\mathcal{D}^{\text{KT}}$, $\mathcal{D}^{\text{UIUC}}$ and $\mathcal{D}^{\text{CRot}}$. The tested patch sizes are 13×13 ($\mathcal{D}^{\text{C}}$ and $\mathcal{D}^{\text{KT}}$), 17×17 ($\mathcal{D}^{\text{UIUC}}$), and 11×11 ($\mathcal{D}^{\text{CRot}}$).

(a) **CUReT (92 samples per class in total)**

Features			Number of training samples per class						
D	C	A	2	10	18	26	34	38	46
$\surd$	$\surd$	$\surd$	**68.07**%	**91.77**%	**97.45**%	**98.31**%	**98.98**%	**99.31**%	**99.37**%
$\surd$	$\surd$		66.33%	91.44%	96.97%	98.13%	98.78%	99.13%	99.28%
$\surd$			64.88%	91.30%	95.71%	97.57%	98.22%	98.53%	99.05%

(b) **KTHTIPS (81 samples per class in total)**

Features			Number of training samples per class						
D	C	A	5	10	20	25	30	35	40
$\surd$	$\surd$	$\surd$	**81.18**%	88.99%	96.10%	**97.74**%	98.38%	98.71%	99.06%
$\surd$	$\surd$		80.90%	**89.49**%	**96.40**%	97.32%	**98.40**%	**99.07**%	**99.29**%
$\surd$			79.72%	88.93%	95.81%	97.45%	98.22%	98.62%	99.01%

(c) **UIUC (40 samples per class in total)**

Features			Number of training samples per class						
D	C	A	1	5	10	13	15	18	20
$\surd$	$\surd$	$\surd$	**61.82**%	90.84%	**96.61**%	**97.42**%	**97.89**%	**98.30**%	**98.56**%
$\surd$	$\surd$		61.62%	**90.96**%	96.00%	97.14%	97.59%	98.13%	98.42%
$\surd$		$\surd$	58.15%	89.55%	95.53%	96.53%	97.10%	97.72%	98.08%
$\surd$			59.00%	89.84%	95.67%	96.69%	97.31%	97.75%	98.30%

(a) **CUReTRot (92 samples per class in total)**

Features			Number of training samples per class						
D	C	A	2	10	18	26	34	38	46
$\surd$	$\surd$	$\surd$	**66.64**%	**89.43**%	**95.75**%	**96.75**%	**97.93**%	**98.28**%	**98.62**%
$\surd$	$\surd$		65.05%	88.87%	95.36%	96.61%	97.67%	98.13%	98.49%
$\surd$			62.28%	87.78%	92.99%	94.98%	95.95%	96.57%	96.87%

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 $\mathcal{D}^{\text{UIUC}}$ when sufficient training data is available. For $\mathcal{D}^{\text{KT}}$ 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. Combing 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.

Table 3. A comparison of the proposed combined SRP features with *12* state-of-the-art approaches on $\mathcal{D}^{\text{C}}$, $\mathcal{D}^{\text{KT}}$ and $\mathcal{D}^{\text{UIUC}}$. The number of training images per class for all results in the table are 46 for $\mathcal{D}^{\text{C}}$, 40 for $\mathcal{D}^{\text{KT}}$ and 20 for $\mathcal{D}^{\text{UIUC}}$. Scores are as originally reported, except for those marked $(*)$ which are taken from Zhang *et al.* [4].

	$\mathcal{D}^{\text{C}}$	$\mathcal{D}^{\text{KT}}$	$\mathcal{D}^{\text{UIUC}}$
SRP Radial-Diff	$\surd$	$\surd$	$\surd$
SRP Circular	$\surd$	$\surd$	$\surd$
SRP Angular-Diff	$\surd$		$\surd$
1. Our Results	**99.37**%	**99.29**%	**98.56**%
2. Varma and Zisserman-MR8 [6]	97.43%		
3. Varma and Zisserman-Patch [5]	98.03%	92.4%$(*)$	97.83%
4. Hayman *et al.* [8]	98.46%	94.8%$(*)$	92.0%$(*)$
5. Lazebnik *et al.* [1]	72.5%$(*)$	91.3%$(*)$	96.03%
6. Mellor *et al.* [15]		89.71%	
7. Zhang *et al.* [4]	95.3%	96.1%	**98.7**%
8. Brodhurst [16]	**99.22**%		
9. Varma and Ray [9]			**98.76**%
10. Crosier and Griffin [7]	98.6%	98.5%	**98.8**%
11. Xu-OTF *et al.* [17]			97.40%
12. Xu-WMFS *et al.* [18]			**98.60**%
13. Liu *et al.* [3]	98.52%	97.71%	96.27%

6. REFERENCES

[1] 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, August 2005.

[2] L. Liu, and P. Fieguth, “Texture classification from random features,” *IEEE Trans. Pattern Anal. Mach. Intell.*, *To Appear*.

[3] L. Liu and P. Fieguth, and G. Kuang, “Compressed sensing for robust texture classification,” in *Asian Conf. on Computer Vision (ACCV)*, 2010.

[4] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, “Local features and kernels for classification of texture and object categories: a comprehensive study,” *Int. J. Comput. Vision*, vol. 73, no. 2, pp. 213-238, 2007.

[5] M. Varma, and A. Zisserman, “A statistical approach to material classification using image patches,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 2032-2047, November. 2009.

[6] M. Varma, and A. Zisserman, “A statistical approach to texture classification from single images,” *Int. J. Comput. Vision*, vol. 62, no. 1-2, pp. 61-81, 2005.

