



**HAL**  
open science

## A scattering transform combination with local binary pattern for texture classification

Vu-Lam Nguyen, Ngoc-Son Vu, Philippe-Henri Gosselin

► **To cite this version:**

Vu-Lam Nguyen, Ngoc-Son Vu, Philippe-Henri Gosselin. A scattering transform combination with local binary pattern for texture classification. International Workshop on Content-based Multimedia Indexing , Jun 2016, Bucharest, Romania. 10.1109/CBML.2016.7500238 . hal-01346956

**HAL Id: hal-01346956**

**<https://hal.science/hal-01346956>**

Submitted on 18 Aug 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A Scattering Transform combination with Local Binary Pattern for Texture Classification

Vu-Lam Nguyen

Ngoc-Son Vu

Philippe-Henri Gosselin

*ETIS - ENSEA/ Université de Cergy-Pontoise/ CNRS UMR 8051  
95000-Cergy, France  
Email: lam.nguyen,son.vu,philippe-henri.gosselin{ @ensea.fr }*

**Abstract**—In this paper, we propose a combined feature approach which takes full advantages of local structure information and the more global one for improving texture image classification results. In this way, Local Binary Pattern is used for extracting local features, whilst the Scattering Transform feature plays the role of a global descriptor. Intensive experiments conducted on many texture benchmarks such as ALOT, CURET, KTH-TIPS2-a, KTH-TIPS2b, and OUTEX show that the combined method outweigh each one which stands alone in term of classification accuracy. Also, our method outperforms many others, whilst it is comparable to state of the art on the experimented datasets.

## 1. Introduction

Texture is the fundamental appearance element of materials or objects which has attracted a great attention of researchers in computer vision. This is due to the fact that texture classification is an important problem in computer vision and image processing, playing a significant role in many applications such as medical image analysis, object recognition, content-based image retrieval.

Texture images can be discriminated based on filter banks through the statistical distributions of their responses or within small scale neighborhoods using the pixel intensities. While the former attracts a more global structure information of images, the latter demonstrates that a good discrimination is able to be achieved through exploiting the distributions of pixel neighborhoods.

Local Binary Pattern (LBP) [1] has drawn considerable attention since its proposal, and has already been used in many other applications, such as image retrieval, face image analysis [2], and so on. However, the conventional version has some limitations, such as small spatial support region, loss of global textural information, also sensitive to noise. Many LBP variants was proposed to overcome those, Completed LBP (CLBP) [3] is one of them. Even though, these variants struggle to get high performance on image dataset with variant in scale, translation, and deformation. On the other hand, Scattering transform introduced

in Scattering NetWork (ScatNet) by Mallat et al. [4], which applies wavelet transform in a deep convolution network, copes well with those characteristics of data. However these scattering transform features do not capture well the small local structure information.

In this paper, we propose to make use the strength of both LBP and Scattering transform in a "Hybrid" descriptor for texture classification, CLPB [3] is used instead of the original version because of its higher performance.

The rest of paper is organized as follows. Section 2 is the proposed approach. Section 3 presents experimental results, and conclusions are given in Section 4.

## 2. Related Work

In this section, LBP, CLBP, scattering transform, and PCA classifier are reviewed.

### 2.1. Brief view of the LBP and CLBP

The LBP method, first proposed by Ojala et al. [1], which encodes the pixel-wise information in textured images. LBP encoding is:

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p, \quad s(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (1)$$

Where  $g_c$  represents the grey value of the center pixel whereas  $g_p$  ( $p = 0, \dots, P - 1$ ) denotes the grey value of the neighbor pixel on a circle of radius  $R$ , and  $P$  is the total number of the neighbors. A given texture image is then represented by histogram of LBP codes. Ojala et al. introduced rotation invariant complement called uniform patterns  $LBP^{riu2}$  which have less than two "one-to-zero or vice versa" transitions.

