STRUCTURED TEXTONS FOR TEXTURE CLASSIFICATION

Pengfei Xu, Xian-Ming Liu, Hongxun Yao, Yanhao Zhang

School of Computer Science and Technology, Harbin Institute of Technology, P.R. China
{pfxu, xmliu, h.yao, yhzhang}@hit.edu.cn

ABSTRACT

In this paper, we are proposing a novel texture descriptor, *Structured Texton*, for image representation. By exploring the local structure existed in the meaningful texture patterns, the structured textons are constructed by the nesting relationship between the different scale local extrema regions, which characterizes not only the local appearance features but also the geometric information. Furthermore, to improve the discriminability of the texture descriptors, higher order texton-words are constructed by means of finding informative patterns from the texton bank, which reduces the ambiguity of the textons through the higher order representation. We conduct two experiments as applications: the texture classification and scene classification. The improvements in the performance demonstrate the effectiveness of the proposed texture descriptors.

***Index Terms*—** Texture Descriptor, Texture Classification, Scene Classification, Structured Textons, Texton

1. INTRODUCTION

The *Texton*, which refers to the fundamental micro-structures in natural images especially the texture perception [1], is important in human cognition and computer vision [2], and is widely used in various applications including image segmentation [3] and texture analysis [4] *et. al.*. It is believed to be one of the basic elements in image analysis [2].

The extraction of textons is usually defined based on the response over a bank of global or local filters. The popular filter are Gabor bases, Laplacian-of-Gaussian and other wavelet transforms as adopted in [3]. Beyond the filters bank, the distribution and relationship between local pixel neighborhoods are also widely employed, especially as the popularity of s-parse coding and pooling techniques [2] [4] [5]. Another kind of textons are based on the local-point detection, which is detected by local point detector such as Laplacian detector [6].

As for the representation of textons, various machine learning methods are utilized, including discriminative [3] [5] and generative [2] methods, seeking for a complete or over-complete bases of texture patterns.

However, as an important factor of visual information, the structure of texton patterns is ignored. As shown in Figure 1

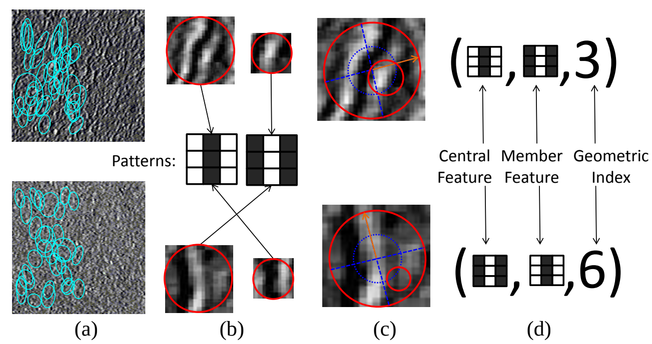


Fig. 1. (a) Two texture images of different categories in CURet dataset [7](“49-017.png” and “16-036.png”). The cyan ellipses denote the interest local regions detected by Laplacian detector. For clarity, only 20% of detected regions are displayed. (b) Two most frequency texture elements of different scales in each image of which the corresponding ellipses are mapped onto unit circles. The two texture images have high confusion due to the texton patterns of texture elements are very similar. (c) To address the problem, the nested relationship and the geometric information between the texture elements are characterized. (d) In our proposed Structured Texton representation, the two images are easily identified since the basic structure textons are very different.

(a) and (b), even the similar basic texton pattern with different structures will exhibit vastly different semantics.

To cope with this problem, we are trying to explore the structure of local texture patterns and hence build a structured texton representation for texture analysis in this paper. As for the structure, two aspects are explored.

First, the structure holds certain spatial and scale distribution, which is referred as the “*Nesting-Structure*”, as the example shown in Figure 1 (a). This “*Nesting*” relationship is widely existed in texture of natural images. On one hand, the texture will exhibit different descriptions on varied scales. On the other hand, as the examples in Figure 1, the texture patterns in real images are usually not uniform, instead they are interlaced distributed with other small texture blobs of different types. To fully explored these two factors, we embedded the nesting-relationship in the texture descriptor, which is ignored in previous researches.

