

## LETTER

# Texture Representation via Joint Statistics of Local Quantized Patterns

Tiecheng SONG<sup>†a)</sup>, *Nonmember*, Linfeng XU<sup>†</sup>, *Student Member*, Chao HUANG<sup>†</sup>, and Bing LUO<sup>†</sup>, *Nonmembers*

**SUMMARY** In this paper, a simple yet efficient texture representation is proposed for texture classification by exploring the joint statistics of local quantized patterns (jsLQP). In order to combine information of different domains, the Gaussian derivative filters are first employed to obtain the multi-scale gradient responses. Then, three feature maps are generated by encoding the local quantized binary and ternary patterns in the image space and the gradient space. Finally, these feature maps are hybridly encoded, and their joint histogram is used as the final texture representation. Extensive experiments demonstrate that the proposed method outperforms state-of-the-art LBP based and even learning based methods for texture classification.

**key words:** texture classification, Local Binary Patterns (LBP), Gaussian derivative filter, ternary coding

## 1. Introduction

Local visual features have been widely studied in various computer vision tasks such as image matching [1], segmentation [2], texture classification [3], co-pattern discovery [4] and face synthesis [5]. In the real world, texture images typically exhibit inter-class and intra-class variations such as rotation, illumination, scale and view changes, making the recognition task very difficult. Therefore, robust texture representation that can handle well these unknown variations is desired for most of these applications.

Up to now, many approaches have been proposed for texture analysis such as co-occurrence statistics, Markov random fields and Gabor filters. Recently, texton dictionary based methods [6], [7] were introduced for texture representation. These methods need to learn a texton dictionary, and involve intensive data-to-cluster computation. The Local Binary Pattern (LBP) was presented in [3] for texture analysis, which is operationally simple and robust to image rotation and monotonous illumination changes. These attractive properties make the LBP a good choice in texture classification [3] and face recognition [8].

A lot of variants based on the LBP were developed. To capture the discriminative patterns, Liao et al proposed the Dominant LBP (DLBP) [9] by considering the most frequently occurred texture patterns. To allow for distant pixel interaction, the global features of Gabor filter responses were employed in [9] to complement the DLBP features. For the same purpose, the Dominant Neighborhood Struc-

ture (DNS) was introduced as the global features [10] to complement the LBP features. To be robust to noise, Tan and Triggs [11] proposed the Local Ternary Pattern (LTP) to quantize the local differences into three levels. However, the split ternary coding adopted in the LTP can lead to some information loss [12]. Recently, the Completed LBP (CLBP) [12] and the Completed Local Binary Count (CLBC) [13] were proposed by jointly encoding the local difference sign, magnitude and the center pixel. Significant results have been achieved by the CLBP and CLBC for texture classification.

In this paper, we propose a simple yet efficient image representation for texture classification. The idea is to characterize a texture by exploring the joint statistics of local quantized (binary and ternary) patterns. To explore information of different domains, we first use the Gaussian first derivative filters to obtain the multi-scale gradient responses. These gradient responses together with the raw image are then quantized and encoded to generate three feature maps. Finally, a hybrid encoding scheme is adopted, and the joint statistical distribution is used to represent the texture. Since the gradient responses are calculated at few scales and the global thresholding is adopted in our ternary quantization, our method is computationally efficient. Moreover, unlike the textons based methods, our method needs no training and needs no costly clustering (via K-means). Extensive experimental results demonstrate the advantages of our proposed method.

## 2. The Proposed Method

In this section, we present a simple yet efficient texture representation by exploring the joint statistics of local quantized patterns (jsLQP). Our main contributions include: (1) The new local ternary patterns (i.e., proposed LTP) based on the multi-scale gradient responses are proposed via the global thresholding. (2) The joint information in the image and gradient domains is efficiently extracted and effectively represented in a hybrid way. As illustrated in Fig. 1, our method can be summarized as three steps: Gaussian derivative filtering, local pattern quantization and encoding, and joint histogram representation. The details are given below.

