LETTER RBM-LBP: Joint Distribution of Multiple Local Binary Patterns for Texture Classification

Chao LIANG^{†,††,†††a)}, Wenming YANG^{†,††,†††b)}, Nonmembers, Fei ZHOU^{†,††,†††c)}, Member, and Qingmin LIAO^{†,††,†††d)}, Nonmember

SUMMARY In this letter, we propose a novel framework to estimate the joint distribution of multiple Local Binary Patterns (LBPs). Multiple LBPs extracted from the same central pixel are first encoded using handcrafted encoding schemes to achieve rotation invariance, and the outputs are further encoded through a pre-trained Restricted Boltzmann Machine (RBM) to reduce the dimension of features. RBM has been successfully used as binary feature detectors and the binary-valued units of RBM seamlessly adapt to LBP. The proposed feature is called RBM-LBP. Experiments on the CUReT and Outex databases show that RBM-LBP is superior to conventional handcrafted encodings and more powerful in estimating the joint distribution of multiple LBPs.

key words: LBP, OLBP, RBM, texture classification

1. Introduction

Texture classification has received considerable attention during the past few decades because it is fundamental for many computer vision tasks, such as image retrieval, image segmentation, face recognition, etc. The appearance of textures changes dramatically as the illumination changes or the image rotates. Local Binary Pattern (LBP) [1], one of the most famous texture features, is robust to illumination changes and is rotation invariant.

For each pixel of an image, LBP compares it with its neighbors sampled on a circle and uses the signs of the differences to compose a binary pattern. Each binary pattern is encoded independently to achieve rotation invariance as well as to reduce the feature dimension. All the encoded patterns of an image compose a histogram as the texture descriptor of the image.

Binary patterns can be encoded through handcrafted encodings or learning-based encodings. Among handcrafted encodings, Rotation Invariant (ri) and Rotation Invariant Uniform Patterns (riu2) encodings [1] are widely used. ri

[†]The authors are with the Department of Electronic Engineering, Graduate School at Shenzhen, Tsinghua University, Beijing, China..

^{††}The authors are with the Shenzhen Key Laboratory of Information Science and Technology, Guangdong, China..

^{†††}The authors are with the Visual Information Processing Lab, Tsinghua-PolyU Biometrics Joint Lab, Guangdong, China..

a) E-mail: chaoliangthu@163.com

b) E-mail: yangelwm@163.com

c) E-mail: flying.zhou@163.com

d) E-mail: liaoqm@tsinghua.edu.cn (Corresponding author) DOI: 10.1587/transinf.2016EDL8072 encoding achieves rotation invariance through circular shift. If one binary pattern equals another binary pattern through bit-wise circular shift, the two patterns are encoded to the same bin. The riu2 encoding further compresses ri encoded patterns by means of merging all nonuniform patterns to the same bin. Subsequently, Ahonen et al. proposed LBP Histogram Fourier features (LBP-HF) [2]; it performs discrete Fourier transform to histograms of uniform patterns to reserve more information than riu2 does. Oriented LBP (OLBP) [3] extends riu2 to Reflection and Rotation Invariant Uniform Patterns (rriu2) to accommodate elliptic neighborhood. LBP-HF, riu2, and rriu2 ignore the discriminative information of different nonuniform patterns based on the fact that nonuniform patterns contribute only about 10% of binary patterns when the radius of sampling is one. But the ratio of nonuniform patterns increases as the radius increases. There are more than 60% nonuniform patterns when the radius is five, which makes it inappropriate to merge all nonuniform patterns to the same bin.

On the contrary, learning-based encodings treat uniform patterns and nonuniform patterns equally. For instance, Dominant LBP (DLBP) [4] counts the most frequently occurred LBP^{ri} in an image and uses the sorted frequencies are the descriptor. However, the dominant patterns of two images may be of different types. Moreover, as the number of neighbor pixels increase, the dimension of DLBP also increases rapidly.

