

Gabor, MDLC, Co-Occurrence 특성의 융합에 의한 언어 인식

장익훈[†], 김지홍^{**}

요 약

본 논문에서는 Gabor 특징과 MDLC 특징, 그리고 co-occurrence 특성의 융합에 의한 질감 특징 기반 언어 인식 방법을 제안한다. 제안된 방법에서는 먼저 시험 영상에 Gabor 변환에 이은 크기 연산자를 적용하여 Gabor 크기 영상을 얻고 그 통계치를 계산하여 결과를 벡터화한다. 이어서 MDLC 연산자를 이용하여 MDLC 영상을 얻고 역시 그 통계치를 계산하여 벡터화한다. 다음으로 시험 영상으로부터 GLCM을 계산하고 이를 이용하여 co-occurrence 특징을 계산한 다음 벡터화한다. 이들 Gabor, MDLC, co-occurrence 특징에 의한 벡터들은 벡터 융합에 의하여 특징 벡터로 사용된다. 분류 단계에서는 얼굴 인식에 주로 사용되는 WPCA를 분류기로 하여 시험 특징 벡터와 가장 유사한 학습 특징 벡터를 찾는다. 제안된 방법의 성능은 15개국 언어의 문서를 스캔하여 얻은 시험 문서 영상 DB에 대한 평균 인식률을 조사하여 알아본다. 실험 결과 제안된 방법은 시험 DB에 대하여 비교적 낮은 특징 벡터 차원으로 매우 우수한 언어 인식 성능을 보여준다.

Language Identification by Fusion of Gabor, MDLC, and Co-Occurrence Features

Ick-Hoon Jang[†], Ji-Hong Kim^{**}

ABSTRACT

In this paper, we propose a texture feature-based language identification by fusion of Gabor, MDLC (multi-lag directional local correlation), and co-occurrence features. In the proposed method, for a test image, Gabor magnitude images are first obtained by Gabor transform followed by magnitude operator. Moments for the Gabor magnitude images are then computed and vectorized. MDLC images are then obtained by MDLC operator and their moments are computed and vectorized. GLCM (gray-level co-occurrence matrix) is next calculated from the test image and co-occurrence features are computed using the GLCM, and the features are also vectorized. The three vectors of the Gabor, MDLC, and co-occurrence features are fused into a feature vector. In classification, the WPCA (whitened principal component analysis) classifier, which is usually adopted in the face identification, searches the training feature vector most similar to the test feature vector. We evaluate the performance of our method by examining averaged identification rates for a test document image DB obtained by scanning of documents with 15 languages. Experimental results show that the proposed method yields excellent language identification with rather low feature dimension for the test DB.

Key words: Language Identification(언어 인식), Texture Feature(질감 특징), Gabor Transform(Gabor 변환), MDLC, Co-Occurrence

※ 교신저자(Corresponding Author): 김지홍, 주소: 부산광역시 부산진구 엄광로 995 동의대학교 영상정보공학과 (608-743), 전화: (051) 890-2267, FAX: (051) 890-2265, E-mail: arim@deu.ac.kr
접수일: 2013년 9월 26일, 수정일: 2013년 12월 5일

완료일: 2013년 12월 18일

[†] 정희원, 경운대학교 항공전자공학과
(E-mail: ihjang@ikw.ac.kr)

^{**} 정희원, 동의대학교 영상정보공학과

1. 서 론

최근 컴퓨터와 모바일 단말기 등의 IT(정보기술) 기기와 그 관련 기술이 비약적으로 발전함에 따라 문서를 IT 기기로 저장하여 관리하는 경우가 많아지고 있다. 이러한 문서 관리 방법 중 하나로 문서 영상을 획득하여 OCR로 그 문자 정보를 인식하여 저장, 관리하는 방법이 있다. OCR로 문자를 인식하는 경우에는 여러 OCR 엔진 중에서 인식하고자 하는 언어에 해당하는 엔진을 사용자가 직접 선택하거나 IT 기기에서 언어를 인식하여 자동으로 선택할 수 있다. 세계가 글로벌화되면서 여러 언어로 된 문서를 접할 기회가 많아지고 문서의 분량도 매우 방대해지고 있어서 후자의 방법이 더 효율적인 것이 되고 있다[1].

문서 영상에서의 문자 인식의 전단계로 사용하기 위한 언어 인식에 관한 기존의 방법들은 언어 별로 문자가 서로 다른 구조를 갖는 점에 근거한 문자 화소의 통계적 특징(statistical feature)을 이용하는 방법[2-4]과 문자 영상으로부터 추출한 질감 특징(texture feature)을 이용하는 방법[5-11]이 주류를 이루어왔다.