### 2.1 Gaussian Derivative Filtering

The image gradients can provide discriminative structural information. Exploring the joint statistics of local quantized

Manuscript received May 21, 2013.

<sup>†</sup>The authors are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China.

a) E-mail: tgwin@gmail.com

DOI: 10.1587/transinf.E97.D.155

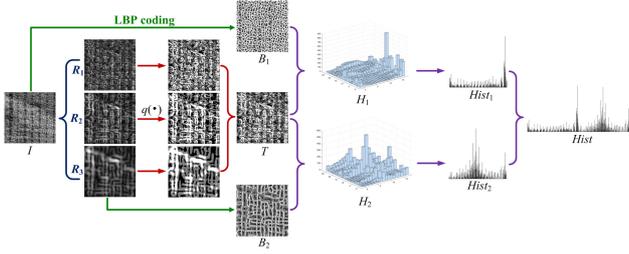


Fig. 1 The pipeline of our proposed method.

patterns in the image and gradient domains is expected to obtain robust and compact texture representation. In this step, we employ the Gaussian first derivative filters to obtain multi-scale gradient response maps. To be robust to the global illumination changes, the input image is normalized to have zero mean and unit standard deviation [7].

Let  $I$  be the normalized input image. The circularly symmetric Gaussian at the position  $(x, y)$  is defined as

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2+y^2)}{2\sigma^2}} \quad (1)$$

We compute the gradient response by

$$R(x, y) = \sqrt{(I \star G_x)^2 + (I \star G_y)^2} \quad (2)$$

where  $\star$  is the convolution operator, and  $G_x$  and  $G_y$  are the first derivative of the Gaussian in the  $x$  and  $y$  directions, respectively.

In our work, we adopt the Gaussian first derivative filters of size  $25 \times 25$  at three scales:  $\sigma_1 = 1$ ,  $\sigma_2 = 2$ ,  $\sigma_3 = 4$  (they are set based on our empirical study, which can achieve good classification results). The obtained three gradient maps  $R_1$ ,  $R_2$  and  $R_3$  are show in Fig. 1. Since these filters are computed at three scales, they are computationally efficient compared with the multi-scale and multi-orientation MR8 [7] filters.

## 2.2 Local Pattern Quantization and Encoding

In this step, we encode the local quantized binary and ternary patterns to generate feature maps both in the image space and the gradient space. It is important to extract the binary and ternary patterns in different domains because they can provide complementary information.

For binary patterns, the LBP operator [3] is applied to the input image and the selected gradient map  $R_3$  ( $\sigma = 4$ ) to encode the micro-structure information. The rotation invariant uniform pattern  $LBP_{P,R}^{riu2}(x, y)$  [3] is adopted to encode each pixel  $(x, y)$  ( $P$  is the total number of neighbors,  $R$  is the radius of the neighborhood). This results in  $P + 2$  possible LBP labels. Thus, the binary coding map  $B_1$  for the input image and  $B_2$  for the gradient map  $R_3$  can be obtained.

For ternary patterns, each gradient map  $R_i$  ( $i = 1, 2, 3$ ) is quantized into three levels via the global thresholding. Unlike the LTP [11] that uses the fixed threshold, our quantization is self-adapted to the image content and it reveals the

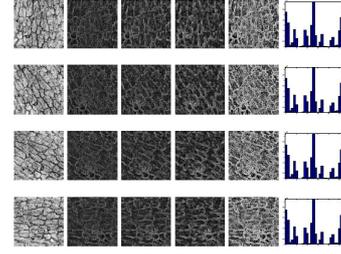


Fig. 2 Illustration of the rotation invariance of the ternary coding maps. Rows 1-4 are for the rotation angles of 0, 30, 60 and 90 degrees, respectively. Left to right: the raw texture, the corresponding gradient maps  $R_1$ ,  $R_2$ ,  $R_3$ , the feature map  $T$ , and the statistical histogram of  $T$ .

