A Rotation-invariant Script Identification based on BEMD and LBP

Jianjia Pan
Department of Computer Science, Hong Kong Baptist University
jjpan@comp.hkbu.edu.hk

Abstract

Script identification is very important to develop the scripts OCR systems. In this paper, we proposed a new algorithm for script identification based on the global and local texture of document images. The BEMD method is used to decompose the image to some components (IMFs) and then the Local Binary Patterns (LBP) method is used to detect the features. Experiments shown the recognition rate based on BEMD-LBP is as well as the LBPV and wavelet based energy feature in 0 angles. At the same time, for the different angles rotation script, the BEMD-LBP feature present some robust adaptive to the rotation script.

1. Introduction

Script recognition is a basic research topic in document analysis, and is also a difficult and time cost problem. The most feature extractors for script rely on the assumption that character can be defined by the local statistical properties of pixel gray levels. Several script analysis systems have been developed.

There have been many script recognition systems, they can be classified into three main approaches, statistics-based [12], [17], [18], texture-based [14], [15] and token-based [13]. The statistics-based approaches identify scripts through the analysis of the distribution of the upward concavity [12] or horizontal projection profile [17], [18]. Texture-based approaches identifies scripts based on the texture detected by some features, such as Gabor filters [14], Gray-level co-occurrence matrix[16] and wavelet-based energy[19]. In [13], the character tokens which specific to different scripts are used for script identification.

The statistics-based approach, which detects upward concavities or horizontal projection profile in an image, is highly sensitive to noise and image quality[15,16]. These approaches usually use the connected components, which should segment the characters before the follow processing. On the other hand, the scripts often have a distinctive visual appearance, so the script document can be considered as a texture. From this, the problem of script identification can be changed into the texture classification problem. For the texture-based approaches, it is not need to extract individual characters, and there is no script-dependent processing.

There have been some texture-based approaches proposed for the script identification, for example, Gabor filter banks[14], wavelet energy[19] and wavelet gray-level co-occurrence matrix[16]. Recently, Empirical mode decomposition (EMD), developed by Huang [1], has been used for the texture analysis and face recognition [2]. EMD is a data driven processing algorithm which applies no predetermined filter. The EMD is based on the local characteriristic scale of the data, which is able to perfectly analyze the nonlinear and nonstationary signals. EMD has present some better quality than Fourier, wavelet and other decomposition algorithms in extracting intrinsic components of textures because of its data driven property [2, 4].

In this paper, we proposed a new algorithm for script identification based on the global and local texture of document images. The key point is using texture analysis to extract the features. The BEMD method is firstly used to decompose the image to some components (IMFs) and then the Local Binary Patterns (LBP) method is used to detect the features. Experiments shown the BEMD-LBP method can identify scripts accurately and is robust to the text line simultaneously compared with other texture-based approaches.

2. Review of BEMD

Empirical Mode Decomposition (EMD) is first proposed by Huang et al. [1] for the processing of non-stationary functions. The tool decomposes signals into components called Intrinsic Mode Functions (IMFs) satisfying the following two conditions:

(a).The numbers of extrema and zero-crossings must either equal or differ at most by one;

(b).At any point, the mean value of the envelope defined by the local maxima and the envelope by the local minima is zero.

Huang [1] have also proposed an algorithm called 'sifting' to extract IMFs from the original signal $f(t)$ as follows:

$$f(t) = \sum_{i=1}^{N} I_i(t) + r_N(t) \qquad (1)$$

Where $I_i(t)$, i=1,…,N are IMFs and $r_N(t)$ is the residue.

The bidimensional EMD (BEMD) process is conceptually the same as the one dimension EMD, except that the curve fitting of the maxima and minima envelope now becomes a surface fitting exercise and the identification of the local extrema is performed in space to take into account for the connectivity of the points.