To embed binary patterns into a low dimensional space without loss of structural information, this letter proposes a two-step encoding framework that combines handcrafted encodings with learning-based encodings. LBPs are first encoded using handcrafted encodings to explore the local structures of textures, and the outputs are fed into a pretrained Restricted Boltzmann Machine to obtain a compact binary code.

2. An Overview of Restricted Boltzmann Machine

The Restricted Boltzmann Machine (RBM) is a two-layer, bipartite, undirected network, where the visible units (v) and hidden units (h) are binary-valued (0 or 1). The probability of a joint configuration (v, h) is defined by its energy function as follows:

$$P(\boldsymbol{v}, \boldsymbol{h}) = \frac{1}{Z} \exp(-E(\boldsymbol{v}, \boldsymbol{h}))$$
(1)

Manuscript received April 1, 2016.

Manuscript revised July 17, 2016.

Manuscript publicized August 19, 2016.

where *Z* is the partition function, and the energy function [5] is given by

$$E(\boldsymbol{v},\boldsymbol{h}) = -\boldsymbol{p}^T \boldsymbol{v} - \boldsymbol{q}^T \boldsymbol{h} - \boldsymbol{v}^T \boldsymbol{W} \boldsymbol{h}$$
(2)

where W is the weight matrix between v and h, and p and q are the biases of v and h, respectively.

Given an instance of v, the corresponding states of h can be obtained through Gibbs sampling, and vice versa. The hidden units can be viewed as binary feature detectors of the visible units. The readers are referred to [6] for details.

3. The Proposed Encoding Scheme

Sampling neighbor pixels on multiple ellipses has been proved to be more discriminative than that on a circle. Therefore, as shown in the bottom row of Fig. 1, for each central pixel, we follow OLBP[3] that samples neighbor pixels on multiple ellipses with different orientations. And then the binary patterns are encoded jointly. A major consideration when designing an encoding scheme is how to make the output invariant to rotation. Moreover, for computational efficiency, low dimensional feature is preferred. The two objectives can be achieved independently.

To achieve rotation invariance, handcrafted encodings are preferred. As illustrated in Fig. 1, we treat each elliptic neighborhood as a block, and apply the Reflection and Rotation invariant (rri) encoding to each block. The rri encoding makes each binary pattern partially invariant to rotation and reduce the feature dimension of each block from 256 to 84 when there are 8 neighbor pixels. It is worth mentioning



Fig. 1 Flowchart of RBM-LBP, meanings of each step is shown in Alg. 1. For simplicity, we use 4 orientations, each with 4 neighbor pixels here, and the within-block encoding is rri. The binary patterns (1111, 0001, 1000, 0100) are encoded to 010.

that the encoded pattern is not converted to an integer value as Ojala *et al.* [1] does. The reason is that all the bits of a binary pattern shall hold the same weight when they are fed into the RBM in the last step.

After within-block encoding, holistic rotation invariance is achieved through block-wise circular shift (BCS), which is given by:

BCS(
$$\boldsymbol{b}, N, P$$
) = min {ROR(\boldsymbol{b}, nP, NP) | $n = 0, 1, \dots, N-1$ }
(3)

where \boldsymbol{b} is the cascaded binary pattern, N is the number of elliptic neighborhoods, P is the number of neighbor pixels for each elliptic neighborhood, and ROR(\boldsymbol{b}, nP, NP) performs a bit-wise circular right shift on the NP-bit number \boldsymbol{b} for nP times.

Although the feature dimension has been reduced through within-block encoding, further feature reduction is necessary. Take N = 8 and P = 8 for example, the dimension of OLBP is 2⁶⁴. Even after within-block encoding using rri, the dimension is still 84⁸, which is intractable. In the former work of OLBP [3], histograms extracted from different elliptic neighborhood are merged into one histogram. The merged histogram is not an estimation of the joint distribution, and leads to the loss of structural information.

