International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:7, No:5, 2013

Texture Feature-Based Language Identification Using Wavelet-Domain BDIP and BVLC Features and FFT Feature

Ick Hoon Jang, Hoon Jae Lee, Dae Hoon Kwon, and Ui Young Pak

***Abstract*—**In this paper, we propose a texture feature-based language identification using wavelet-domain BDIP (block difference of inverse probabilities) and BVLC (block variance of local correlation coefficients) features and FFT (fast Fourier transform) feature. In the proposed method, wavelet subbands are first obtained by wavelet transform from a test image and denoised by Donoho's soft-thresholding. BDIP and BVLC operators are next applied to the wavelet subbands. FFT blocks are also obtained by 2D (two-dimensional) FFT from the blocks into which the test image is partitioned. Some significant FFT coefficients in each block are selected and magnitude operator is applied to them. Moments for each subband of BDIP and BVLC and for each magnitude of significant FFT coefficients are then computed and fused into a feature vector. In classification, a stabilized Bayesian classifier, which adopts variance thresholding, searches the training feature vector most similar to the test feature vector. Experimental results show that the proposed method with the three operations yields excellent language identification even with rather low feature dimension.

***Keywords*—**BDIP, BVLC, FFT, language identification, texture feature, wavelet transform.

I. Introduction

In recent years, due to the rapid development of computers and smart devices we can easily convert paper documents into document files and store, manage, and share them for paperless office. Among several processes for such tasks, character recognition using an OCR (optical character recognition) system may be a core one. Most of state-of-the-art OCR systems have banks of multi-language OCR engines. Prior to inputting a document into an OCR system, the OCR engine proper for the document is usually selected by the user. With the trend of globalization of the world, we have more frequent chances to deal with multi-language documents so that the selection of the corresponding OCR engine by language identification may be smarter than the user selective one [1].

For convenience, we hereafter consider that the term "language identification" stands for script and/or language identification since script identification accomplishes language identification for scripts of a language. Up to now, many researches on language identification as a preprocessing of character recognition have been published. There are two major categories in typical language identification: statistics-based approach [2]-[4] and texture feature-based one [5]-[9]. The former utilizes the statistics of character pixels in a document image based on the fact that the structure of characters differs from that of another language. On the other hand, the latter considers a block of texts in a document image as a texture block so that the problem of language identification may be viewed as that of texture classification.

Hochberg *et al*. [2] proposed a language identification method to distinguish one of 13 languages from the others using cluster-based templates. In the method, frequent characters or word shapes of each language are clustered, and a representative template for each cluster is created. A language whose training templates produce the best match with the templates extracted from a test image is decided to be the identified one. Spits [3] suggested a method to determine one of six languages using vertical distributions of upward concavities of characters, normalized pixel distributions of binary character cells, and frequencies of occurrence of particular word shapes. Shijian and Tan [4] tried to identify the language of a document image by converting it into a document vector that characterizes the shape and frequency of its characters or words and finding the best matched vector. It was reported in an experiment for six languages that this method yields good performances on document images having noise and degradation.

In texture feature-based methods, one of the most important procedures is the extraction of texture features that can effectively represent the characters of each language. Pearke and Tan [5] extracted texture features using GLCM (gray-level co-occurrence matrix) and Gabor filters from document images. Tan [6] also exploited rotation invariant texture features using Gabor filters. Chan and Coghill [7] also adopted Gabor filters for extraction of texture features. Busch *et al*. [8] employed texture features of wavelet log co-occurrence. In [5]-[8], seven, six, 16, and eight languages are identified, respectively. An overview on script and/or language identification can be found in [1].

Lee *et al*. [9] adopted wavelet-domain BDIP, BVLC, and

This work was supported in part by the Agency for Defense Development with Project No. UD110091ED.

I. H. Jang is with the Department of Avionic Electronics Engineering, Kyungwoon University, Gumi, 730-739 Korea (phone: +82-54-479-1213; fax: +82-54-479-1149; e-mail: ihjang@ikw.ac.kr).

H. J. Lee is with the Division of Computer and Information Engineering, Dongseo University, Busan, 617-716 Korea (e-mail: hjlee@dongseo.ac.kr).