Guo et al. [3] suggested a variant called CLBP by which the image local differences are decomposed into two complementary components, the signs ( $s_p$ ) and the magnitudes ( $m_p$ ):

$$s_p = s(g_p - g_c), \quad m_p = |g_p - g_c| \quad (2)$$

where  $g_p$ ,  $g_c$  and  $s(x)$  are defined as in (1). Two operators called CLBP-Sign ( $CLBP_S$ ) and CLBP-Magnitude ( $CLBP_M$ ), respectively, are proposed to encode them, where the  $CLBP_S$  is equivalent to the conventional LBP, and the  $CLBP_M$  measures the local variance of magnitude. The  $CLBP_M$  is defined as follows:

$$CLBP\_M_{P,R} = \sum_{p=1}^{P-1} t(m_p, c) 2^p, \quad t(x, c) = \begin{cases} 1, & x \geq c \\ 0, & x < c \end{cases}$$

where threshold  $c$  is the mean value of  $m_p$  of the whole image. CLBP-Center ( $CLBP_C$ ) operator extracts the local central information as  $CLBP\_C_{P,R} = t(g_c, c_I)$  where threshold  $c_I$  is set as the average grey level of the whole image. Over all descriptor gained by combining the three operators  $CLBP_S$ ,  $CLBP_M$  and  $CLBP_C$ .

## 2.2. Review of Scattering Transform

Scattering transform was introduced by Mallat in [5] as a Scattering Network (ScatNet). It is implemented by a deep convolution network, in which wavelet transformation followed by modulus non-linearities operators are consecutively computed. A Scattering presentation of a texture image which preserves enough information so that the invariance to rotation, translation, deformation, and shear can be obtained with linear projection at the classifying level using PCA classifier as stated in [6].

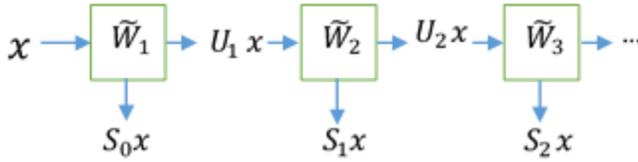


Figure 1. Scattering representation is computed by a cascade of wavelet-modulus operators  $\tilde{W}_m$ . Every  $\tilde{W}_m$  has one input and two outputs which are the invariant scattering coefficients  $S_m x$  and the next layer of covariant wavelet modulus coefficients  $U_{m+1} x$ . The latter is used for further transformations.

An input image  $x$  (Figure 1) is used to calculate  $S_0 x$  and  $U_1 x$ , which in turn form the first wavelet modulus operator  $\tilde{W}_1$ ,

$$\text{with } \tilde{W}_1(x) = (S_0(x), U_1(x))$$

$$\text{where } S_0 x(u) = x \star \phi_j(u) = \sum_v x(v) \phi_j(u-v)$$

$$\text{and } U_1(x) = x \star \psi_{j,\theta}(x)$$

$\phi_j(u) = 2^{-2j} \phi(2^{-j}u)$  is a Gaussian low pass filter. This leads to the averaged image  $S_0 x$  is almost invariant to rotations and translations up to  $2^j$  pixels, while it loses the high frequencies of  $x$ . These will be gotten back by the convolution with high pass wavelet filters. Then the wavelet  $\psi$  is rotated by  $\theta$  angles and dilated by  $2^j$  in order to obtain rotation covariant coefficients.

Where  $S(x)$  is called scattering coefficient of the network,  $U(x)$  is the wavelet coefficient, and  $\star$  is a convolution operator. Finally, the scattering features vector of an image are obtained by concatenating scattering coefficients of all network layers  $S(x) = (S_0(x), S_1(x), S_2(x))$ .

Sifre et al. showed in [4] that maximum number layers of the network is 3, if this number exceeds 3 then the energy will decay, and so no more useful signal for discrimination.

## 2.3. PCA Classifier

A generative classifier called Principal Component Analysis (PCA) [6] was proved to have decent performance for ScatNet in case of small training dataset. PCA Classifier is described as following.

Given a test image  $X$ ,  $\tilde{S}X$  denotes the scattering transform of  $X$  and its dilated version  $\mathcal{D}_j X$ .

$$\tilde{S}X = \left( \sum_{0 \leq j < H} 1 \right)^{-1} \sum_{0 \leq j < H} \tilde{S} \mathcal{D}_j X. \quad (3)$$

The representation of  $\tilde{S}X$  used at test time is therefore a scattering transform.