The main process of the BEMD can be described as:

(a).Locate the maximum and minimum points in the image *I(k)*;

(b).Interpolation the surface between the all maxima (resp. munima) to build the envelope *Xmax(k)* and *Xmin(k)*;

(c).Compute the mean envelope function

$$Xm(k) = (Xmax(k) + Xmin(k))/2; \quad (2)$$

(d).Update the *I(k)=I(k-1)-Xm(k)* ;

(e).Check the stopping criterion

$$SD = \frac{1}{N}\sum_{k=0}^{K}\frac{(I_{i,j-1}(k) - I_{i,j}(k))^2}{I_{i,j-1}^2(k)} \quad (3)$$

if SD is larger than a threshold ε, repeat the steps (a)-(e) with *I(k)* as the input, other wise, *I(k)* is an IMF *d(k)*;

(f).Update the residual *I(k)=I(k-1)-d(k)*;

(g).Input the *I(k)* to steps(a)-(e) until it can not be decomposed, and the last residual *I(k)=r(n)*.

After the BEMD, the decomposition of the image can be written as following form:

$$I(n) = \sum_{k=1}^{K} d_k(n) + r(n) \quad (4)$$

The *d(k)* is the IMFs (intrinsic mode functions) of the images, and *r(n)* is the residual function.

3. A new BEMD based on self-similar

With the intension of some difficult in implement BEMD, we used some methods to improve the BEMD. The local extrema are detected based on its neighbor and the extended parts are rebuilt based on self-similarity.

3.1 Local extrema detection

Detection the local extrema means finding the local maxima and minima points from given images. In the normal BEMD methods [1,4], the mathematical morphology method is used to local the extrema, but we find the extrema points will be reduced fast. It means that, after two or three surface interpolations, the image will be too smooth to local any significative extrema points. Neighbor location method [7] is used to detect the extrema in our method.

Definition 1: f[i,j] is a maximum (or. minimum) if it is larger (or. lower) than the value of *f* at the nearest neighbors of *[i,j]*.

Let X be an $M\times N$ 2D matrix represented by

$$X = \begin{bmatrix} x_{11}x_{12}\cdots x_{1N} \\ x_{21}x_{22}\cdots x_{2N} \\ \vdots \ \ \vdots \ \ \cdots \ \ \vdots \\ x_{M1}x_{M2}\cdots x_{MN} \end{bmatrix} \quad (5)$$

x_{mn} is the element of X located in the mth row and nth column.

Let the window size for local extrema determination be $(2w+1)\times(2w+1)$, Then,

$$x_{mn} = \begin{cases} Local\ Maximum & if\ x_{mn} > x_{ij} \\ Local\ Minimum & if\ x_{mn} < x_{ij} \end{cases} \quad (6)$$

Where

$$x_{ij} = \{x \mid (m-w):i:(m+w),(n-w):j:(n+w)\}$$
$$i \neq m,\ j \neq n \quad (7)$$

From the experimental, we find 3×3 window results in an optimum extrema map for a given image. The larger windows are also used in some conditions to reduce the computation, but as the mathematical morphology method, the extrema points will be reduced fast.

3.2 Surface interpolation method

Another difficulty in the BEMD comes from generating a smooth fitting surface to the identified maxima and minima. There are several interpolation methods for BEMD. Nunes [2, 8] used the radial basis function (RBF) for surface interpretation. Linderhed [9] used the spline for surface interpretation to develop two-dimensional EMD data. Damerval [10] used a third way based on Delaunay triangulation to obtain an upper surface and a lower surface. Delaunay triangulation can effectively reduce the interpolation computation. Our interpolation method is based on a Delaunay triangulation.

3.3 Self-similar for BEMD Boundary Processing

A self-similar object is exactly or approximately similar to a part of itself, which means the whole has the same shape as one or more of the parts. Many objects in the real world are statistically self-similar: parts of them show the same statistical properties at many scales. Self-similarity is a typical property of fractals.

The self-similar feature means that, irrespective of the complexity of the shape of an object, by looking deeply into its structure, an observer can be the same (or similar) shapes on contractible scales. In the boundary process, we use this self-similar property to build the extend boundary. The basic idea is: the extend part can find a self-similar part in the original image.