D. H. Kwon and U. Y. Pak are with the Agency for Defense Development, Daejeon, 305-152 Korea, (e-mail: dhkwon@add.re.kr, puy@add.re.kr).

NRMA (normalized magnitude) features for language identification. The authors demonstrated in their experiments that these features provide good performances for ten languages with very low feature dimension. The BDIP and BVLC operators had been proposed for image retrieval [10]. It had been shown that they yield very good performances not only in image retrieval [10], [11] but also in other areas of texture classification [12], face recognition [13], ROI determination [14], and language identification. BDIP is a kind of nonlinear operator which gives an average of normalized local gradients. It is known to effectively represent local bright variations. BVLC is a maximum deviation of normalized local correlations according to orientations, which is known to measure variation of texture smoothness well.

In texture classification, several texture features are usually fused for better performance. Fusion of features gives the increase of the dimension of feature vector and so the growth of computational loads, which may bring about the curse of dimensionality. Accordingly it is more desirable that we establish a texture classification method for language identification which shows a high identification rate but has a lower vector dimension as possible.

In this paper, we present a texture feature-based language identification method using wavelet-domain BDIP and BVLC features and FFT feature. In the proposed method, a test document image is decomposed into wavelet subbands. The noise in detail subbands is then reduced by Donoho's soft-thresholding [15]. The BDIP and BVLC operators are next applied to the smooth subband and the denoised detail subbands to extract the corresponding BDIP and BVLC subbands. The test image is also partitioned into blocks. 2D FFT is implemented for each block and some significant coefficients in each FFT block are selected. The magnitude operator is then applied to each significant coefficient. Moments of BDIP and BVLC subbands and those of FFT magnitudes are fused into a feature vector, which is finally classified by a stabilized Bayesian classifier.

The remainder of this paper is organized as follows. In Section II, we explain GLCM and Haralick features, spatial-domain BDIP and BVLC, BDIP and BVLC subbands in the wavelet domain, and FFT features. In Section III, we describe the proposed language identification method. Section IV discusses experimental results and Section V shows the conclusion.

II. Texture Feature-Based Language Identification and Texture Features

A. Texture Feature-Based Language Identification

Fig. 1 shows a block diagram of typical texture feature-based language identification. Let us assume that there are K language groups in the image DB for training, each of which has J document images. In training phase, texture features are extracted from each image $I_{k,j}$, $k = 1, \cdots, K$, $j = 1, \cdots, J$ in the image DB and formed as a feature vector $\mathbf{f}_{k,j}$. In a typical training scheme, the set of representative feature vectors $\{\bar{\mathbf{f}}_k \mid k = 1, \cdots, K\}$ for all language groups is calculated by averaging feature vectors $\mathbf{f}_{k,j}$. The set of covariance matrix **Ck** of feature vectors is also computed if necessary and the statistics are stored in the feature DB. In test phase, a test image I for language identification enters the system, which extracts its feature vector **f**. The classifier then searches the representative feature vector $\bar{\mathbf{f}}_c$ which is most similar to the test feature vector **f** among the training set $\{\bar{\mathbf{f}}_k \mid k = 1, \cdots, K\}$ and outputs its group index c as

$$c = \arg\min_{k \in \{1, \cdots, K\}} d(\mathbf{f}, \bar{\mathbf{f}}_k) \tag{1}$$

where d (·) denotes the distance between two vectors and c the index of identified language.

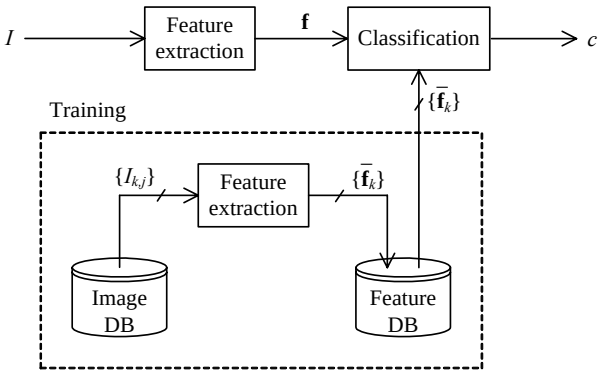

Fig. 1 Block diagram of a typical texture feature-based language identification

B. GLCM and Haralick Features

Let us denote the joint probability mass function (PMF) between a pair of two pixels, whose gray levels are individually i and j and whose distance and angle are r and θ, as $P(i, j; r, \theta)$. Then the estimate of the joint PMF $\hat{P}(i, j; r, \theta)$ is often called GLCM or GLCP [16]. The averaged GLCM $\hat{P}(i, j)$ is obtained averaging $\hat{P}(i, j; r, \theta)$ over all possible (r, θ) as follows:

$$\hat{P}(i, j) = E[\hat{P}(i, j; r, \theta)] \tag{2}$$

where $E[\cdot]$ denotes the expectation operator. Haralick *et al.* [16] suggested 28 features such as entropy, variance, moment, and correlation etc. and Holmes *et al.* [17] used the average estimate of the joint PMF $\hat{P}(i, j)$ in (2) to extract some of Haralick features.

C. BDIP and BVLC in Spatial Domain

In order to represent the BDIP in the spatial domain, let us denote I_p be the brightness of an image I at a pixel $p = (x, y)$ and R_p a local region centered at p. Then the BDIP for I at p is defined as [10]

$$D_p = \frac{E[\hat{I}_p - I_{p+q} \mid q \in R_p]}{\hat{I}_p} \tag{3}$$

where $E[\cdot|\cdot]$ denotes the conditional expectation, and $\hat{I}_p$ represents the maximum brightness in R_p, which is expressed as

$$\hat{I}_p = \max_{q \in R_p} I_{p+q}. \tag{4}$$

In (3), the numerator of D_p denotes the average of gradients in the local region R_p. So one can see that the BDIP denotes an average of normalized local gradients. For stabilization, $\hat{I}_p$ is clipped as $\hat{I}_p = \max(\hat{I}_p, \delta_D)$.

The BVLC for I at p is defined as [10]

$$V_p = \max_{d \in O} \rho_p(d) - \min_{d \in O} \rho_p(d) \tag{5}$$

where $\rho_p(d)$ denotes the local correlation coefficient of I at p according to orientation d as follows:

$$\rho_p(d) = \frac{E[(I_{p+d+q} - \mu_{I_{p+d}})(I_{p+q} - \mu_{I_p}) \mid q \in R_p]}{\sigma_{I_{p+d}} \sigma_{I_p}}, \quad d \in O \tag{6}$$

where μ_{I_p} and σ_{I_p} denote the mean and standard deviation of I in a local region R_p. The symbols $\mu_{I_{p+d}}$ and $\sigma_{I_{p+d}}$ denote the mean and standard deviation in a local region R_{p+d} centered at $p+d$, the notation O the orientation set of $O = \{(-m,0),(m,0),(0,-m),(0,m)\}$. From (5) and (6), one can see that the BVLC denotes a maximum deviation of normalized local correlations according to orientations. The $\sigma_{I_{p+d}}$ and σ_{I_p} are also clipped with δ_V for stabilization

D.BDIP and BVLC Subbands in Wavelet Domain

As for wavelet transform, applying one-dimensional low-pass and high-pass filters separately in horizontal and vertical directions to an image I produces a set of one smooth subband and three detail subbands at the first scale, denoted as $\{W^{(1,LL)}, W^{(1,LH)}, W^{(1,HL)}, W^{(1,HH)}\}$, which are usually down-sampled and shown in Fig. 2. In a similar way, $W^{(1,LL)}$ is further decomposed into another set of four subbands at the second scale, $\{W^{(2,LL)}, W^{(2,LH)}, W^{(2,HL)}, W^{(2,HH)}\}$. Executing these procedures up to the Lth scale yields a set of wavelet subbands $WI = \{W^{(L,LL)}, \{W^{(l,b)}\}, l = 1, \cdots, L, b \in \{LH, HL, HH\}\}$.

