

# MONOGENIC-LBP: A NEW APPROACH FOR ROTATION INVARIANT TEXTURE CLASSIFICATION

Lin Zhang, Lei Zhang<sup>1</sup>, Zhenhua Guo, and David Zhang

Biometrics Research Centre, Dept. of Computing, The Hong Kong Polytechnic University, Hong Kong

## ABSTRACT

Analysis of two-dimensional textures has many potential applications in computer vision. In this paper, we investigate the problem of rotation invariant texture classification, and propose a novel texture feature extractor, namely Monogenic-LBP (M-LBP). M-LBP integrates the traditional Local Binary Pattern (LBP) operator with the other two rotation invariant measures: the local phase and the local surface type computed by the 1<sup>st</sup>-order and 2<sup>nd</sup>-order Riesz transforms, respectively. The classification is based on the image's histogram of M-LBP responses. Extensive experiments conducted on the CURET database demonstrate the overall superiority of M-LBP over the other state-of-the-art methods evaluated.

*Index Terms*— Texture classification, monogenic signal, LBP

## 1. INTRODUCTION

Texture analysis is an active area in the fields of computer vision and pattern recognition, and has many potential applications [1].

Early methods for texture classification focus on the statistical analysis of texture images. The representative methods include the co-occurrence matrix [2] and the filtering based approaches [3]. These methods could achieve good classification results as long as the training and the test samples have identical or similar orientations. Kashyap and Khotanzad were among the first researchers to study the rotation invariant texture classification using a circular autoregressive model [4]. After that, many other models were explored, for example, the multiresolution autoregressive model [5] and the Gaussian Markov random field [6]. In [1], Ojala *et al.* proposed to use the Local Binary Pattern (LBP) histogram for rotation invariant texture classification. Varma and Zisserman [7] proposed a statistical learning based algorithm, namely Maximal Response 8 (MR8), using a group of filter banks, where a rotation invariant texton library is build first from a training set and then an unknown texture image is classified

according to its histogram of texton frequencies. Very recently, under the same framework, Varma and Zisserman [8] proposed a new statistical learning based algorithm, in which, instead of responses of filter banks, compact image patches were used directly to represent local patterns. In addition, scale and affine invariant texture classification is another active research topic and also attracts much attention recently [9].

As one of the effective rotation invariant texture classification methods, LBP is widely used since it is simple yet powerful. However, LBP tends to oversimplify the local image structures. Thus, we want to find some other rotation invariant features to supplement LBP in order to improve the classification accuracy while preserving its simplicity. According to [10], the local phase corresponds to a qualitative measure of a local structure (step, peak, etc) and it is a robust feature with respect to noise and illumination changes. We adopt the monogenic signal theory, which is an isotropic 2-D extension of the 1-D analytic signal, to extract the local image phase information in a rotation invariant way. Besides, we utilize the monogenic curvature tensor to extract the local surface type information, which is another rotation invariant metric. Then, we combine the uniform LBP, the local phase information and the local surface type information together as a novel texton feature, namely Monogenic-LBP (M-LBP). Experiments are conducted to show the efficacy and the efficiency of the M-LBP.

The rest of this paper is organized as follows. Section 2 introduces the monogenic signal and the monogenic curvature tensor. Section 3 describes the M-LBP based texture classification. Section 4 reports the experimental results. Conclusions are presented in Section 5.

## 2. THE MONOGENIC SIGNAL AND THE MONOGENIC CURVATURE TENSOR

For a 2-D image, the intrinsic dimension expresses the number of degrees of freedom necessary to describe the local structure [10]. The monogenic signal is an isotropic 2-D extension of the traditional 1-D analytic signal. It is an effective tool to analyze 1D (intrinsic 1 dimension) 2-D signals (such as lines and edges) in a rotation invariant

---

<sup>1</sup> Corresponding author: [cslzhang@comp.polyu.edu.hk](mailto:cslzhang@comp.polyu.edu.hk).

manner. It is built upon the 1<sup>st</sup>-order Riesz transform. In 2D case, the spatial representation of the Riesz kernel is

$$(R_x(\mathbf{x}), R_y(\mathbf{x})) = \left( \frac{x}{2\pi|\mathbf{x}|^3}, \frac{y}{2\pi|\mathbf{x}|^3} \right), \mathbf{x} = (x, y) \in \mathbb{R}^2 \quad (1)$$

and its transfer function in the Fourier domain is

$$(H_u(\mathbf{u}), H_v(\mathbf{u})) = \left( -i \frac{u}{|\mathbf{u}|}, -i \frac{v}{|\mathbf{u}|} \right), \mathbf{u} = (u, v) \in \mathbb{R}^2 \quad (2)$$