Let  $P_{U_c} \tilde{S}X$  denotes the orthogonal projection of  $\tilde{S}X$  in the scattering space  $U_c$  of a given class  $c$ . The principal components space  $U_c$  is approximately computed from the singular value decomposition (SVD) of the matrix of centered training sample  $\tilde{S} \mathcal{D}_j X_{c,i} - \mu_c$  with all possible samples  $i$  dilated by  $2^j$  for a given class  $c$ . The PCA classification computes the class  $\hat{c}(X)$  base on the minimum distance  $\|(Id - P_{U_c})(\tilde{S}X - \mu_c)\|$  from  $\tilde{S}X$  to the space  $\mu_c + U_c$ , (Figure 2)

$$\hat{c}(X) = \arg \min_c \|(Id - P_{U_c})(\tilde{S}X - \mu_c)\|^2 \quad (4)$$

where  $\mu_c$  is the average of scattering transform  $\tilde{S}X_{c,i}$  for all training samples  $X_{c,i}$  of class  $c$ .

$$\mu_c = \left( \sum_{0 \leq j < H} 1 \right)^{-1} \sum_{0 \leq j < H} \tilde{S} \mathcal{D}_j X_{c,i} \quad (5)$$

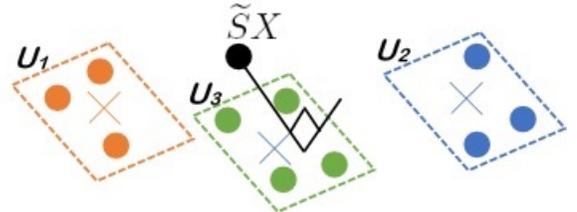


Figure 2. PCA-classifier classifies a test image  $X$  based on the minimum distance from scattering transform  $\tilde{S}X$  to subspace  $\mu_c + U_c$ .

The value  $H$  quantifies the ranges of scale invariance, e.g for a training set with dilated versions of each  $X_{c,i}$  by different scaling factors  $2^j$  for  $0 \leq j < H$ .  $H$  is the range

of scale invariance, limited by the size images. Typically,  $H = 2$  and sample  $j$  at half integer which leads to 4 scaling factors  $\{1, \sqrt{2}, 2, 2\sqrt{2}\}$ , and the dilated samples  $\mathcal{D}_{c,i}(u) = X_{c,i}(2^j u)$ .

### 3. Proposed Method

LBP proposed by Ojala et al. [1] and its variants such as CLBP [3], scLBP [7], BF+CLBP [8], and MRELBP [9] have the high performance on datasets such as OUTEX [10] because they capture well the small local structure information of the image data. On the other hand, scattering transform method [4], [5], [6] extracts a wider range of signals for its features. According to the observation of those, we intuitively think that if there is a descriptor which employs both small local and global structure information it would make texture classification have promising results.

It can be vividly seen that LBP family are vulnerable to image scale because of the fixed radius from which the neighborhood chosen for thresholding. A measure to this is that we make use of a multi-scale LBP to extract local structure cues from texture images, and the wide range complementary signals from Scatnet. Regarding to the weak point of scattering transform, a lack of small local structure information is compensated by LBP features, so we keep it intact. Overall idea is broadly illustrated in Figure 3.

Given an input image  $x$ , then  $y_1 = CLBP_S(x)$ ,  $y_2 = CLBP_M(x)$ ,  $y_3 = CLBP_C(x)$  are sequentially various types of CLBP features extracted from the image.

Let  $f_1 = h(y_1, y_2, y_3)$  denotes the 3D joint histogram ( $CLBP_S/M/C$ ) as proposed in [3], and  $f_2 = g(Sx)$  is a concatenation of scattering coefficients then  $f = c(f_1, f_2)$  is the concatenating operator which forms the descriptor of our proposal. This will be not only tolerant to scaling but also preserve enough small local structure signals for discrimination. It should be noticed that we focus mainly upon building a descriptor based on scattering transform rather than techniques of Scatnet [4] such as multi-scale average, and multi-scale training to augment classification results.

$$input_x \rightarrow \left\{ \begin{array}{l} SCatNet \\ CLBP \end{array} \right\} \rightarrow PCA \text{ Classifier}$$

Figure 3. ScatNet, CLBP Combination

## 4. Experimental Validation

### 4.1. Experimental Settings

We test the effectiveness of our method by doing experiments on eight texture databases: ALOT [11], CUREt [12], KTHTIPS2-a&b [13], and Outex [10].

ALOT dataset consists of 250 classes with variety of viewing angles, illumination, and color. Each material has 100 samples. There are different versions of ALOT, they are full/half resolution, color, etc. We choose the Grey

value version with the resolution of 384 by 256 to test our descriptor.