Hochberg 등[2]은 각 언어 별로 자주 사용되는 단어의 형태를 찾아내어 클러스터링(clustering) 시킨 다음 클러스터별 대표 틀(template)을 생성하고 인식하고자 하는 언어와 가장 유사한 클러스터를 찾아내어 해당 언어를 인식된 언어로 하였다. 이 방법에서는 13개국 언어를 인식하였다. Spitz[3]는 문자의 윗방향 요면(upward concavity)의 수직적 분포, 이진화된 문자 셀(character cell)의 정규화된 화소 분포, 그리고 단어 형태의 발생 빈도를 이용하여 6개국 언어를 인식하였다. Shijian과 Tan[4]은 문서 영상을 문자 또는 단어의 형태와 발생 빈도를 특징지우는 문서 벡터(document vector)로 만든 다음 가장 유사한 벡터를 찾는 방식으로 언어를 인식하는 방법을 제안하였다. 이 방법은 6개국 언어에 대한 실험에서 잡음과 왜곡이 있는 문서 영상에서도 우수한 성능을 보이는 것으로 알려졌다.

질감 특징을 이용한 언어 인식 방법에서는 각 언어의 문자가 시각적으로 다르게 나타난다는 사실에 근거하여 문서 영상으로부터 문자 블록을 얻은 다음 이를 질감 패턴으로 보고 그 특징을 추출하여 질감 분류(texture classification) 방식으로 언어를 인식한

다[5-11]. 그러므로 질감 특징을 이용한 방법에서 가장 중요한 것은 각 언어에 해당하는 문자를 표현할 수 있는 특징을 추출하는 것이라고 할 수 있다. 언어 인식을 위한 기존의 질감 특징으로는 Gabor, GLCM(gray-level co-occurrence matrix)에 의한 co-occurrence, BDIP(block difference of inverse probabilities)와 BVLC(block variance of local correlation coefficients), MDLC(multi-lag directional local correlation) 등이 있다.

Gabor 필터는 가우시안(Gaussian) 함수를 복소지수 함수로 변조한 함수를 모 함수(mother function)로 하여 이를 스케일링(scaling)하고 회전하여 얻어진다[12]. 이는 방향성과 주파수 선택성을 가지며 영상에서 국부적인 특정 주파수 성분과 방향성 정보를 효과적으로 표현할 수 있다. Gabor 필터는 질감 분류를 위한 특징 추출로 많이 사용되어 왔으며 [13], 언어 인식을 위한 질감 특징 추출로 가장 많이 사용된 것으로 알려져있다[1]. Pearke와 Tan[5]은 기존의 질감 분류에서 많이 사용되는 Gabor 필터를 이용하여 문서 영상에서 질감 특징을 추출하였다. Tan[6]은 Gabor 필터를 이용하여 회전(rotation)에도 변하지 않는 질감 특징을 추출하였고, Chan과 Coghil[7] 또한 Gabor 필터를 이용하였다. [5]에서는 7개국 언어, [6]에서는 6개국 언어, [7]에서는 16개국 언어를 각각 인식하였다.

GLCM[14]은 두 화소간의 결합 통계치로서 질감 분류, 얼굴 인식, 언어 인식 등을 위한 특징 추출에 많이 이용되어 왔다. [5]에서는 GLCM의 일부 원소들로 구성되는 co-occurrence 특징을 언어 인식에 사용하였다. Busch 등[8]은 Wouver 등[15]이 질감 분류(texture classification)에 적용한 8개의 통계치에 의한 co-occurrence 특징 중에서 7개를 도입하고 여기에 1개의 특징을 추가하여 구성된 8개의 특징을 언어 인식에 적용한 바 있다. 이 특징들은 웨이브렛 영역에서 로그 연산자와 함께 사용하여 추출되었으므로 WLC(wavelet log co-occurrence)라고 명명되었다. Hiremath와 Shivashankar[9]는 GLCM에 의한 3개의 통계치를 co-occurrence 특징으로 선택하고 이를 웨이브렛 영역에서 추출하여 언어를 인식하고자 하였다. [8]과 [9]에서는 8개국 언어를 각각 인식하였다.

BDIP와 BVLC는 영상의 국부적인 밝기 변화와

질감의 평활성에 관련된 질감 특징을 각각 추출한다 [10]. 이들 연산자는 언어 인식 뿐만 아니라 질감 분류와 영상 검색, 얼굴 인식, 관심영역 결정(ROI determination)에도 적용되어 우수한 성능을 보여주었다. MDLC는 영상의 국부영역에서 방향과 거리차(lag)에 따른 상관계수를 얻는 연산자이며 국부 영역에서 밝기의 유사성을 효과적으로 표현한다[11]. MDLC는 얼굴 인식을 위하여 제안되었으며 언어 인식에도 우수한 성능을 보여주었다. 언어 인식에 대한 전반적인 개요는 [1]을 참조하면 된다.