Second, the correlation among the texture patterns further improve the discriminative power of textons, which is motivated by the success of visual phrase [8], and it is also validated in texture classification [9]. Considering the structure in the representation makes the obtained textons hold semantics and be more informative, as shown in Figure 1 (c) and (d), compared with traditional texture descriptors.

The structured textons are useful for various applications, including texture classification, logo detection, image synthesis, and image matching. In this paper, the structured texton is utilized to texture classification and scene classification respectively.

2. THE STRUCTURED TEXTONS

In this section, the construction of structured texton is introduced in details. The structured textons are constructed after building the the texture elements, and the discriminative high order structured textons are selected by the semantic ranking.

2.1. Building the Texture Elements

To get the representation of texture elements, we use the Laplacian blob detector to detect the affine regions, which are thought of as texture elements having a characteristic elliptic shape and a distinctive appearance pattern [10]. Similar with [10], we normalize these detected regions by mapping the corresponding ellipses onto a unit circle, and compute the SIFT-Like descriptors $d_i \in \mathbb{R}^{128}$ [11] for each detected region i. The scale and principle orientation information of each region are also preserved.

A dictionary is further obtained by performing the K-Means clustering on all the $\{d_i\} \in \mathbb{R}^{128}$ into M clusters, to quantize the descriptors of texture regions into the visual word ID v_i, following [12]. Some high frequency words of the dictionary are shown in Figure 2 (c) and (d). Finally, the texture regions are quantized into the collection of *Texture Elements* as $E = \{e_i\}_{i=1}^N$, $e_i = (x_i, y_i, s_i, \phi_i, r_i, v_i)$, where (x_i, y_i) denotes the position of the texture region, $s_i \in \mathbb{R}$ its scale, $\phi_i \in [-\pi, \pi)$ its orientation, and r_i the bounding circle of the SIFT local patch (mapped to a unit circle).

2.2. The Structured Textons

Inspired by the observation that local interest regions are usually in different scales and overlap with each other (as shown in Figure 1(a)), we found it is quite natural and effective to utilize the *nesting relationship*, which captures both the spatial and scale structural information among local interest regions to explore the structure of texture patterns.

The *nesting relationship* structure is demonstrated in Figure 1 (c), which consists a center feature c and a group of member features M_c nested by the bound r_c. More specific, we use the nesting structure to capture the regular and meaningful texture patterns, as called the *Structured Textons*. The geometric information is further incorporated into our structured texton, by iteratively overlaying a circular grid centered at each region in the central circle. The grid has a radius

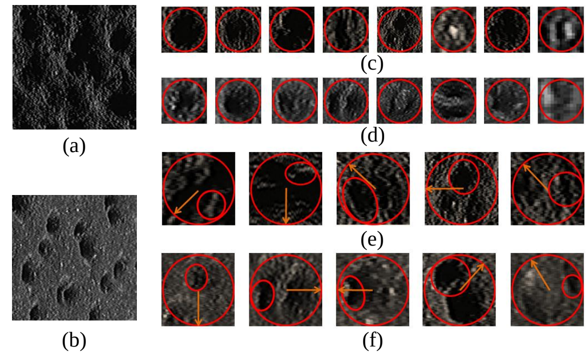


Fig. 2. (a) and (b) two sample images randomly selected from CURet dataset("Sample59" and "Sample60" in CURet dataset); (c) and (d) the high frequency textons of the two texture categories, respectively; (e) and (f) the discriminative 2-order structured textons.

proportional to the feature's scale, and the rotation invariance of the configurations is achieved by rotating the circular grid by the feature's orientation(as shown in Figure 3(a)), the bounding circle r_c is divided to $R \times S$ regions, denoted as $G = \{g_i\}_{i=0}^{R\times S-1}$.

Finally, the *Structured Textons* are formulated as the triple:

$$T = \{(c_i, M_i)\}, \tag{1}$$

where c_i $(i = 1, \ldots, N)$ denotes the ith central region, $M_i = \{(e_j, g_j)\}$ denotes the member feature set which is covered by c_i, and $g_j \in G$ the geometric label of member point e_j. It embeds both the spatial, scale and geometric information in the representation of textures.