The concrete algorithm is as follows:

Assume the size of original image I is $N\times N$. The size of the extend block is $M\times M$. After extending, the extended image is $(N+2M)\times(N+2M)$ with middle $N\times N$ block the original image. The original image I is divided to $M\times M$ size blocks. For each extended block $part_e$, its three neighbor blocks in the original image are defined as its neighbor blocks $part_n$. And then in the image I, find the blocks which are the most similar to the $part_n$. The similar judgment criterion is based on the

MAD (Mean Absolute Difference) for representing the distances different between boundary blocks and the matched blocks. At last, the block with most similar neighbor blocks is used as the extend block.
After the self-similar based extension boundary processing, the boundary interference of the BEMD will be reduced, and the IMF components is more significant.

4. LBP based on BEMD

To classify the rotation script images, we proposed to use the LBP to extract the local features from the IMFs. Local Binary Patterns (LBP) is introduced as a powerful local descriptor for microstructures of images [20]. The LBP operator labels the pixels of an image by thresholding the 3 × 3-neighborhood of each pixel with the center value and considering the result as a binary string or a decimal number.

4.1 Local Binary Patterns (LBP)

The LBP operator was originally developed for texture description. The operator assigns a label to every pixel of an image by thresholding the 3x3-neighborhood of each pixel with the center pixel value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor. Figure 1 shows an example of the LBP operator[24].
The form of the resulting 8-bit LBP code can be defined as follows:

$$LBP(x_c, y_c) = \sum_{n=0}^{7} s(i_n - i_c)2^n \qquad (8)$$

where i_c corresponds to the gray value of the center pixel (x_c, y_c), into the gray values of the 8 neighborhood pixels, and function s(x) is dedined as:

$$s(x) = \begin{cases} 1 & if \quad x \geq 0 \\ 0 & if \quad x \leq 0 \end{cases} \qquad (9)$$

From the above processing, the LBP present that it will be not affected by any monotonic gray-scale transformation which preserves the pixel intensity order in a local neighborhood. Each bit of the LBP code has the same significance level and that two successive bit values may have a totally different meaning.

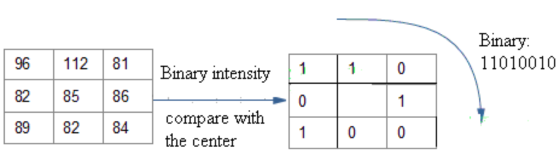


Figure 1 The LBP operator

To deal with textures at different scales, the LBP operator was later extended to use neighborhoods for different sizes [20]. The local neighborhood is extended to as a set of sampling points evenly spaced on a circle centered at the pixel to be labeled allows any radius and number of sampling points[22]. If a sampling point is not in the center of a pixel, it will be rebuilt by bilinear interpolation. The notation (P;R) is defined as the pixel neighborhoods which means P sampling points on a circle of radius of R. Figure 2 shows an example of circular neighborhoods.

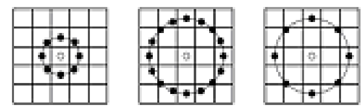


Figure2 The circular (8,1) (16,2) (8,2) neighborhoods

Another extension to the original operator is the definition of so called *uniform patterns* [20]. A local binary pattern is called *uniform* if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular [22]. For example, the patterns 11111111 (0 transitions), 00011000 (2 transitions) and 11100011 (2 transitions) are uniform, the patterns 11001001 (4 transitions) and 01010111 (6 transitions) are not. In the LBP histogram, uniform patterns are used so that the histogram has a separate bin for every uniform pattern and all non-uniform patterns are assigned to a single bin[21].