We propose to train an RBM model to reduce the dimension of multiple LBPs. LBP is defined on binary states and each bit of the binary pattern contains position information, henceforth, classical feature reduction methods, such as Principal Components Analysis (PCA) and Locally Linear Embedding (LLE), are not proper for LBP. On the contrary, RBM is defined on binary states, and the hidden units of an RBM can be viewed as binary feature detectors. Moreover, RBM is unsupervised feature learner. The objective of training the RBM is to maximize the log probabilities over all training samples (the outputs of BCS). This leads to the learning rule using stochastic gradient ascent as:

$$\Delta W_{ij} = \epsilon \sum_{h} v_i h_j P(h|v) - \epsilon \sum_{v,h} v_i h_j P(v,h)$$
(4)

where ϵ is the learning rate. The first term on the right side is driven by the training data (binary patterns). The second term on the right side is driven by the model itself, and is

Algorithm	1	RBM-LBP feature extraction.	

Input:

Oriented binary patterns of a pixel

- Output:
- RBM encoded binary patterns 1: Within-block encoding

For each orientation, the binary pattern is encoded using ri or rri.

- Block-wise circular shift The cascaded binary pattern is circular shifted block-wisely to its minimal value.
- 3: Encoding using RBM

The circular shifted binary pattern is fed into a pre-trained RBM and obtain *S* corresponding binary states of hidden units through Gibbs sampling.

intractable to compute.

In our implementation, the contrastive divergence (CD method [7] is used to approximate the gradient. The CD method apply "one-step" reconstructions of v and h through Gibbs sampling to approximate the second term in Eq. (4).

After training, the marginal distribution of the visible units is nearly the same as the distribution of the training samples. In addition, a side effect of using CD is that v and **h** can reconstruct each other with high probability. In other words, the distribution of the visible units and the distribution of the hidden units are coupled, and one can be represented by the other. Therefore, we propose to use the states of the hidden units to represent those of the visible units (the cascaded binary patterns). If the number of hidden units (N_h) is smaller than the number of visible units (N_p) , the feature dimension is reduced. It is worth noting that the states of the hidden units still form a binary pattern, which means the dimension of the output is 2^{N_h} rather than N_h . For example, we train an RBM on Outex TC10 database [8] using the configuration in Fig. 1. The most frequently occurred binary pattern after BCS is (1100, 1100, 1100, 1100), and the corresponding RBM output is (1.000, 1.000, 0.997). After Gibbs sampling, the most probable state is 111, but 110 is possible, too. Another frequently occurred binary pattern is (0001, 0001, 0001, 0001), and it is assigned to state 000 with probability near to one. The Hamming distance between the two binary patterns is very large (12 out of 16 bits are different), henceforth, the encodings 111 and 000 make sense. It is worth noting that the within-block encoding and BCS are necessary because the RBM knows nothing about the local structures of an image.

We call the proposed feature RBM-LBP. The steps of extracting RBM-LBP are summarized in Alg. 1. We follow LBP that the histogram of all the RBM-LBPs of an image is computed as the image descriptor.

4. Experiments and Discussion

To evaluate the effectiveness of the proposed feature, we test LBP, LBP-HF, DLBP, OLBP and RBM-LBP on CUReT database [9] and Outex database [8]. Both CUReT and Outex need the descriptor to be rotation invariant.

4.1 Settings of Experiments

The dimensions of each feature we compare are listed in Table 1. The dimension of DLBP varies with database and

Table 1Comparison of feature dimensions.

descriptor	number of neighbors	encoding	feature dimension
LBP	8	riu2	10
LBP-HF	8	-	38
DLBP	16	ri	≈ 70
DLBP	24	ri	≈ 600
OLBP	64	rriu2	21
RBM-LBP	64	-	2^{N_h}

we give an approximation here. In all settings, the nearest neighbor classifier with chi-square distance as the metric is utilized.