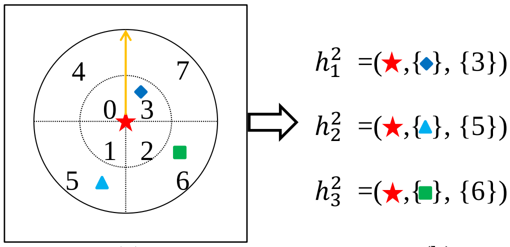
In order to reduce the dimension of the structure texton representation, and also preserve the correlations between the center feature c_i and member features M_i within structured textons, we decompose a structure texton $\tau \in T, \tau = (c_i, M_i)$ into several high order textons, $T_{ci} = \{(c_i, M_{ik})\}$, where $M_{ik} \subseteq M_i$ denotes the subset of M_i.

We further formulate the above decomposition into the n-order structure texton set ($n \geq 1$), $T_{ci}^n \subseteq T_{ci}$, s.t., $|M_{ik}| = n - 1$. If $n = 1$, the high order texton becomes simply texton (i.e., only the central texton c_i). For simplification, we only consider the 2-order texton in this paper.

2.3. Discriminative Structured Texton Generation

Due to the high repeatability and large amount the the 2nd order structured textons, we extract the discriminative structured textons using a data-driven approach, by observing the distribution of the whole feature space.

We consider the candidate high order frequencies within a certain category and the overall categories in training dataset. According to the TF-IDF weighting in information retrieval theory, a candidate high order texton is considered important



(a) (b)

Fig. 3. Toy examples of geometric configure in a structured texton and high order texton generation. Stars, triangles, rectangles and rhombs represent four kinds of local textons.(a) a structured texton consisting of the central feature and three member features with geometric grid index; (b) The structured texton is decomposed into three 2-order textons.

to an category if it appears more frequently appeared in it and less frequently in others. Based on this strategy, the importance of a high order texton t_i to the category C is computed as follows,

$$S_i^{(C)} = logN_i^{(C)}/logN_i, \quad (2)$$

where $N_i^{(C)}$ denotes the number of t_i in category C and N_i is the number of t_i in all categories in the entire feature space. Consequently, the top ranked structured textons in Eq. 2 are thought to be discriminative for category C and selected as the basic texture patterns.

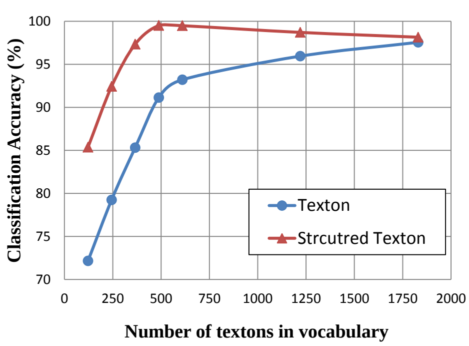


Fig. 4. Texture classification results of our proposed structured texton and texton method with different number of vocabularies on the CURet dataset.

3. EXPERIMENTS

To evaluate the discrimination and effectiveness of the proposed *Structured Texton*, we apply the proposed framework to two tasks: texture classification and scene classification.

3.1. Preliminary

We apply our proposed structured texton into classification task. First, we extract top K Laplacian regions (ranked by the responses of each region, $K = 400$ in our experiments) in each image in the training set, and then construct the structured textons according to the nesting relationship. Second, we generate the top V discriminative 2-order structured textons for each category and align all the discriminative 2-order structured textons of all C categories into a the structured texton dictionary. Finally, the structured textons in each image are quantized into a histogram of vocabulary words to characterize its texture pattern. Similar with the most previous methods, we train SVM classifier with the RBF kernel and classify the texture and scene images using the LibSVM [13].

3.2. Texture Classification

For the task of texture classification, three datasets are involved to evaluate our method, including 1) **Brodatz texture album** [14]: contains 111 texture images and each original image is extracted 9 samples as non-overlapping regions. We use 3 training samples and the rest for testing; 2) **KTH-TIPS dataset** [15]: consists of 10 texture classes with 81 images per class, in which images are captured at nine scales, viewed under three different illumination directions and three different poses. We use 40 images for training per class and the rest for testing; and 3) **CUReT texture dataset** [7]: consists of 61 texture classes with 92 images for each class that are captured under different illuminations with seven different viewpoint directions. We use 46 for training and 46 for testing per class.