4.2 LBP based on BEMD

In this paper, a script image is firstly decomposed by BEMD into several sub-images IMFs, and then, the LBP is used to extract those IMFs. We can use them as a set of the features to classify the script characters.
BEMD is based on the local characteriristic scale of the data, which is able to perfectly analyze the nonlinear and nonstationary signals. The details in the global and local information of the different script are extracted.
LBP has several properties that favor its usage in rotation-invariant script identification. Because of the invariance of the LBP features, the LBP can be suit for the considerable gray-scale variations in images and no normalization of input images is needed. Secondly, the LBP features are very fast to compute. Thirdly, LBP is a nonparametric method, which means that no prior knowledge about the distributions of images is needed. The operator does not require many parameters to be set.
We use the following notation for the script features:
Firstly, the original image I is decomposed to its IMFs d_k:

$$I(n) = \sum_{k=1}^{K} d_k(n) + r(n) \qquad (10)$$

Secondly, the $LBP^u(P,R)$ is used to detect the uniform patterns of the IMFs. The subscript represents using the operator in a (P;R)neighborhood. Superscript *u* stands for using only uniform patterns.

Thirdly, the different $LBP_i^u(P,R)$ for IMF_i are combined with weighting rules:

$$LBP^u(P,R) = \sum_{i=1}^{n} w_i * LBP_i^u(P,R) \quad (11)$$

Where $LBP_i^u(P,R)$ indicate the LBP corresponding to the IMF_i . Superscript u reflects the use of rotation invariant 'uniform' patterns. (P;R) is used for pixel neighborhoods which means P sampling points on a circle of radius of R. w_i indicate the corresponding weights, the sum of the those weights is 1.

4.3 The classifier based on SVM

Support Vector Machines (SVM) has become a hot research topic in machine learning because of its excellent statistical learning performance. It has been widely applied to pattern recognition. Simply, the principle of constructing the optimal separating hyperplane is that the distance between each training sample and the optimal separating hyperplane is maximum.[20]

Let $\{(x_i, y_i)\}$ ($1 \le i \le N$) be a linearly separable set. Where, $x_i \in R^d$, $y_i \in \{-1,1\}$, and y_i are labels of categories. The general expression of the linear discrimination function in d-dimension space is defined as $g(x) = w*x + b$, and the corresponding equation of the separating hyperplane is as follows: $w*x + b = 0$.

Normalize $g(x)$ and make all the x_i meet $g(x) \ge 1$, that is, the samples which are closed to the separating hyperplane meet $|g(x)| = 1$.Hence, the separating interval is equal to $2/\|w\|$, and solving the optimal separating hyperplane is equivalent with minimizing $\|w\|$.The object function is as follows:

$$\text{Min}\ \Phi(w) = \frac{1}{2}\|w\|^2 \quad (12)$$

Subject to the constraints:

$$y_i(w \bullet x_i + b) \ge 1, i=1,\ldots,N \quad (13)$$

When adopting Lagrangian algorithm and introducing Lagrangian multipliers $\alpha = \{\alpha_1,\ldots,\alpha_N\}$, the problem mentioned above can be converted into a quadratic programming problem and the optimal separating hyperplane can also be solved.

Where, $w = \sum_i \alpha_i y_i x_i$,

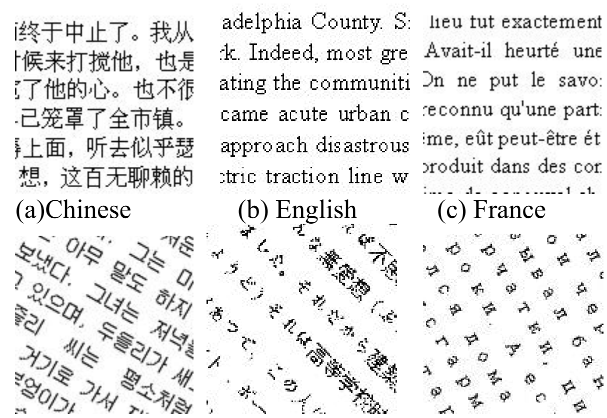
x_i is the sample only appearing in the separating interval planes. These samples are named support vectors and the classification function is defined as follows:

$$f(x) = \text{sgn}\left(\sum_i \alpha_i y_i x_i \bullet x + b\right) \quad (14)$$

For our experiments, we use the RBF kernel because it offers better discrimination than the linear kernel, while using less parameter than the polynomial kernel.