The BDIP subband $WD_p^{(l,b)}$ is defined as [10]

$$WD_p^{(l,b)} = \frac{E[\hat{W}_p^{(l,b)} - W_{p+q}^{(l,b)} \mid q \in R_p]}{\hat{W}_p^{(l,LL)}}, \quad l = 1, \cdots, L,\ b \in \{LL, LH, HL, HH\} \tag{7}$$

where $\hat{W}_p^{(l,b)}$ denotes the maximum in a local region R_p. The BVLC subband $WV_p^{(l,b)}$ is also defined as [10]

$$WV_p^{(l,b)} = \max_{d \in O} \rho_p^{(l,b)}(d) - \min_{d \in O} \rho_p^{(l,b)}(d), \quad l = 1, \cdots, L,\ b \in \{LL, LH, HL, HH\} \tag{8}$$

where $\rho_p^{(l,b)}(d)$ denotes the local correlation coefficients of $W^{(l,b)}$ at p according to orientation d. It is given as

$$\rho_p^{(l,b)}(d) = \frac{E[(W_{p+d+q}^{(l,b)} - \mu_{W_{p+d}^{(l,b)}})(W_{p+q}^{(l,b)} - \mu_{W_p^{(l,b)}}) \mid q \in R_p]}{\sigma_{W_{p+d}^{(l,b)}} \sigma_{W_p^{(l,b)}}}, \quad d \in O \tag{9}$$

where $\mu_{W_p^{(l,b)}}$ and $\sigma_{W_p^{(l,b)}}$ denote the local mean and standard deviation of $W^{(l,b)}$ at p, respectively. Similarly, $\mu_{W_{p+d}^{(l,b)}}$ and $\sigma_{W_{p+d}^{(l,b)}}$ denote the local mean and standard deviation of $W^{(l,b)}$ at $p+d$, respectively.

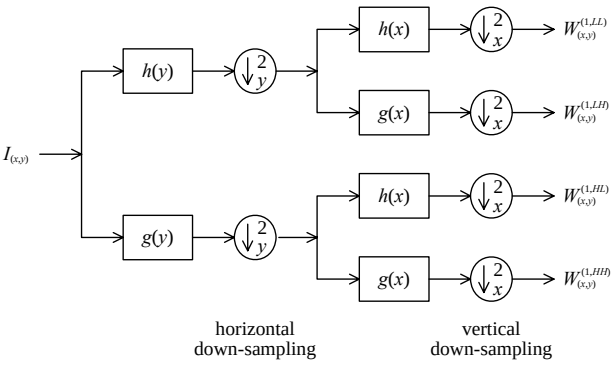

Fig. 2 Block diagram of wavelet transform up to the first scale

E. FFT Features

For 2D FFT computation, an image I is partitioned into blocks of $B \times B$ size. Let us denote $I^{(m,n)}$ be a $B \times B$ image block of block position (m, n). An FFT block $FI^{(m,n)}$ is obtained by applying 2D FFT [18] to $I^{(m,n)}$. Considering the conjugate symmetry property and the frequency distribution of 2D FFT, some significant coefficients in $FI^{(m,n)}$ are usually selected using a zonal selection. The magnitudes of selected FFT coefficients averaged over the entire image I are used as features, which is represented as follows:

$$FM = \frac{1}{N_B} \sum_{(m,n)} |FI^{(m,n)}| \quad (10)$$

where FM denotes a block of the averaged FFT magnitudes and N_B the number of blocks in the image I.

III. Proposed Texture Feature-Based Language Identification

Fig. 3 shows the block diagram of the proposed texture feature-based language identification. When a test image for language identification is input, its wavelet subbands are first obtained by wavelet transform, and the detail subbands is then denoised by Donoho's soft-thresholding as in [9]. The BDIP and BVLC operators are next performed on the smooth subband and the denoised detail subbands. The two moments of global mean and standard deviation for each BDIP and BVLC subband are computed and formed as feature vectors $\mathbf{f}_{WD}$ and $\mathbf{f}_{WV}$. The test image is also partitioned into blocks, FFT blocks for each block are obtained by 2D FFT, and some significant coefficients in each FFT block are zonally selected. Magnitude operator is applied to the selected FFT coefficients. A moment of mean for each FFT magnitude are calculated and formed as feature vector $\mathbf{f}_{FM}$. Finally, the three feature vectors are fused into a feature vector $\mathbf{f} = [\mathbf{f}_{WD}, \mathbf{f}_{WV}, \mathbf{f}_{FM}]$. In classification, a stabilized Bayesian classifier searches the most similar vector to the test feature vector, which is stored in the feature DB, and outputs the searched index.