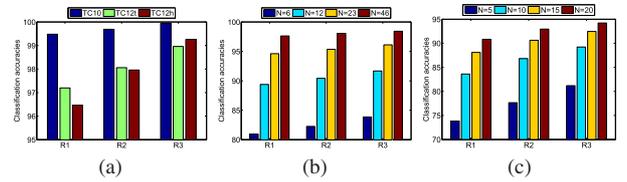


Fig. 3 Classification accuracies of the proposed method with different gradient maps on the (a) Outex database [14], (b) CURET database [7], and (c) UIUC [15] database, respectively.  $N$  denotes the number of training images used for classifier learning for each texture class.

relative gradient intensity of each pixel against the overall gradient intensity. Specifically, the quantization is accomplished by

$$t_i(x, y) = q(R_i(x, y)) = \begin{cases} 2, & R_i(x, y) > (1 + \tau)m_i \\ 0, & R_i(x, y) < (1 - \tau)m_i \\ 1, & \text{otherwise} \end{cases} \quad (3)$$

where  $\tau$  is a control parameter ( $\tau = 0.4$ ), and  $m_i$  is a global threshold depending on the mean of overall gradient responses:

$$m_i = \frac{1}{|I|} \sum_{x,y} R_i(x, y) \quad (4)$$

Subsequently, the ternary feature map  $T$  is generated by jointly encoding the ternary labels as follows

$$T(x, y) = \sum_{i=1}^3 t_i(x, y)3^{i-1}, \quad T(x, y) \in [0, 26] \quad (5)$$

Figure 2 illustrates the rotation invariance of the ternary coding maps for texture samples captured at different rotation angles. It can be seen that the generated gradient maps  $R_1$ ,  $R_2$  and  $R_3$  and their jointly encoded feature map  $T$  are statistically rotationally invariant. Figure 3 shows the classification accuracies of our method with different gradient maps, which can be selected to produce the binary coding map  $B_2$ . It demonstrates that the gradient map  $R_3$  ( $\sigma = 4$ ) is a better choice.

## 2.3 Joint Histogram Representation

In this step, we explore the joint pixel-wise information that

corresponds to different feature domains to characterize a texture. The texture features are represented as a joint histogram which is built in a hybrid way.

Specifically, the labels of  $B_1$  and the ternary feature map  $T$  are jointly encoded into a 2-D histogram:

$$H_1(l, p) = \sum_{x,y} \delta((T(x, y), B_1(x, y)) == (l, p)) \quad (6)$$

where  $(l, p)$  ( $l \in [0, 26], p \in [0, P + 1]$ ) is the entry of the histogram  $H_1$ , and

$$\delta(z) = \begin{cases} 1, & \text{if } z \text{ is true} \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

Similarly, the  $B_2$  and  $T$  are also jointly encoded to form a 2-D histogram  $H_2$ . Finally,  $H_1$  and  $H_2$  are both converted into 1-D histograms and concatenated to represent the texture (see Fig. 1).

In this way, the pixel-wise information of different feature domains is explored via the joint statistics of local quantized binary and ternary patterns. In our work, we set  $R = 2, P = 16$ . Therefore, we obtain a  $27 \times (P + 2) \times 2 = 972$  dimensional histogram representation.

### 3. Experimental Results

In this section, we compare the proposed jsLQP with state-of-the-art LBP based and learning based (VZ\_MR8 [6] and VZ\_Joint [7]) methods on the Outex [14], CURET [7] and UIUC [15] databases. Following the experimental setup in [12] and [13], the chi-square dissimilarity metric [3] and the nearest neighborhood classifier are adopted. The results for VZ\_MR8 and VZ\_Joint on the Outex and CURET databases are quoted from [12], and the results on the UIUC database are taken from [7].