For an image  $f(\mathbf{x})$ , the monogenic signal is defined as the combination of  $f$  and its Riesz transform

$$\begin{aligned} f_M(\mathbf{x}) &= (f(\mathbf{x}), R_x\{f\}(\mathbf{x}), R_y\{f\}(\mathbf{x})) \\ &= (f, R_x * f, R_y * f) \end{aligned} \quad (3)$$

where  $*$  stands for the convolution. For 1D signals (see Fig. 1-a for an example), the local orientation can be calculated as [10, 11]

$$\theta = \arctan \frac{R_y\{f\}}{R_x\{f\}}, \theta \in [0, \pi) \quad (4)$$

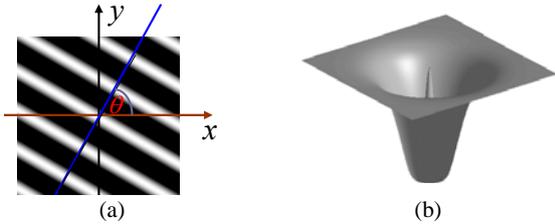
Moreover, for 1D signal  $f(x)$ , it can be proved that [11]

$$\begin{aligned} &\sqrt{R_x^2\{f\}(0,0) + R_y^2\{f\}(0,0)} \\ &= |(h_1 * f_\theta)(0)| = \left| -\frac{1}{\pi} \int_{t \in \mathbb{R}} \frac{f(t \cos \theta, t \sin \theta)}{t} dt \right| \end{aligned} \quad (5)$$

where  $h_1 * f_\theta$  is actually the partial Hilbert transform of  $f$  along its main orientation  $\theta$  and  $h_1 = 1/\pi x$  is the Hilbert transform kernel. Thus, the local phase of the 1D signal  $f(x)$  can be defined analogously to the 1-D analytic signal as

$$\varphi = \text{atan} 2 \left( \sqrt{R_x^2\{f\} + R_y^2\{f\}}, f \right), \varphi \in [0, \pi) \quad (6)$$

It is easy to see that, by such a definition, the local phase is calculated in a rotation invariant way.



**Fig. 1:** (a) Illustration for the local orientation of an ideal 1D signal; (b) an example of the LOP filter in the frequency domain.

The monogenic curvature tensor [11], based on higher order Riesz transforms can be used to analyze some specific 2D signals. For our needs, we only make use of the even part of the monogenic curvature tensor, which is defined as [11]

$$T_e = \begin{bmatrix} R_x\{R_x\{f\}\}, R_x\{R_y\{f\}\} \\ R_x\{R_y\{f\}\}, R_y\{R_y\{f\}\} \end{bmatrix} \quad (7)$$

It can be seen that  $T_e$  is composed of 2<sup>nd</sup>-order Riesz transforms. Its determinant is

$$\det(T_e) = (R_x\{R_x\{f\}\})(R_y\{R_y\{f\}\}) - (R_x\{R_y\{f\}\})^2 \quad (8)$$

If we embed the image  $f(x, y)$  into the Monge patch  $s(x, y) = \{x, y, f(x, y)\}$ , where  $s(x, y)$  is a 2-D surface embedded in a 3-D ambient Euclidean space, then the Gaussian curvature  $K$  of  $s$  can partially determine the surface type of  $s$  [12]. Surfaces with  $K > 0$  correspond to elliptic patches while  $K < 0$  indicates a hyperbolic patch. According to [11], the sign of  $\det(T_e)$  is the same as the sign of  $K$ . Thus, the sign of  $\det(T_e)$  can reflect the shape type of  $s$  embedded from the image  $f$  and it is a rotation invariant metric. In this paper we will use the sign of  $\det(T_e)$  to represent the local surface type information.

It can be seen that the monogenic signal and the related results, are build upon some assumptions on the signal models. In this paper, we simply suppose that these assumptions can be satisfied locally and therefore the results stated above are applied directly to the texture images.