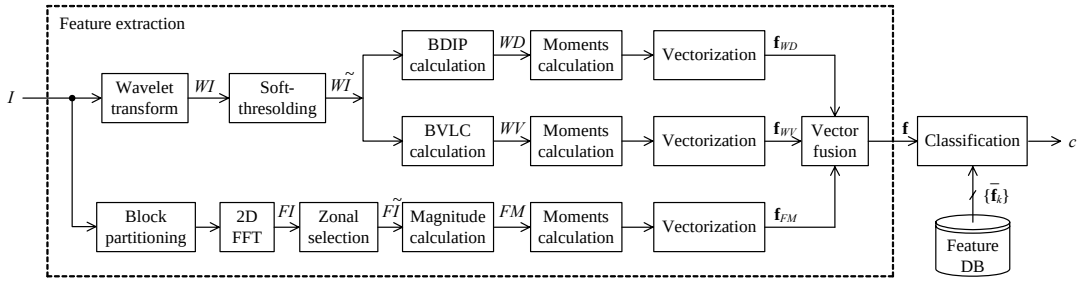

Fig. 3 Block diagram of the proposed texture feature-based language identification

A. Soft-Thresholding in Detail Subbands

Donoho's soft-thresholding for a detail subband $W^{(l,b)}$ at p can be represented as

$$\tilde{W}_p^{(l,b)} = \begin{cases} \mathrm{sgn}(W_p^{(l,b)})(|W_p^{(l,b)}| - \delta_l), & |W_p^{(l,b)}| > \delta_l \\ 0, & \text{otherwise,} \end{cases}$$
$$l = 1, \cdots, L,\ b \in \{LH, HL, HH\} \quad (11)$$

where sgn(·) denotes the sign of the quantity and δ_l the threshold at the lth level. We denote the set of wavelet subbands with the soft-thresholding in (11) as $\tilde{WI} = \{W^{(L,LL)}, \{\tilde{W}^{(l,b)}\}, l = 1, \cdots, L,\ b \in \{LH, HL, HH\}\}$. We decide the threshold δ_l proportional to the standard deviation of noise $\hat{\sigma}_n$ estimated from the blank lines of the image I as $\delta_l = \alpha_l \cdot \hat{\sigma}_n$.

B. Formation of Feature Vectors

The feature vector $\mathbf{f}_{WD}$ is formed from the two moments, global mean $\mu_{WD^{(l,b)}}$ and standard deviation $\sigma_{WD^{(l,b)}}$ of each BDIP subband $WD^{(l,b)}$ as follows:

$$\mathbf{f}_{WD} = [\mathbf{f}_{WD^{(l,b)}} \mid l = 1, \cdots, L,\ b \in \{LL, LH, HL, HH\}] \quad (12)$$

$$\mathbf{f}_{WD^{(l,b)}} = [\mu_{WD^{(l,b)}}, \sigma_{WD^{(l,b)}}]. \quad (13)$$

In a similar way, $\mathbf{f}_{WV}$ is also produced by computing the global mean $\mu_{WV^{(l,b)}}$ and standard deviation $\sigma_{WV^{(l,b)}}$, by gathering the corresponding vector $\mathbf{f}_{WV^{(l,b)}}$ over all l and b's. In this procedure, BVLC is computed in the eight orientations of

$$O = \{(u,v) \mid u,v \in \{-m, 0, m\}, (u,v) \neq (0,0)\}. \quad (14)$$

In order to describe the feature vector of FFT magnitude, let us denote FM_b be a coefficient in the averaged FFT magnitude block FM represented in (10) at a position $b = (s, t)$. The $\mathbf{f}_{FM}$ is generated by collecting the FM_b over all b's in a selection zone as follows:

$$\mathbf{f}_{FM} = [FM_b \mid b \in Z] \quad (15)$$

where Z denotes a set of positions in the selection zone. Finally, the $\mathbf{f}_{WD}$, $\mathbf{f}_{WV}$, $\mathbf{f}_{FM}$ are simply fused so that a feature vector $\mathbf{f}$ is obtained as $\mathbf{f} = [\mathbf{f}_{WD}, \mathbf{f}_{WV}, \mathbf{f}_{FM}]$.