#### 3.1 Experimental Results on the Outex Database

For the Outex database, experiments are conducted on two test suites: Outex\_TC\_00010 (TC10) and Outex\_TC\_00012 (TC12), each with 24 classes of textures collected under three illuminations (“horizon”, “inca” and “t184”) and nine rotation angles ( $0^\circ, 5^\circ, 10^\circ, 15^\circ, 30^\circ, 45^\circ, 60^\circ, 75^\circ$  and  $90^\circ$ ). There are 20 texture samples per class under each setting. The  $24 \times 20$  samples with illumination “inca” and angle  $0^\circ$  are used for classifier training. For TC10, the other samples under eight rotation angles with illumination “inca” are used for testing. For TC12, all samples under illumination “t184” or “horizon” are used for testing.

The comparison results for different methods are given in Table 1. The following remarks can be made. Firstly, the proposed jsLQP yields the best classification performance (with the average accuracy 99.39%) compared with other methods for all the test sets. It demonstrates the robustness of our method to rotation and illumination changes. Secondly, our method with compact representation outperforms the multi-scale CLBP by 2.73%. Specifically, the proposed

**Table 1** Classification accuracies (%) on the Outex database.

	TC10	TC12		Aver
		t184	horizon	
LTP (R=2, P=16)	96.95	90.16	86.94	91.35
LTP (R=3, P=24)	98.20	93.59	89.42	93.74
Proposed LTP	96.90	95.65	95.46	96.00
CLBP_S (R=2, P=16)	89.40	82.27	75.21	82.29
CLBC_S (R=2, P=16)	88.67	82.57	77.41	82.88
CLBP_S (R=3, P=24)	95.08	85.05	80.79	86.97
CLBC_S (R=3, P=24)	91.35	83.82	82.75	85.97
CLBP_S/M/C (R=2, P=16)	98.72	93.54	93.91	95.39
CLBC_S/M/C (R=2, P=16)	98.54	93.26	94.07	95.29
CLBP_S/M/C (R=3, P=24)	98.93	95.32	94.54	96.26
CLBC_S/M/C (R=3, P=24)	98.78	94.00	93.24	95.67
CLBC_CLBP (R=2, P=16)	98.83	93.59	94.26	95.56
CLBC_CLBP (R=3, P=24)	98.96	95.37	94.72	96.35
Multi-scale CLBP (R=1, 2, 3)	99.17	95.23	95.58	96.66
Multi-scale CLBC (R=1, 2, 3)	99.04	94.10	95.14	96.09
DLBP+NGF [9] (R=2, P=16)	99.1	93.2	90.4	94.2
DLBP+NGF [9] (R=3, P=24)	98.2	91.6	87.4	92.4
DNS+LBP [10] (R=2, P=16)	98.90	93.22	92.13	94.75
DNS+LBP [10] (R=3, P=24)	99.27	94.40	92.85	95.51
VZ_MR8	93.59	92.55	92.82	92.99
VZ_Joint	92.00	91.41	92.06	91.82
<b>Proposed jsLQP</b>	<b>99.95</b>	<b>98.96</b>	<b>99.26</b>	<b>99.39</b>

jsLQP has 972 dimensions while the multi-scale CLBP has  $2 \times (10 \times 10 + 18 \times 18 + 26 \times 26) = 2200$  dimensions.

#### 3.2 Experimental Results on the CURET Database

The CURET database<sup>†</sup> contains 61 classes of textures. There are 92 images per class that are captured at different view-points and illumination conditions. Following [12], [13],  $N$  training images are randomly chosen from each class, and the remaining  $92 - N$  images are used for testing.

The classification results averaged over 100 random splits of the training and test sets are presented in Table 2. For this database, the proposed jsLQP again performs best. Particularly, it has 1.43%, 2.55%, 4.02% and 5.6% performance improvements over the multi-scale CLBP for 46, 23, 12 and 6 training samples, respectively. Our method also outperforms the learning based VZ\_MR8 [6] and VZ\_Joint [7] by a large margin. Note that the dimension of our jsLQP is much lower than that of VZ\_MR8 and VZ\_Joint (i.e., 972 versus 2440). It demonstrates the power of the joint statistics of local quantized patterns. In addition, the overall performance of the CLBP\_S is better than that of the CLBC\_S on this database. It indicates that the sign components of the LBP employed by our method are informative and discriminative.