### 3. MONOGENIC-LBP (M-LBP)

#### 3.1. M-LBP feature extraction

In this section, we present a new texton feature, namely Monogenic-LBP (M-LBP). The idea is that we want to combine the local phase information, the local surface type information, and the traditional LBP to improve the classification accuracy. Specifically, for LBP, we use the “uniform” form, defined as [1]

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^P s(g_p - g_c) & \text{if } U(LBP_{P,R}) \leq 2 \\ P+1 & \text{otherwise} \end{cases} \quad (9)$$

where

$$\begin{aligned} U(LBP_{P,R}) &= |s(g_{P-1} - g_c) - s(g_0 - g_c)| \\ &\quad + \sum_{p=1}^{P-1} |s(g_p - g_c) - s(g_{p-1} - g_c)| \end{aligned} \quad (10)$$

Superscript “*riu2*” means the use of rotation invariant “uniform” patterns that have  $U$  value of at most 2;  $s$  is the sign function;  $g_c$  corresponds to the gray value of the center pixel of the local neighborhood and  $g_p (p = 0, \dots, P-1)$  correspond to the gray values of  $P$  equally spaced pixels on a circle of radius  $R$ .

Instead of using the value of the local phase  $\varphi$  directly, we quantize it into  $M$  discrete levels to get the “phase code” as

$$\varphi_c = \lceil \varphi / (\pi / M) \rceil \quad (11)$$

where  $\lceil x \rceil$  is the operator to return the smallest integer not smaller than  $x$ . In this paper, we experimentally set  $M$  as 5. Thus, it is easy to see that  $\varphi_c$  is an integer within 1 ~ 5.

Similarly, we do not use the exact value of  $\det(T_e)$  either; instead, we binarize it as

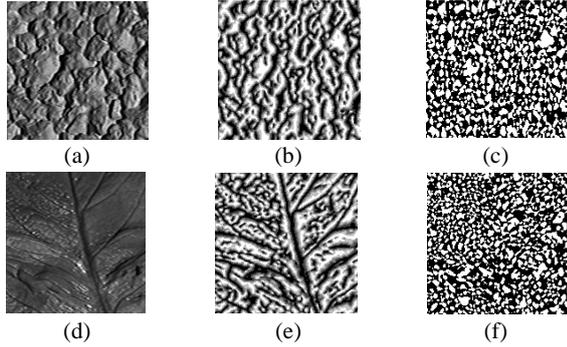
$$S_C = \begin{cases} 0, & \det(T_e) \leq 0 \\ 1, & \text{else} \end{cases} \quad (12)$$

By combining  $\varphi_c$ ,  $S_C$ , and  $LBP_{P,R}^{riu2}$ , we can obtain a new 3-D texton feature vector  $(\varphi_c, S_C, LBP_{P,R}^{riu2})$  and we name it as Monogenic-LBP (M-LBP).

In practice the signals are of finite length. Therefore, we need to perform band-pass filtering to the image before applying the Riesz transforms to it. With respect to the band-pass filter, we use the Laplacian of Poisson (LOP) filter proposed in [11], whose transfer function in the Fourier domain is

$$\mathcal{F}\{LOP\}(\mathbf{u}) = -4\pi^2 |\mathbf{u}|^2 \exp(-2\pi|\mathbf{u}|\lambda), \mathbf{u} \in \mathbb{R}^2 \quad (13)$$

where  $\lambda$  is used to control the center frequency. An example of the LOP filter with  $\lambda = 3.5$  in the Fourier domain is shown in Fig. 1-b. In Fig. 2, two texture images and their associated  $\varphi_C$  and  $S_C$  extracted with  $\lambda = 3.5$  are shown.



**Fig. 2:** (a) and (d) are two texture images; (b) is the  $\varphi_C$  extracted from (a); (c) is the  $S_C$  extracted from (a); (e) is the  $\varphi_C$  extracted from (d); (f) is the  $S_C$  extracted from (d).

In real applications, a multiresolution analysis can usually lead to better results. With our scheme, a multiresolution analysis can be accomplished by combining the information provided by multiple operators of varying  $(P, R, \lambda)$ . Specifically, we use 3 M-LBP operators, denoted by  $M-LBP_i$  ( $i=1, 2, 3$ ), in this paper and the parameters are experimentally set as  $(8, 1, 3.5)$ ,  $(16, 3, 7)$  and  $(24, 5, 14)$ .