C. Stabilized Bayesian Classifier

We adopt a stabilized Bayesian classifier in which the distance between the feature vector $\mathbf{f}$ of a test image I and the representative feature vector $\bar{\mathbf{f}}_k$ of the kth language in the feature DB is defined as [9]

$$d(\mathbf{f},\bar{\mathbf{f}}_k)=(\mathbf{f}-\bar{\mathbf{f}}_k)\,\tilde{\mathbf{C}}_k^{-1}(\mathbf{f}-\bar{\mathbf{f}}_k)^T+\ln|\tilde{\mathbf{C}}_k| \quad (16)$$

where $\tilde{\mathbf{C}}_k$ denotes the covariance matrix of $\bar{\mathbf{f}}_k$ with variance thresholding. Equation (16) can be derived when $\mathbf{f}$ has a Gaussian distribution. The diagonal elements of $\tilde{\mathbf{C}}_k$ are thresholded as

$$\tilde{C}_k(i,j)=\begin{cases}C_k(i,j), & i\neq j \text{ or } C_k(i,i)\geq\delta\\ \delta, & C_k(i,i)<\delta\end{cases} \quad (17)$$

where $C_k(i,j)$ denotes an element of the covariance matrix $\mathbf{C}_k$ and δ a threshold.

IV. Experimental Results

The performance of our language identification method was evaluated on a document image DB written in the ten languages of English, Greek, Russian, Hebrew, Persian, Hindi, Thai, Chinese, Japanese, and Korean. In establishing the image DB, we scanned paper documents to acquire eight mother images for each language, each of which includes two font types with normal, skewed by 1.5 and 3.0 degrees, and down-scaled by 0.8:1 and is divided into 50 subimages of 128×128 size. Some subimages of each language were used for training images and other 200 ones for test images. Fig. 4 shows sample subimages of ten languages used in our experiments.

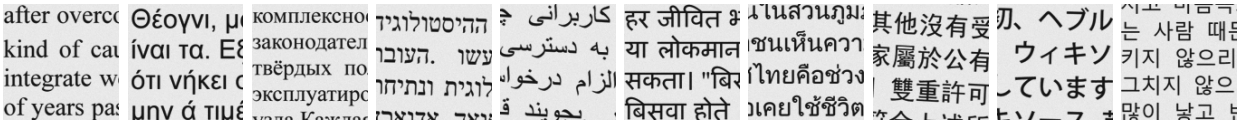

Fig. 4 Sample subimages of ten languages: English, Greek, Russian, Hebrew, Persian, Hindi, Thai, Chinese, Japanese, and Korean from left to right

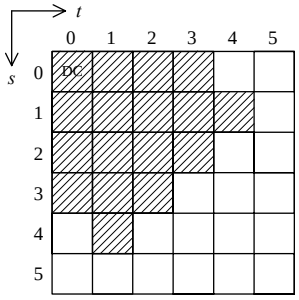

Fig. 5 Zonal selection of 6×6 FFT coefficients

We compare the performance of our method with those of a language identification using Haralick features adopted in [8] and that using wavelet-domain BDIP and BVLC features. For the extraction of texture features in the wavelet domain, we performed the wavelet transform up to the first scale (L = 1) with down-sampling using Haar filters. Odd-symmetric extension was chosen for image boundary processing. The proportional constant α_1 for Donoho's soft-thresholding was experimentally selected as α_1 = 1.75. The size of local region for the BDIP and BVLC subbands was chosen as 3×3 and the parameter m for the BVLC according to the eight orientations as m = 1. We also chose the block size of FFT as 6×6 and selected 17 significant FFT coefficients with the zonal selection as shown in Fig. 5. For the computation of FFT, we inputted a sketch image of an image I whose brightness at a pixel p is normalized with the maximum brightness in a local region R_p as

$$I'_p=\frac{I_p}{\bar{I}_p}. \quad (18)$$

In addition, we experimentally determined the clipping thresholds δ_D and δ_V for the stabilization of BDIP and BVLC as δ_D = 2 and $\delta_V^2=10^{-3}$, respectively. We decided the threshold δ for the stabilized Bayesian classifier as the value which covers ε% of feature variances sorted in descending order over all the training images.