We trained the RBM model using CD method for 5 epoches in mini-batches of 100, with a learning rate of 0.1, a momentum of 0.95, and a weight decay rate of 10^{-5} . The weight matrix is initialized using random numbers sampled from a zero-mean Gaussian distribution with a standard deviation of 0.01, and the initial value of the biases are set to **0**. And we extract 500 binary patterns randomly per training image as the training samples.

For simplicity, we abbreviate the names of the features as follows: LBP_{*P,r*} represents LBP with *P* neighbor pixels sampled on a circle of radius *r*. The marks for LBP-HF and DLBP are the same as LBP's. RBM-LBP_{*P,a,b,N,N_h*} represents RBM encoded patterns with *N* ellipses, each with *P* sampling pixels, *a* and *b* are the lengths of the semi-major axis and the semi-minor axis of the ellipses, and *N_h* is the number of hidden units. If rri within-block encoding is used, the angle between two adjacent ellipses is set to π/N . If we bypass the within-block encoding step (with "origin" as the superscript), the angle is set to $2\pi/N$.

4.2 Experiment #1

The CUReT database contains 61 texture classes, each with 92 images. These images are acquired at different illumination directions and the the viewing angles are different.

As we can see in Table 2, the recognition rate of RBM-LBP_{8,1,1,1,3} is only 65.66%, which is 14% lower than LBP^{riu2}'s. The result show that encoding using RBM is not as powerful as riu2. The result of DLBP is lower than that of the original paper because we use the nearest neighbor classifier rather than a support vector machine.

As the number of neighbor pixels increases, the recognition rates of RBM-LBP_{8,5,1,8,8} and RBM-LBP_{8,5,1,8,12} are higher than all handcrafted encoding schemes. In most cases, the recognition rate of RBM-LBP_{8,5,1,8,12} is higher than RBM-LBP_{8,5,1,8,8}'s, which means that more hidden

descriptor	number of training samples			
accomptor	46	23	12	6
LBP ₈₁	79.94	73.99	66.83	57.88
LBP ^{riu2}	73.00	67.52	61.34	53.54
LBP-HF _{8,1}	89.74	84.24	77.01	67.20
LBP-HF _{8,5}	89.90	84.60	77.64	68.26
DLBP ^{ri} _{16.2}	80.43	74.47	67.48	58.76
DLBP ^{ri} _{24.3}	83.92	78.42	71.80	63.40
OLBP ^{rriu2} , 8,5,1,8	91.28	85.80	78.64	69.03
RBM-LBP ^{ri} 8.1.1.1.3	65.66	60.10	54.20	47.38
RBM-LBP ^{ri} 8.5.5.1.3	39.12	37.42	47.48	32.23
RBM-LBP ^{origin} 8.5.1.8.8	95.23	91.45	85.93	77.78
RBM-LBP ^{rri} 8,5,1,8,8	92.36	87.80	81.43	72.48
RBM-LBP ^{origin} 8.5.1.8.12	95.66	92.15	86.94	78.97
RBM-LBP ^{rri} 8,5,1,8,12	92.81	88.66	82.94	74.57

Table 3Classification rate (%) on TC10 and TC12.

	TC10	TC12 t184	TC12 horizon
LBP ^{riu2} ₈₁	85.05	66.13	64.05
LBP ^{riu2}	71.20	63.96	63.47
LBP-HF _{8,1}	83.23	76.53	78.31
LBP-HF _{8,5}	75.44	71.69	71.74
DLBP ^{ri} _{16.2}	84.61	71.94	65.81
DLBP ^{ri} _{24.3}	89.45	76.53	73.82
OLBP ^{rriu2} , 8,5,1,8	98.18	91.18	85.95
RBM-LBP ^{ri} 8,1,1,1,3	74.90	56.90	54.49
RBM-LBP ^{ri} _{8,5,5,1,3}	50.57	42.89	43.36
RBM-LBP ^{origin} 8.5.1.8.8	97.50	94.72	94.17
RBM-LBP ^{rri} 8,5,1,8,8	98.85	93.52	92.66
RBM-LBP ^{origin} 8,5,1,8,12	92.37	92.89	94.33
RBM-LBP ^{rri} 8,5,1,8,12	99.35	96.32	95.69