### 3.2. Classification method

A 3-D normalized histogram  $h_i$  ( $i=1, 2, 3$ ) can be constructed for each image by counting the frequencies of textons  $M-LBP_i$  ( $i=1, 2, 3$ ) over the whole image. Then, we concatenate  $h_1$ ,  $h_2$ , and  $h_3$  together to form a larger histogram  $h$ , and regard it as the descriptor of the image. It can be seen that the bins of  $h$  is of size 540 ( $5 \times 2 \times 10 + 5 \times 2 \times 18 + 5 \times 2 \times 26$ ).

We use the  $\chi^2$  statistic to measure the dissimilarity of sample and model histograms. Thus, a test sample  $T$  will be assigned to the class of model  $L$  that minimizes

$$D(T, L) = \sum_{n=1}^N (T_n - L_n)^2 / (T_n + L_n) \quad (14)$$

where  $N$  is the number of bins, and  $T_n$  and  $L_n$  are the values of the sample and model histogram at the  $n^{\text{th}}$  bin, respectively. With respect to the classifier, we use the nearest neighborhood classifier.

## 4. EXPERIMENTS AND DISCUSSIONS

The code of the proposed method can be downloaded at <http://www4.comp.polyu.edu.hk/~cslzhang/code.htm>. We conducted experiments on a modified CURET database [13] provided at [14]. It contains 61 textures and each texture has 92 images obtained under different viewpoints and illumination directions. Before feature extraction, all the images are converted to gray scale and are normalized to have a mean of 0 and a standard deviation of 1.

We compared the performance of the proposed M-LBP method with the other three state-of-the-art rotation invariant texture classification methods, MR8 [7], Joint [8], and the original LBP [1]. In MR8, 40 textons were clustered from each of the 61 texture classes using the training samples and thus the texton dictionary was of the size 2440 ( $61 \times 40$ ). In Joint, the size of the image patch was selected as  $7 \times 7$ , and also 40 textons were clustered from each of the 61 texture classes. In LBP, we combined the information extracted by three operators  $LBP_{8,1}^{riu2}$ ,  $LBP_{16,3}^{riu2}$ , and  $LBP_{24,5}^{riu2}$  together. The classification accuracies were obtained with  $\chi^2$  distance and the nearest neighborhood classifier.

We performed experiments on four different settings to simulate four situations:

1. T46A: The training set for each class was selected by taking one from every two adjacent images. Hence, there were 2,806 ( $61 \times 46$ ) models and 2,806 test samples. This setting was used to simulate the situation of large and comprehensive training set.

2. T23A: The training set for each class was selected by taking one from every four adjacent images. Hence, there were 1,403 ( $61 \times 23$ ) models and 4,209 ( $61 \times 69$ ) test samples. This setting was used to simulate the situation of small but comprehensive training set.

3. T46F: The training set for each class was selected as the first 46 images. Hence, there were 2,806 models and 2,806 test samples. This setting was used to simulate the situation of large but less comprehensive training set.

4. T23F: The training set for each class was selected as the first 23 images. Hence, there were 1,403 models and 4,209 test samples. This setting was used to simulate the situation of small and less comprehensive training set.

The classification accuracies and the feature sizes for the four methods are listed in Table 1. In addition, we also care about the classification speeds. At the classification stage, the histogram of the test image will be built at first and then it will be matched to all the models generated from the training samples. In Table 2, we list the time for one test histogram construction and for one matching at the classification stage by each method. All the algorithms were implemented with Matlab 7.4 except that a C++ implemented  $kd$ -tree (encapsulated in a MEX function) was used in MR8 and Joint to accelerate the labeling process. Experiments were performed on a Dell Inspiron 530s PC with Intel 6550 processor and 2GB RAM.

**Table 1.** Classification accuracies (%) and feature sizes

	Histogram Bins	T46A	T23A	T46F	T23F
LBP	54	95.47	93.09	85.64	78.50
MR8	2440	97.65	96.15	88.70	77.83
Joint	2440	97.36	95.37	87.13	78.55
<b>M-LBP</b>	<b>540</b>	<b>97.86</b>	<b>96.72</b>	<b>89.67</b>	<b>81.21</b>

**Table 2.** Time consumption (msec) at the classification stage

	Histogram construction	One matching
LBP	87	0.022
MR8	4960	0.089
Joint	13173	0.089
M-LBP	221	0.035