While CUREt database contains 61 texture classes, 205 images per class, acquired at different viewpoints, illumination, and orientations. There are 118 images shot from a viewing angle of less than 60 degrees. We choose a subset 92 images from 118 for each class with totally  $61 \times 92 = 5612$  images are selected. According to this approach a sufficiently large region could be cropped ( $200 \times 200$ ) across all texture classes. All cropped regions are converted to grey scale.

The material databases KTHTIPS2a, KTHTIPS2b [13], with 3 viewing angles, 4 illuminants, and 9 different scales, producing 432 images per class with totally 11 classes.

Outex database contains textural images which are captured from a wide variety of real material surfaces. We consider the two commonly used test suites, Outex\_TC\_00010 (TC10) and Outex\_TC\_00012 (TC12), containing 24 classes with up to 200 texture images per class. This database is built by taking images under three different illuminations ("horizon", "inca", and "t184").

The last database we use for testing our proposal is UIUC [14] which contains 25 classes with 40 samples per class. We follow the standard classification protocol of UIUC. The mean classification accuracy and standard deviation over 10 random splits between training and testing with 20 samples per class chosen for training.

PCA classifier [6] is used in our experimentation. For CLBP, parameters chosen as following, the number of neighbors ( $P=8$ ), and Radius ( $r=1,2,3$ ) for three different scales. Whereas, arguments of ScatNet for scattering transform are selected such that number of scales in filter banks ( $J=4$  or  $5$  depend on the resolution of images in datasets), Orientations of filter bank ( $L=8$ ), number of ScatNet Layers ( $M=3$ ).

### 4.2. Classification Results

Intensive experiments were conducted on eight texture datasets, the results are compared with state-of-the-art of those, we chose the highest results reported by CLBP [3], BF + CLBP [8], scLBP [7], and MRELBP [9] for the comparison. Table 1 shows that our proposal has the accuracy which is comparable to some, while the novel descriptor is consistently better than the ones it inherits from, CLBP and Scattering transform coefficient, on the experimented data sets. It should be noticed that we do not use multi-scale ScatNet because our focus in this paper is the complementation of CLBP and ScatNet. So it is not surprising when multi-scale CLBP has a better classification rate than ScatNet in some cases even in a multi-scale dataset such as UIUC.

For the experiments on ALOT and CUREt database, we follow the training and testing scheme used in [8], [9], a half of class samples chosen for training while the remaining for testing. Splits are implemented 10 times independently, the average accuracy over 10 randomly partitions are selected. The correct classified rate of our method reaches State-of-the-art on these two data sets, the proportion are above

TABLE 1. CLASSIFICATION ACCURACY(%) ON CURET, OUTEX, KTH-TIPS2A, KTH-TIPS2B, ALOT, AND COMPARING WITH STATE OF THE ART RESULTS

	CLBP	ScatNet	C-ScatNet(ours)	scLBP [7]	BF+CLBP/CLBC [8]	MRELBP [9]
ALOT	97.13 ± 0.18	98.11 ± 0.32	<b>99.57 ± 0.17</b>	-	-	99.08
CURET	99.51 ± 0.18	97.76 ± 0.57	<b>99.74 ± 0.17</b>	99.33	97.65	99.02
OUTEX TC10	99.32	98.39	<b>99.51</b>	-	99.40	99.87
OUTEX TC12 "f"	96.04	96.44	<b>98.66</b>	97.73	95.37	99.49
OUTEX TC12 "h"	97.25	97.45	<b>98.94</b>	98.56	94.72	99.75
KTH-TIPS2a	89.39±1.74	89.52±1.74	<b>92.33±1.28</b>	78.53	-	-
KTH-TIPS2b	63.3±1.3	65.08±1.20	<b>69.03±1.41</b>	-	-	77.91
UIUC	96.02 ± 1.25	93.84 ± 1.48	<b>97.50 ± 0.73</b>	98.45	-	-

99% which gains a competitive advantage over the results reported in [7], [8], [9].