5. Experimental results

The proposed algorithm for script identification was tested on a database containing six different script types (English, France, Chinese, Japanese, Russian and Korean). Each script has 250 samples for training, and 500 samples for testing. Some examples of the images are shown in Figure. 3. Each document is scanned on the gray level of 0-255, each 128*128 pixels in size, extracted for each script class.



(a)Chinese (b) English (c) France

(d) Korean 30° (e) Japanese 45° (f) Russian 60°

Figure 3 some samples of the script image

The classifier is based on SVM described as above[23]. The proposed feature is compared with the wavelet based energy feature [19] and the LBPV feature [21].

Table 1 shows the result of the identification. As the table shown, the recognition rate based on BEMD-LBP is as well as the wavelet based energy feature in 0 angles. For the English and the France, which are both Latin word, the similarity of their textures is large than other scripts, the recognition rate of BEMD-LBP and LBPV are lower than the wavelet based energy feature. At the same time, for the different angles rotation script, the BEMD-LBP feature present some robust adaptive to the rotation script. The wavelet energy features shown to be sensitive to the script rotation.

Angles	English	France	Chinese	Japanese	Russian	Korean	Average
0^o BEMD-LBP	92.51	91.53	**97.38**	**96.24**	**96.52**	**98.56**	95.45
0^o wavelet[19]	**95.31**	**94.63**	94.21	93.58	93.73	94.71	94.36
0^o LBPV[21]	88.55	89.42	94.60	93.78	94.77	95.13	92.70
30^o BEMD-LBP	74.55	73.38	**84.38**	**83.53**	**82.13**	**86.58**	80.75
30^o wavelet[19]	**75.14**	**74.27**	78.54	77.43	75.39	80.27	76.84
30^o LBPV[21]	70.43	69.26	82.24	82.61	71.25	81.51	76.21
45^o BEMD-LBP	**50.54**	**49.36**	**58.53**	**57.87**	**55.44**	**60.78**	55.42
45^o wavelet[19]	40.23	39.76	44.38	43.93	42.35	43.57	42.37
45^o LBPV[21]	43.14	42.48	49.14	49.76	48.22	50.50	47.20
60^o BEMD-LBP	**70.16**	**72.61**	**79.40**	**78.53**	**77.50**	**79.11**	76.21
60^o wavelet[19]	65.37	67.93	68.30	73.75	72.42	76.73	70.75
60^o LBPV[21]	65.10	64.83	75.58	69.46	74.22	77.20	71.06
90^o BEMD-LBP	88.45	87.73	**93.56**	**92.10**	**91.83**	**94.27**	91.32
90^o wavelet[19]	**89.53**	**88.27**	90.40	92.33	88.52	89.18	89.70
90^o LBPV[21]	85.14	84.57	90.21	85.68	85.94	90.44	86.99

Table 1 Recognition rate for different angles script by proposed method and wavelet energy [19], LPBV[21]

6. Conclusion

Script identification is very important for development of multi-script OCR systems. In this paper, a new global-local feature is proposed to identify the word-wise printed script. Firstly, the BEMD decompose the image to different IMFs, which present different scale information of the original image. And then the LBP method is used to detect the local information of IMFs. The experimental shown the combined BEMD-LBP features is robust adaptive to the rotation script compared to wavelet energy features.

In the following work, we will improve the features to be adaptive to the Latin font. On the other hand, this new texture feature will be applied to classify the nature texture images, and other classers will be used for the recognition.

References

[1] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, et. “The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis”. Proceedings of Royal Society. Lond, (A) vol.454, pp:903-1005.1998

[2] J.C.Nunes, Y. Bouaoune, E. Delechelle, O.Niang, Ph. Bunel. “Image analysis by bidimensional empirical mode decomposition”. Image and Vision Computing Vol.21(12) , pp:1019-1026, November 2003

[3] C.Damerval, S.M eignen, and V.Perrier. “A fast algorithm for bidimensional EMD”, IEEE signal processing letters, vol.12,no.10, pp 701-704, 2005

[4] N.E.Huang, M.L.C.Wu, S.R.Long, “A confidence limit for the EMD and Hilerbet spectral analysis”, Proceeding of the royal society A, vol 459., pp 2317-345, 2003

[5]P. Flandrin, G. Rilling and P. Goncalves. “Empirical mode decompositionas a filter bank”. IEEE Signal Processing Letters, vol.11(2),pp:112-114, 2004.