The evaluation measurement follows [12], in which the reported classification rate is the average over 100 random training set/testing set partitions for all the datasets.

3.2.1. Comparison with the simple texton-based method

To evaluate the discrimination improvement of structured texton, we compare the proposed method with the simple texton-based method, which clusters the descriptors of detected regions into a vocabulary and represents each image with a histogram according to the vocabulary. The SVM classifier with RBF kernel is used for training and classifying in both simple texton-based method and our proposed structured texton method. Figure 4 illustrates the classification accuracies on the two methods with different size of vocabularies on the CURet dataset. From Figure 4, one can observe that the proposed method outperforms the simple texton-based method consistently. The structured texton method achieve the best accuracy, 99.51%, when the size of vocabulary is 488. The results demonstrate the effectiveness and discrimination of strucuted texton representation method.

3.2.2. Comparison with the state-of-the-art methods

We also compared our method against the state-of-the-art methods, 1) Multi-feature combination method (**Multi-feature**) [12], which combines SIFT, SPIN and RIFT as the descriptors of local regions to classify the texture image; 2) Texton-based SVM method (**Texton-SVM**) [16], in which training set is extended by re-scaling the samples for texture classification; and 3) Covariance structure feature

Table 1. Texture classification results (%) on Brodatz, KTH-TIPS, and CURet Datasets.

Methods	Brodatz	KTH-TIPS	CURet
Multi-feature [12]	95.9	96.1	95.3
Texton SVM [16]	-	90.56	98.00
Covariance structure [17]	96.14	95.74	-
Our proposed method	**97.42**	**99.57**	**99.51**

method (**Covariance structure**) [17], which uses multivariate log-gaussian cox processes to characterize the covariance structure in texture images.

Tab. 1 illustrates the results of different methods on all the benchmark datasets. Our proposed structured texton method not only outperforms the other texton-based methods (i.e., Multi-feature and Texton-SVM) but also achieves significant improvement than Covariance Structure method, which demonstrates the effectiveness of the proposed Structured texton representation. Furthermore, on the more challenging CURet dataset containing the the variance on scale, affine and rotation, the proposed *structured textons* also achieve fairly good performance, which also demonstrate its robustness.

3.3. Scene Classification

We extend our Structured texton to the scene classification task on the widely used 15 class scene dataset [18]. Images in the dataset are about 250×300 resolution, with 210 to 410 images per class. This dataset contains a wide range of outdoor and indoor scene environments, and the 15 scene categories are shown in Figure 5. In our experiment, 100 images in each category are used for training and the remaining images constitute the testing set, the same with previous researches.

We compare the proposed method with the spatial structures/spatial constraint-based state-of-the-art methods, including multi-level Spatial Pyramid Matching method (SPM) [18], the global GIST feature-based method (GIST) [19], and CENTRIST [20].

Table 2 illustrates the average accuracy in all categories of different methods. All related results are cited from original paper except GIST whose result is cited from [20]. Our proposed method outperforms the spatial structures/spatial constraint-based state-of-the-art methods with 3%-9% improvements in classification performance. The reason is that the proposed structured textons preserve the structure characteristic of different category scene images.

Fig. 5 shows a confusion matrix between the fifteen scene categories of our proposed method. Structured texton effectively classifys the different scenes, while confusion occurs between category pairs such as bedroom/living room, and coast/open country, which have very similar components.

3.4. Discussion

To summarize, the experimental results show that the *Structured Texton*:

Table 2. Classification performance on the 15 class scene dataset.

Methods	Performance
Single-level SPM(#voca.=200) [18]	79.4 ± 0.3
Single-level SPM(#voca.=400) [18]	79.7 ± 0.5
Multi-level SPM(#voca.=200) [18]	81.1 ± 0.3
Multi-level SPM(#voca.=400) [18]	81.4 ± 0.5
GIST [19]	73.28 ± 0.67
CENTRIST [20]	83.88 ± 0.76
Our proposed method	$\mathbf{86.60 \pm 0.68}$