units are useful to achieve higher recognition rate. To our surprise, RBM-LBP $_{8,5,1,8,8}^{\text{origin}}$ is superior to RBM-LBP $_{8,5,1,8,8}^{\text{rri}}$ on CUReT. The reasons is that even if we bypass the withinblock encoding, the BCS operation still make the binary pattern rotation invariant.

4.3 Experiment #2

The Outex database has 24 classes of textures acquired under three illuminants and nine rotation angles. The Outex database has three test suites: TC10, TC12 "horizon", and TC12 "t184". For TC10, all the samples are taken under illuminant "inca". The training samples of TC10 are taken at angle 0°, while the rest samples are used for testing. For TC12 "horizon" and TC12 "t184", the training samples are the same as those of TC10, while the testing samples are acquired under illuminants "horizon" and "t184", respectively.

Table 3 lists the experimental results on Outex database. The results are similar to experiments on CUReT except that RBM-LBP^{rri}_{8,5,1,8,8} is superior to RBM-LBP^{origin}_{8,5,1,8,8}. Because the training images and testing images in Outex are acquired in different rotation angles, more rotation invariance is needed than CUReT database. In this situation, the rri within-block encoding and BCS operation jointly make RBM-LBP robust to rotation.

5. Conclusion

In this letter, we propose a two-step encoding scheme called

RBM-LBP to encode multiple binary patterns jointly. RBM-LBP combines the advantages of handcrafted encodings with those of learning-based encodings. Experiments on CUReT database and Outex database show that RBM-LBP is powerful for mapping the joint distribution of binary patterns into a low dimensional space and achieves higher recognition rate than treating each binary pattern independently.

Acknowledgements

This work was supported by the National Natural Science Foundation of China Grant No. 61471216, and the Special Foundation for the Development of Strategic Emerging Industries of Shenzhen under Grant No. ZDSYS20140509172959974 and JCYJ20150331151358138.

References

- 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.
- [2] T. Ahonen, J. Matas, C. He, and M. Pietikäinen, "Rotation invariant image description with local binary pattern histogram Fourier features," Image Analysis, ed. A.B. Salberg, J. Hardeberg, and R. Jenssen, Lecture Notes in Computer Science, vol.5575, pp.61–70, Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.
- [3] C. Liang, W. Yang, F. Zhou, and Q. Liao, "Reflection and rotation invariant uniform patterns for texture classification," IEICE Trans. Inf. & Syst., vol.E99-D, no.5, pp.1400–1403, May 2016.
- [4] S. Liao, M.W.K. Law, and A.C.S. Chung, "Dominant local binary patterns for texture classification," IEEE Trans. Image Process., vol.18, no.5, pp.1107–1118, May 2009.
- [5] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities," Proc. National Academy of Sciences, vol.79, no.8, pp.2554–2558, April 1982.
- [6] G.E. Hinton, "A practical guide to training restricted Boltzmann machines," Neural Networks: Tricks of the Trade, ed. G. Montavon, G. Orr, and K.R. Müller, Lecture Notes in Computer Science, vol.7700, pp.599–619, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.
- [7] G.E. Hinton, "Training products of experts by minimizing contrastive divergence," Neural Comput., vol.14, no.8, pp.1771–1800, Aug. 2002.
- [8] T. Ojala, T. Maenpaa, M. Pietikainen, J. Viertola, J. Kyllonen, and S. Huovinen, "Outex — New framework for empirical evaluation of texture analysis algorithms," Proc. 16th International Conference on Pattern Recognition, 2002, vol.1, pp.701–706, IEEE, 2002.
- [9] 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.