#### 3.3 Experimental Results on the UIUC Database

The UIUC database has 25 texture classes, each having 40 textures imaged under significant viewpoint and scale

<sup>†</sup><http://www.robots.ox.ac.uk/~vgg/research/texclass/index.html>

**Table 2** Classification accuracies (%) on the CURET database.

$N$	46	23	12	6
LTP (R=2, P=16)	91.56	86.15	78.79	68.65
LTP (R=3, P=24)	92.51	87.69	80.89	71.31
Proposed LTP	89.86	84.28	77.33	68.36
CLBP_S (R=2, P=16)	84.95	79.28	72.02	62.91
CLBC_S (R=2, P=16)	78.92	73.99	67.86	59.88
CLBP_S (R=3, P=24)	87.14	81.74	75.26	66.84
CLBC_S (R=3, P=24)	78.19	73.29	67.74	60.51
CLBP_S/M/C (R=2, P=16)	95.65	91.74	85.45	75.91
CLBC_S/M/C (R=2, P=16)	95.48	91.19	84.27	74.00
CLBP_S/M/C (R=3, P=24)	95.76	91.98	85.90	76.60
CLBC_S/M/C (R=3, P=24)	95.05	90.66	83.55	73.13
CLBC_CLBP (R=2, P=16)	96.01	92.01	85.57	75.76
CLBC_CLBP (R=3, P=24)	96.11	92.29	85.95	76.26
Multi-scale CLBP (R=1, 2, 3)	97.00	93.57	87.66	78.26
Multi-scale CLBC (R=1, 2, 3)	96.76	93.01	86.34	76.17
DNS+LBP [10] (R=2, P=16)	95.00			
DNS+LBP [10] (R=3, P=24)	94.52			
VZ_MR8	97.79	95.03	90.48	82.90
VZ_Joint	97.66	94.58	89.40	81.06
<b>Proposed jsLQP</b>	<b>98.43</b>	<b>96.12</b>	<b>91.68</b>	<b>83.86</b>

**Table 3** Classification accuracies (%) on the UIUC database.

$N$	20	15	10	5
LTP (R=2, P=16)	78.75	75.60	70.41	60.06
LTP (R=3, P=24)	82.08	78.91	73.42	62.56
Proposed LTP	84.35	81.68	77.61	69.22
CLBP_S (R=2, P=16)	60.49	56.67	51.19	42.09
CLBC_S (R=2, P=16)	62.80	59.07	53.26	43.14
CLBP_S (R=3, P=24)	64.06	60.16	54.16	44.63
CLBC_S (R=3, P=24)	67.10	63.02	57.32	47.22
CLBP_S/M/C (R=2, P=16)	91.32	89.46	86.34	79.01
CLBC_S/M/C (R=2, P=16)	90.98	89.44	86.71	79.79
CLBP_S/M/C (R=3, P=24)	91.42	89.37	85.93	78.01
CLBC_S/M/C (R=3, P=24)	91.51	89.82	86.93	79.70
CLBC_CLBP (R=2, P=16)	91.46	89.82	86.92	80.07
CLBC_CLBP (R=3, P=24)	92.22	90.38	87.26	79.97
Multi-scale CLBP (R=1, 2, 3)	91.56	89.75	86.62	79.23
Multi-scale CLBC (R=1, 2, 3)	92.30	90.69	87.81	80.80
VZ_MR8	92.94	91.16	88.29	81.12
VZ_Joint	<b>97.83</b>	<b>96.94</b>	<b>95.18</b>	<b>90.17</b>
<b>Proposed jsLQP</b>	<b>94.22</b>	<b>92.46</b>	<b>89.23</b>	<b>81.19</b>

changes. As in [12], [13],  $N$  images per class are randomly chosen for training and the remaining  $40-N$  images are kept for testing. The average accuracies over 100 random splits of the training and test sets are listed in Table 3.