Regarding to KTH-TIPS2a and KTH-TIPS2b databases, standard classification protocols [13] is applied, only unseen data is used for test. With KTH-TIPS2a, three out of four samples are used for training, the remaining is used for testing. The results are reported as the mean over four test runs. Our method has an improvement up to 9% comparing to the original descriptors, see Table 1. Whereas, on KTH-TIPS2b data set, one sample per class is used for training and the remaining three unseen samples are used for testing. As can be seen from Table 1, our method outperforms CLBP, scattering transform descriptor up to 8%. Our testing protocol on this dataset is different from the one in [9], which three samples per class are used for training, and one for testing. This can lead to a different result.

In case of Outex database, experiment conducted on 2 test suites, Outex\_TC10 and Outex\_TC12. In more detail, for Outex\_TC10 training is data with illuminants (inca) and rotation (0°) while testing is on the same illuminants and rotation at { 5°, 10°, 15°, 30°, 45°, 60°, 75°, 90° }. Similar protocol is applied for Outex\_TC12 but with different rotation angles. We first show how far our method can improve comparing with CLBP and ScatNet standing alone. For TC10 (Table 1), there is a slight improvement in comparison with the original versions at 99.51% (recent state-of-the-art reported in [9] is 99.87%). Higher improvement gained in Outex\_TC12 at 98.94% accuracy.

Finally, we get around 4% classification enhance on UIUC (from 93.84% to 97.50% when using LBP-Scattering Transform combined descriptor).

## 5. Conclusion

In this paper, we have proposed a method to compensate the local structure information for ScatNet, namely Complementary Scattering Network (C-ScatNet). This novel descriptor exploits the integration of local structure information of LBP features and the global ones extracted by ScatNet in order to enhance distinctiveness of texture while preserving the robustness to variations in illumination, rotation, and noise. Our experiments show that the C-ScatNet achieves high competitive results on eight different available datasets. Overall, Scatnet and LBP are not concurrent, but complementary. Future study can be drawn on the same

domain with scale variation tolerance by using multi-scale average and training technique.

## References

- [1] T. Ojala, M. Pietikainen, and T. Maenpaa, "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, Jul. 2002.
- [2] N.-S. Vu, "Exploring patterns of gradient orientations and magnitudes for face recognition," *IEEE Trans. Information Forensics and Security*, vol. 8, no. 2, pp. 295–304, 2013.
- [3] Z. Guo, L. Zhang, and D. Zhang, "A completed modeling of local binary pattern operator for texture classification," *IEEE Trans. Image Processing*, vol. 19, no. 6, pp. 1657–1663, Jun. 2010.
- [4] L. Sifre and S. Mallat, "Rotation, scaling and deformation invariant scattering for texture discrimination," in *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, 2013, pp. 1233–1240.
- [5] S. Mallat, "Group Invariant Scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.
- [6] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 35, no. 8, pp. 1872–1886, 2013.
- [7] J. Ryu, S. Hong, and H. S. Yang, "Sorted consecutive local binary pattern for texture classification," *IEEE Transactions on Image Processing*, vol. 24, no. 7, pp. 2254–2265, 2015.
- [8] N. Vu, T. P. Nguyen, and C. Garcia, "Improving texture categorization with biologically-inspired filtering," *Image Vision Comput.*, vol. 32, no. 6-7, pp. 424–436, 2014.
- [9] L. Liu, P. W. Fieguth, M. Pietikäinen, and S. Lao, "Median robust extended local binary pattern for texture classification," in *2015 IEEE International Conference on Image Processing, ICIP 2015, Quebec City, QC, Canada, September 27-30, 2015*, 2015, pp. 2319–2323.
- [10] T. Ojala, T. Menp, M. Pietikinen, J. Viertola, J. Kyllnen, and S. Huovinen, "Outex - new framework for empirical evaluation of texture analysis algorithms," in *16th Int. Conf. on Pattern Recognition*, 2002, pp. 701–706.
- [11] G. J. Burghouts and J. Geusebroek, "Material-specific adaptation of color invariant features," *Pattern Recognition Letters*, vol. 30, no. 3, pp. 306–313, 2009.
- [12] K. J. Dana, B. van Ginneken, S. K. Nayar, and J. J. Koenderink, "Reflectance and texture of real-world surfaces," *ACM Trans. Graph.*, vol. 18, no. 1, pp. 1–34, 1999.
- [13] B. Caputo, E. Hayman, M. Fritz, and J.-O. Eklundh, "Classifying materials in the real world," *Image Vision Comput.*, vol. 28, no. 1, pp. 150–163, 2010.
- [14] 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, 2005.