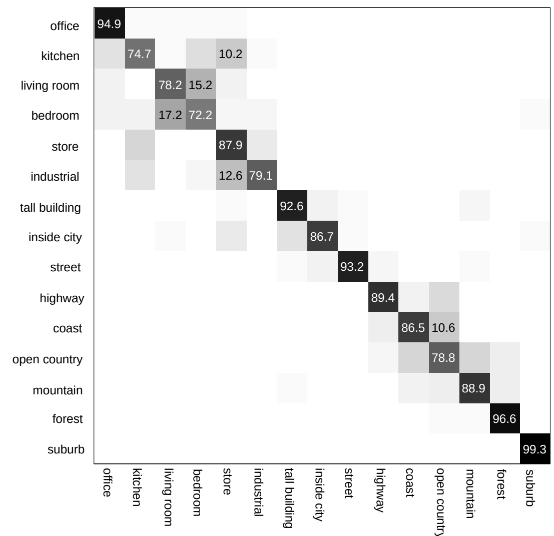


Fig. 5. Confusion matrix of the 15 class scene dataset. Only rates higher than 10.0% are shown in the figure

1. holds the structural information including spatial, scale distributions, geometrical, and correlations between textons, which leads to the improvement in performance in both texture classification and scene classification;

2. discovers the discriminative texture patterns to avoid the false alarm in texture analysis caused by high repeatability of textures in natural images.

4. CONCLUSION

In this paper, we have presented a novel texture descriptor, structured textons for image representation. Structure texton is advantageous in several aspects. First, it explores the meaningful local structure patterns in texture images; Second, we propose high order structured texton to preserve structure characteristic of texture elements; Third, we proposed a category-based method to select discriminative high order texton for classification task. Experimental results show that Structured texton achieves promising performance in both texture classification and scene classification applications.

5. REFERENCES

[1] B. Julesz, "Textons, the elements of texture perception, and their interactions.," *Nature*, 1981.

[2] S.C. Zhu, C.E. Guo, Y. Wang, and Z. Xu, "What are textons?," *International Journal of Computer Vision*, vol. 62, no. 1, pp. 121–143, 2005.

[3] J. Malik, S. Belongie, J. Shi, and T. Leung, "Textons, contours and regions: Cue integration in image segmentation," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. IEEE, 1999, vol. 2, pp. 918–925.

[4] M. Varma and A. Zisserman, "Texture classification: Are filter banks necessary?," in *Computer vision and pattern recognition, 2003. Proceedings. 2003 IEEE computer society conference on*. IEEE, 2003, vol. 2, pp. II–691.

[5] B.A. Olshausen, D.J. Field, et al., "Sparse coding with an overcomplete basis set: A strategy employed by vi?," *Vision research*, vol. 37, no. 23, pp. 3311–3326, 1997.

[6] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1265–1278, 2005.

[7] K.J. Dana, B. Van Ginneken, S.K. Nayar, and J.J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Transactions on Graphics (TOG)*, vol. 18, no. 1, pp. 1–34, 1999.

[8] J. Yuan, Y. Wu, and M. Yang, "Discovery of collocation patterns: from visual words to visual phrases," in *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*. IEEE, 2007, pp. 1–8.

[9] X. Qi, R. Xiao, J. Guo, and L. Zhang, "Pairwise rotation invariant co-occurrence local binary pattern," *Computer Vision–ECCV 2012*, pp. 158–171, 2012.

[10] S. Lazebnik, C. Schmid, and J. Ponce, "A sparse texture representation using local affine regions," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 8, pp. 1265–1278, 2005.

[11] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

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

[13] Chih-Chung Chang and Chih-Jen Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.

[14] P. Brodatz, *Textures: a photographic album for artists and designers*, vol. 66, Dover New York, 1966.

[15] E. Hayman, B. Caputo, M. Fritz, and J.O. Eklundh, "On the significance of real-world conditions for material classification," *Computer Vision-ECCV 2004*, pp. 253–266, 2004.

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

[17] H.G. Nguyen, R. Fablet, and J.M. Boucher, "Visual textures as realizations of multivariate log-gaussian cox processes," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 2945–2952.

[18] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE, 2006, vol. 2, pp. 2169–2178.

[19] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International Journal of Computer Vision*, vol. 42, no. 3, pp. 145–175, 2001.

[20] J. Wu and J.M. Rehg, "Centrist: A visual descriptor for scene categorization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1489–1501, 2011.