The following observations can be made. Firstly, the proposed jsLQP has about 1.92%, 1.77%, 1.42% and 0.39% performance gains over the multi-scale CLBC (1990 dimensional features) for 20, 15, 10 and 5 training samples, respectively. Secondly, the CLBC performs better than the CLBP when  $R = 2$  and  $R = 3$  on this database. Despite the weakness of the LBP, our method still shows very good performance for scale and viewpoint changes. This is further confirmed on the UIUC database, where the proposed 27-D LTP via the global thresholding outperforms the original LTP [11]. The reasons are two-fold. The first is to employ

the multi-scale gradient responses that account for the scale changes. The second is to utilize the local quantized patterns that are robust to viewpoint variations. Thirdly, the jsLQP outperforms state-of-the-art learning based VZ\_MR8, and it is second to VZ\_Joint. It is noticed that both the VZ\_MR8 and the VZ\_Joint have 2500 dimensions. More importantly, our method is training-free, and involves no costly data-to-cluster assignments. It suggests that exploring the joint statistics of local quantized patterns is effective and efficient for robust texture representation.

#### 4. Conclusion

In this paper, we have proposed a simple yet efficient texture representation by exploring the joint statistics of local quantized (binary and ternary) patterns. The binary and ternary patterns are efficiently extracted based on the raw image and the multi-scale gradient responses. The information in the image and gradient domains is sufficiently explored in a hybrid way. Our method is computationally simple in that the gradient responses are calculated at few scales and the global thresholding is adopted in the proposed ternary quantization. Moreover, our method is training-free, and needs no costly data-to-cluster assignments. Compared with existing LBP based and even learning based algorithms, impressive classification results have been achieved. As future work, we will improve our method by developing local CLBC-like patterns that are robust to significant viewpoint changes.

#### References

- [1] T. Song and H. Li, "Local polar dct features for image description," IEEE Signal Process. Lett., vol.20, no.1, pp.59–62, 2013.
- [2] F. Meng, H. Li, G. Liu, and K.N. Ngan, "Object co-segmentation based on shortest path algorithm and saliency model," IEEE Trans. Multimedia, vol.14, no.5, pp.1429–1441, Oct. 2011.
- [3] 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, July 2002.
- [4] H. Li and K.N. Ngan, "A co-saliency model of image pairs," IEEE Trans. Image Process., vol.20, no.12, pp.3365–3375, Dec. 2011.
- [5] H. Li, G. Liu, and K.N. Ngan, "Guided face cartoon synthesis," IEEE Trans. Multimedia, vol.13, no.6, pp.1230–1239, Dec. 2011.
- [6] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," Int. J. Comput. Vis., vol.62, no.1-2, pp.61–81, April 2005.
- [7] 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.
- [8] W. Zhou, A. Ahrary, and S. Kamata, "Image description with local patterns: An application to face recognition," IEICE Trans. Inf. & Syst., vol.E95-D, no.5, pp.1494–1505, May 2012.
- [9] S. Liao, M. Law, and A. Chung, "Dominant local binary patterns for texture classification," IEEE Trans. Image Process., vol.18, no.5, pp.1107–1118, May 2009.
- [10] F. Khellah, "Texture classification using dominant neighborhood structure," IEEE Trans. Image Process., vol.20, no.11, pp.3270–3279, Nov. 2011.
- [11] 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, June 2010.
- [12] Z. Guo, L. Zhang, and D. Zhang, “A completed modeling of local binary pattern operator for texture classification,” *IEEE Trans. Image Process.*, vol.19, no.6, pp.1657–1663, June 2010.
- [13] Y. Zhao, D.S. Huang, and W. Jia, “Completed local binary count for rotation invariant texture classification,” *IEEE Trans. Image Process.*, vol.21, no.10, pp.4492–4497, Oct. 2012.
- [14] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, “Outex - new framework for empirical evaluation of texture analysis algorithms,” *Proc. ICPR*, pp.701–706, 2002.
- [15] 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.
-