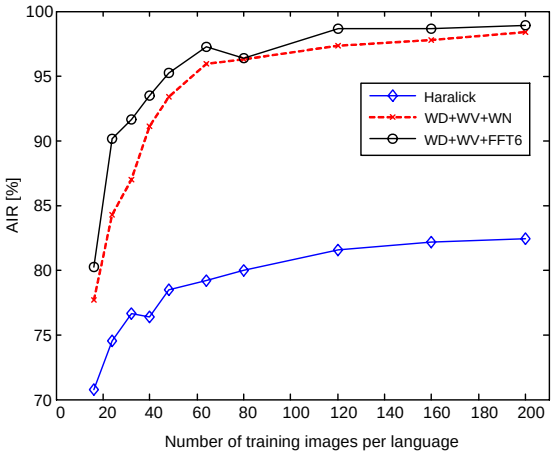

Fig. 6 AIR performances of Haralick features, WD+WV+WN features, and the proposed method according to the number of training images per language in case of 200 test images per language

The performance of language identification was measured as the average identification rate (AIR), which is defined as the ratio of the total number of test images to that of correctly identified ones. Fig. 6 shows the AIR performances of Haralick features and the two fused features according to the number of training images per language, where the number of test images per language is 200. In this figure, Haralick denotes the eight

features of energy, entropy, inertia, contrast, local homogeneity, cluster shade, cluster prominence, and information measure of correlation using GLCM, where the distances include $r=1$ and 2 and the orientations $\theta=0$, 45, 90, 135°. The WD+WV+WN represents the fusion of wavelet-domain BDIP, BVLC, and NRMA features proposed in [9] and WD+WV+FFT6 the proposed fusion of WD+WV and 6×6 block FFT features. The ε for the variance thresholding was chosen as 87~100, where the larger the number of training images, the larger the value of ε.

We can see from Fig. 6 that the Haralick and the WD+WV+WN demonstrate AIR performances of 70.75%~82.35% and 77.65%~98.35%, respectively. On the contrary, the proposed WD+WV+FFT6 gives AIRs of 80.25%~98.90% so that its performance improvement over the Haralick and the WD+WV+WN are 9.5%~18.05% and 0.15%~5.85%, respectively.

Table I shows the AIR results according to the types of texture features in case that the number of training images per languages are 200. As shown in Table I, we can see that Haralick features yield 82.45% AIR. We can also see that WD, WV, WN, FFT6 features give 84.60%, 84.25%, 76.50%, and 95.85% AIRs, respectively, and the fused features WD+WV and WD+WV+WN 94.45% and 98.35%, respectively. It is also shown that the feature WD+WV+FFT6 produces 98.90% performance so that it is 16.45% and 0.55% better than Haralick and WD+WV+WN features, respectively.

TABLE I
AIRS ACCORDING TO FEATURES IN CASE OF 200 TRAINING IMAGES

Types of texture feature	Haralick	WD	WV	WN	FFT6	WD+WV	WD+WV+WN	WD+WV+FFT6
Feature dimension	8	8	8	6	17	16	22	33
AIR [%]	82.45	84.60	84.25	76.50	95.85	94.45	98.35	98.90

Table II shows a confusion matrix for the result of WD+WV+FFT6 in Table I. One can see from this table that our method yields somewhat low error of 0.5~2.0% for English, Greek, Russian, and Chinese and 6.0% for Japanese. However, it gives no error for Hebrew, Persian, Hindi, Thai, and Korean.

TABLE II
CONFUSION MATRIX FOR THE PROPOSED METHOD IN CASE OF 200 TRAINING IMAGES, WHERE AL (IL) DENOTES ACTUAL (IDENTIFIED) LANGUAGE

AL \ IL	Eng	Gre	Rus	Heb	Per	Hin	Tha	Chn	Jap	Kor
Eng	99.0	0.5	1.5	0	0	0	0	0	0	0
Gre	0.5	99.5	0	0	0	0	0	0	0	0
Rus	0.5	0	98.5	0	0	0	0	0	0	0
Heb	0	0	0	100	0	0	0	0	0	0
Per	0	0	0	0	100	0	0	0	0.5	0
Hin	0	0	0	0	0	100	0	0	1.0	0
Tha	0	0	0	0	0	0	100	0	0	0
Chn	0	0	0	0	0	0	0	98.0	4.5	0
Jap	0	0	0	0	0	0	0	1.5	94.0	0
Kor	0	0	0	0	0	0	0	0.5	0	100
Error [%]	1.0	0.5	1.5	0	0	0	0	2.0	6.0	0