[6]Bhagavatula,R., MariosSavvides, and M. “Acoustics. Analyzing Facial Images using Empirical Mode Decomposition for Illumination Artifact Removal and Improved Face Recognition.” IEEE International Conference on Speech and Signal Processing, 2007. Vol. 1, pp. 505-508. April 2007

[7] Sharif M. A. Bhuiyan, Reza R. Adhami, and Jesmin F. Khan. “Fast and Adaptive Bidimensional Empirical Mode Decomposition Using Order-Statistics Filter Based Envelope Estimation”. EURASIP Journal on Advances in Signal Processing, vol. 2008, Article ID 728356, 18 pages, 2008.

[8] Nunes J. C., Guyot S., and Deechelle E. “Texture analysis based on local analysis of the Bidimensional Empirical Mode Decomposition”. Machine Vision and Applications vol.16(3), pp. 177-188, 2005.

[9] Linderhed, A., “Variable sampling of the empirical mode decomposition of two-dimensional signals,” International Journal of Wavelets, Multiresolution and Information Processing., vol.3, 435–452, 2005

[10] Damerval, C., S. Meignen, and V. Perrier, “A fast algorithm for bidimensional EMD”, IEEE Signal Processing Letter, vol.12(10), 701–704, October, 2005.

[11] Zhao Na, “A Periodic Extension Approach for HHT Empirical Mode Decomposition”, Computer Simulation, vol.25(12) 2008

[12] A.L. Spitz, “Determination of Script and Language Content of Document Images,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 3, pp. 235-245, Mar. 1997.

[13] J. Hochberg, L. Kerns, P. Kelly, and T. Thomas, “Automatic ScriptIdentification from Images Using Cluster-Based Templates,” IEEETrans. Pattern Analysis

and Machine Intelligence, vol. 19, no. 2,pp. 176-181, Feb. 1997.
[14] T.N. Tan, “Rotation Invariant Texture Features and Their Use in Automatic Script Identification,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 7, pp. 751-756, July 1998.
[15] A. Busch, W.W. Boles, and S. Sridharan, “Texture for Script Identification,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 27, no. 11, pp. 1720-1732, Nov. 2005.
[16] P.S. Hiremath, S. Shivashankar, "Wavelet based co-occurrence histogram features for texture classification with an application to script identification in a document image" Pattern Recognition Letters, pp. 1182–1189, vol.29, 2008
[17] U. Pal and B.B. Chaudhury, “Identification of Different Script Lines from Multi-Script Documents,” Image and Vision Computing, vol. 20, no. 13-14, pp. 945-954, 2002.
[18] A.M. Elgammmal and M.A. Ismail, “Techniques for Language Identification for Hybrid Arabic-English Document Images,” Proc. Sixth Int’l Conf. Document Analysis and Recognition, pp. 1100-1104, 2001.
[19] Li Zeng, Yuanyan Tang ,Tinghuai Chen, "Multi-scale wavelet texture-based script identification method " JISUANJI XUEBAO. Vol. 23, no. 7, pp. 699-704. July 2000
[20] T. Ojala, M. Pietikäinen, and T. T. Mäenpää, “Multiresolution gray-scale and rotation invariant texture classification with Local Binary Pattern,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.
[21] Zhenhua Guo, Lei Zhang, David Zhang "Rotation invariant texture classification using LBP variance(LBPV) with global matching" Pattern Recognition, 706–719 vol 43. 2010
[22] Guoying Zhao and Matti Pietikainen, "Dynamic Texture Recognition Using Local Binary Patterns with an Application to Facial Expressions", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 29, no. 6, pp. 915-928, 2007
[23] Library for Support Vector Machines http://www.csie.ntu.edu.tw/~cjlin/libsvm/
[24] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen, "Face recognition with local binary patterns",Lecture Notes in Computer Science, ECCV ,pp.469-481,2004