With Table 1 and Table 2, we evaluate the four schemes from three aspects: the classification accuracy, the feature size, and the runtime classification speed. First, with respect to the classification accuracy, M-LBP outperforms all the other 3 approaches under all the 4 experimental settings. Especially, it performs significantly better than the other ones in “T23F”. By contrast, although MR8 and Joint achieves similar performance with M-LBP in “T46A”, they perform much worse than M-LBP in “T23F”. This is because both MR8 and Joint require a training stage, which depends much on the training samples; when the training set is small and not comprehensive (as in “T23F”), the classification accuracy will drop largely. Thus, this indicates that the proposed M-LBP is more suitable for real applications where training samples are limited and not comprehensive. Second, M-LBP requires a moderate feature size, bigger than LBP but much smaller than MR8 and Joint. Although the feature size of M-LBP is bigger than LBP, considering the gain in the classification accuracy, it is deserved. Third, the four schemes have very different classification speeds. LBP and M-LBP work much faster than MR8 and Joint. In MR8 and Joint, in order to build the histogram of the test image, every pixel on the test image need to be labeled to one item in the texton dictionary, which is quite time consuming. Such a process is not required in LBP and M-LBP. Besides, an extra training period is needed in MR8 and Joint to build the texton dictionary, which is also not required in LBP and M-LBP. Thus, in general, the proposed M-LBP has the merits of high classification accuracy, small feature size and fast classification speed.

## 5. CONCLUSION

We present a novel training free rotation invariant texture classification method, namely M-LBP. It combines two rotation invariant measures, the local phase and the local surface type extracted by the 1<sup>st</sup>- and 2<sup>nd</sup>- order Riesz transforms, with the traditional uniform LBP operator. Experimental results validate that M-LBP can achieve

higher classification accuracy than the other methods evaluated, especially in the cases when the training set is small and not comprehensive. Moreover, compared with the two state-of-the-art training based methods, MR8 and Joint, M-LBP has the advantage of smaller feature size and faster classification speed, which makes it a more suitable candidate in real applications.

## ACKNOWLEDGEMENT

This research is supported by the Hong Kong RGC General Research Fund (PolyU 5330/07E) and The Hong Kong PolyU Internal Fund (A-SA08).

## 6. REFERENCES

- [1] T. Ojala, M. Pietikäinen, and T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”, *IEEE Trans. PAMI*, vol. 24, pp. 971-987, 2002.
- [2] R.M. Haralik, K. Shanmugam, and I. Dinstein, “Texture features for image classification”, *IEEE Trans. SMC*, vol. 3, pp. 610-621, 1973.
- [3] T. Randen and J.H. Husy, “Filtering for texture classification: a comparative study”, *IEEE Trans. PAMI*, vol. 21, pp. 291-310, 1999.
- [4] R.L. Kashyap and A. Khotanzed, “A model-based method for rotation invariant texture classification”, *IEEE Trans. PAMI*, vol. 8, pp. 472-481, 1986.
- [5] J. Mao and A.K Jain, “Texture classification and segmentation using multiresolution simultaneous autoregressive models”, *Pattern Recognition*, vol. 25, pp. 173-188, 1992.
- [6] H. Deng and D.A. Clausi, “Gaussian MRF rotation-invariant features for image classification”, *IEEE Trans. PAMI*, vol. 26, pp. 951-955, 2004.
- [7] M. Varma and A. Zisserman, “A statistical approach to texture classification from single images”, *IJCV*, vol. 62, pp. 61-81, 2005.
- [8] M. Varma and A. Zisserman, “A statistical approach to material classification using image patch exemplars”, *IEEE Trans. PAMI*, vol. 31, pp. 2032-2047, 2009.
- [9] Y. Xu, H. Ji, and C. Fermüller, “Viewpoint invariant texture description using fractal analysis”, *IJCV*, vol. 83, pp. 85-100, 2009.
- [10] M. Felsberg and G. Sommer, “The monogenic signal”, *IEEE Trans. SP*, vol. 49, pp. 3136-3144, 2001.
- [11] O. Fleischmann, “2D signal analysis by generalized Hilbert transforms”, Thesis, University of Kiel, 2008.
- [12] M.P. do Carmo, *Differential geometry of curves and surfaces*, Prentice-Hall, 1976.
- [13] K.J. Dana, B.V. Ginneken, S.K. Nayar, and J.J. Koenderink, “Reflectance and texture of real world surfaces”, *ACM Trans. Graphics*, vol. 18, pp. 1-34, 1999.
- [14] [http://www.robots.ox.ac.uk/~vgg/research/texclass/data/curetc\\_ol.zip](http://www.robots.ox.ac.uk/~vgg/research/texclass/data/curetc_ol.zip), 2008.