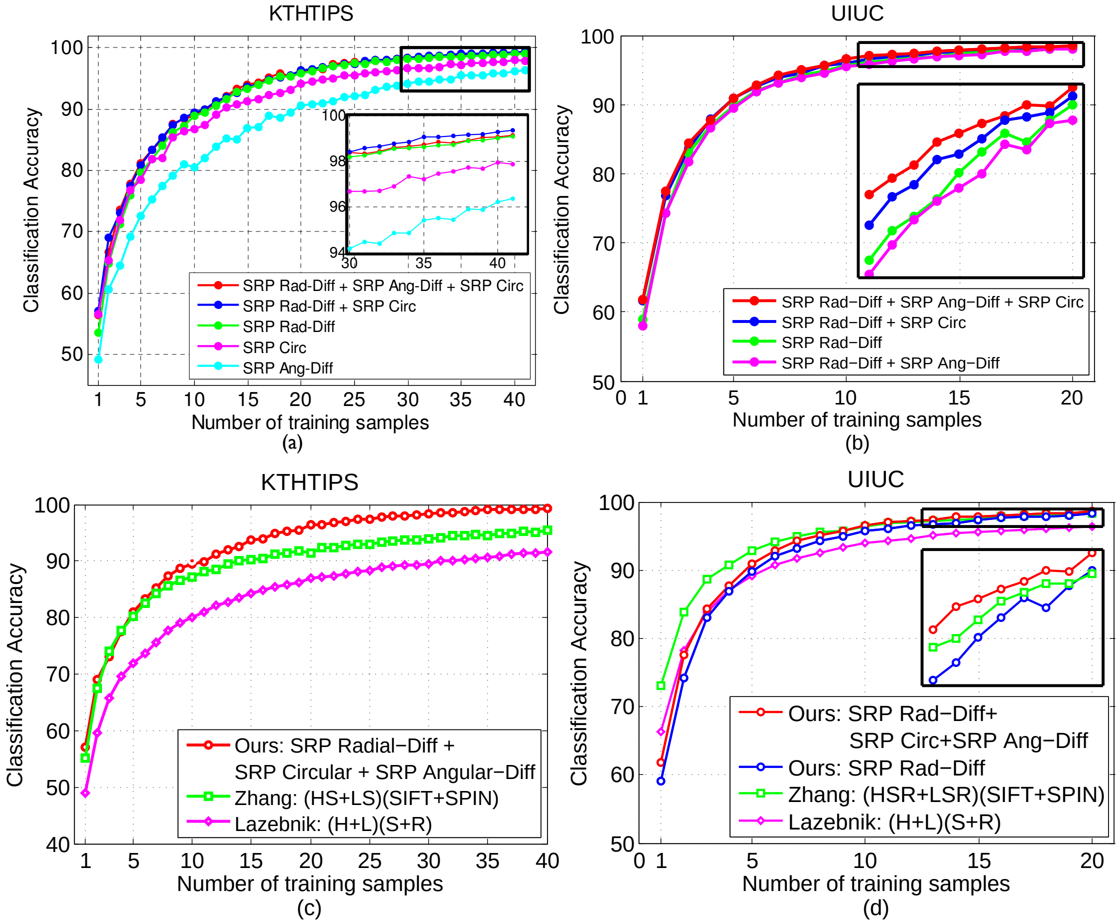

Fig. 2. Classification rate vs. number of training samples on datasets $\mathcal{D}^{\mathrm{KT}}$ and $\mathcal{D}^{\mathrm{UIUC}}$: the left image compares single and combined classifiers, and the right image compares our proposed classifier with two state-of-the-art approaches from [4] and [1].

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

[8] E. Hayman, B. Caputo, M. Fritz, and J.-O. Eklundh, "On the significance of real-world conditions for material classification," in *European Conf. on Computer Vision (ECCV)*, 2004, vol. 4, pp. 253-266.

[9] M. Varma and R. Garg, "Locally invariant fractal features for statistical texture classification," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2007, pp. 1-8.

[10] M. Varma, and R. Garg, "Locally invariant fractal features for statistical texture classification," in *IEEE Int. Conf. on Computer Vision (ICCV)*, 2007, pp. 14-21. ICCV , 2007.

[11] E. Candès, and T. Tao, "Near-optimal signal recovery from random projections: universal encoding stratigies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406-5425, December, 2006.

[12] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289-1306, April, 2006.

[13] W. B. Johnson, and J. Lindenstrauss, "Extensions of lipschitz mappings into a Hilbert space," in *Conf. in Modern Analysis and Probability*, 1984, pp. 189-206.

[14] S. Dasgupta, and A. Gupta, "An elementary proof of a theorem of Johnson and Lindenstrauss," *Random Structures and Algorithms*, vol. 22, no. 1, pp. 60-65, 2003.

[15] 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, January, 2008.

[16] R. E. Broadhurst, "Statistical estimation of histogram variation for texture classification," in *Proceedings of the fourth internaltional workshop on texture analysis and synthesis*, 2005, pp. 1597-1604.

[17] Y. Xu, S. Huang, H. Ji, and C. Fermüller, "Combining powerful local and global statistics for texture description," in *IEEE Int. Conf. of Computer Vision and Pattern Recognition (CVPR)*, 2009, pp. 573-580.

[18] Y. Xu, X. Yang, H. Ling and H. Ji, "A new texture descriptor using multifractal analysis in multi-orientation wavelet pyramid," in *IEEE Int. Conf. of Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 161-168.