V. CONCLUSION

In this paper, a texture feature-based language identification method using wavelet-domain BDIP and BVLC features and 6×6 block FFT features has been proposed. In the propose method, the FFT features includes 17 significant FFT coefficients. For classification, the stabilized Bayesian classifier which includes variance thresholding was adopted. Experimental results have revealed that our method AIR performance of 80.25%~98.90% for the test document image DB so that its performance improves as 0.15%~5.85% compared to that of the fusion of wavelet-domain BDIP, BVLC, and NRMA so that it yields excellent performance.

REFERENCES

[1] D. Ghosh, T. Dube, and A. P. Shivaprasad, "Script recognition - a review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, Jan. 2010.
[2] J. Hochberg, L. Kerns, P. Kelly, and T. Thomas, "Automatic script identification from document images using cluster-based templates," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 2, pp. 176-181, Feb. 1997.
[3] A. L. Spitz, "Determination of the script and language content of document images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 3, pp. 235-245, Mar. 1997.
[4] L. Shijian and C. L. Tan, "Script and language identification in noisy and degraded document images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 14-24, Jan. 2008.
[5] G. S. Pearke and T. N. Tan, "Script and language identification from document images," in *Proc. IEEE Workshop on Document Image Analysis 97*, San Juan, Puerto Rico, Jun. 1997, pp. 10-17.
[6] T. N. Tan, "Rotation invariant texture features and their use in automatic script identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 20, no. 7, pp. 743-756, Jul. 1998.
[7] W. Chan and G. Coghill, "Text analysis using local energy," *Pattern Recognit.*, vol. 34, no. 12, pp. 2523-2532, Dec. 2001.
[8] A. Busch, W. W. Boles, and S. Sridharan, "Texture for script identification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1720-1732, Nov. 2005.
[9] W. S. Lee, N. C. Kim, and I. H. Jang, "Texture feature-based language identification using wavelet-domain BDIP, BVLC, and NRMA features," in *Proc. IEEE International Workshop on Machine Learning for Signal Processing 2010*, Kittilä, Finland, Aug./Sep. 2010, pp. 444-449.
[10] Y. D. Chun, S. Y. Seo, and N. C. Kim, "Image retrieval using BDIP and BVLC moments," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 9, pp. 951-957, Sep. 2003.
[11] Y. D. Chun, N. C. Kim, I. H. Jang, "Content-based image retrieval using multiresolution color and texture features," *IEEE Trans. Multimedia*, vol. 10, no. 6, pp. 1073-1084, Oct. 2008.
[12] H. J. So, M. H. Kim, and N. C. Kim, "Texture classification using wavelet-domain BDIP and BVLC features," in *Proc. 17th European Signal Processing Conf.*, Glasgow, Scotland, Aug. 2009, pp. 1117-1120.
[13] H. J. So, M. H. Kim, Y. S. Chung, and N. C. Kim, "Face detection using sketch operators and vertical symmetry," *FAQS-2006, Lecture Notes in Artificial Intelligence,* vol. 4027, pp. 541-551, Jun. 2006.
[14] T. D. Nguyen, S. H. Kim, and N. C. Kim, "An automatic body ROI determination for 3D visualization of a fetal ultrasound volume," *KES-2005, Lecture Notes in Artificial Intelligence,* vol. 3682, pp. 145-153, Sep. 2005.
[15] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform.*

Theory, vol. 41, no. 3, pp. 613-627, May 1995.
[16] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610-621, Nov. 1973.
[17] Q. A. Holmes, D. R. Neusch, and R. A. Shuchman, "Textural analysis and real-time classification of sea-ice types using digital SAR data," *IEEE Trans. Geosci. Remote Sensing,* vol. GE-22, no. 2, pp. 113-120, Mar. 1984.
[18] A. K. Jain, *Fundamentals of Digital Image Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989, ch. 5.