질감 특징은 여러 가지 특징을 융합하여 사용할 수 있다. 이러한 특징 융합에 의한 질감 분류 방법은 사용되는 특징 벡터의 차원이 커지면 계산량이 증가하고 경우에 따라서는 성능이 오히려 감소할 수 있다. 따라서 질감 분류 방법에 사용되는 방법을 문자 영상에 적용할 때는 특징 벡터의 차원을 고려하여 가능하면 낮은 차원의 특징으로 높은 인식률을 보이도록 하여야 한다. 그러기 위해서는 문자 영상의 특성을 잘 표현하면서 융합되었을 때 특징 벡터가 상호 보완하면서 시너지 효과를 얻을 수 있고 차원은 낮은 질감 특징들을 선택하여야 한다. 본 저자 등[10]은 Gabor 특징과 웨이블릿 영역의 BDIP와 BVLC 특징을 융합하여 언어 인식에 적용한 바 있다. 그리고 Gabor 특징과 MDLC 특징을 융합한 언어 인식[11]도 제안하여 [10]에서 제안한 방법보다 성능이 우수한 것을 실험으로 보인 바 있다. [10]과 [11]에서는 각각 10개국 언어를 인식하였다.

본 논문에서는 Gabor 특징과 MDLC 특징에 영상 내의 화소 간의 통계적 특성 8가지를 표현하여 이들 특징과는 그 특성이 다르고 벡터 차원도 낮은 co-occurrence 특징을 융합한 질감 특징 기반 언어 인식 방법을 제안한다. 제안된 방법에서는 먼저 시험 영상에 질감 분류에 우수한 성능을 보이는 것으로 알려진 Manjunath와 Ma[12]가 제안한 Gabor 필터에 이은 크기 연산자를 적용하여 국부적인 특징 주파수 성분과 방향성 정보와 관련된 질감 특징을 추출한다. 그리고 시험 영상에 MDLC 연산자를 적용하여 국부 영역에서 밝기의 유사성에 관련된 질감 특징을 추출한다. 이어서 추출된 Gabor 크기 영상과 MDLC 영상 특징의 전역 평균과 표준 편차를 구하고 이들을 각 특징별로 모아서 벡터화한다. 이어서 co-occurrence 특징을 추출하고 이들을 벡터화하여 Gabor와 MDLC

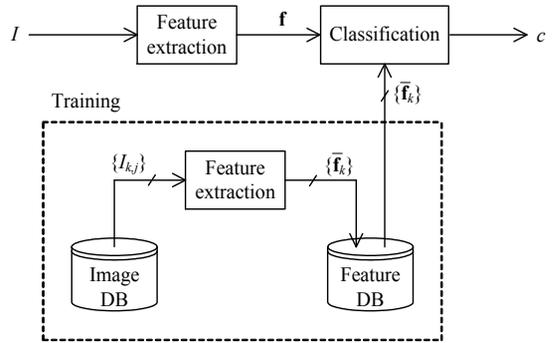


그림 1. 기존의 질감 특징을 이용한 언어 인식의 블록도

특징에 의한 벡터와 융합하여 얻어지는 벡터를 특징 벡터로 사용한다. 분류기로는 얼굴 인식에 주로 사용되는 것으로 알려진 WPCA(whitened principal component analysis)[16]를 사용한다.

2. 질감 특징을 이용한 언어 인식과 질감 특징

2.1 질감 특징을 이용한 언어 인식

그림 1은 기존의 질감 특징을 이용한 문자 영상에서의 언어 인식의 블록도를 보여준다. 편의상 학습(training)을 위한 영상 DB에는 K 개의 언어를 가지면서 각 언어별로는 J 개의 영상이 포함된 것으로 하자. 먼저 학습 단계에서는 영상 DB에 저장되어 있는 각 영상 $I_{k,j}$, $k=1, \dots, K$, $j=1, \dots, J$ 로부터 질감 특징을 추출한 다음 이들로 부터 특징 벡터 $f_{k,j}$ 를 생성한다. 기존의 학습에서는 주로 특징 벡터 $f_{k,j}$ 를 평균하여 각 언어에 대한 대표 특징 벡터 집합 $\{f_k | k=1, \dots, K\}$ 를 얻는다. 그리고 필요하다면 특징 벡터에 대한 공분산 행렬 집합 $\{C_k | k=1, \dots, K\}$ 를 계산하고 통계치를 특징 DB에 저장한다. 시험(test) 단계에서는 언어 인식을 위한 시험 영상 I 가 들어오면 이로부터 특징 벡터 f 를 추출하고 분류기에서 다음과 같이 DB에 저장되어 있는 특징 벡터 $\{f_k | k=1, \dots, K\}$ 중에서 f 와 가장 유사한 벡터 \bar{f}_c 를 찾아 그 인자(index) c 를 출력한다.

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

여기서 $d(\cdot)$ 는 두 벡터간의 거리(distance)를 나타낸다. 이때 c 는 인식된 언어의 인자가 된다. 그리고

계산된 공분산 행렬 C_k 는 (1)식의 거리 $d(\cdot)$ 의 측정
에 사용될 수가 있다.

2.2 Gabor 특징

2차원 Gabor 필터에서는 다음과 같이 표현되는 가
우시안 함수를 복소 지수 함수로 변조한 함수를 모
함수로 사용한다[12].

$$g(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left[-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2}\right) + 2\pi j Wx\right] \quad (2)$$

여기서 σ_x 와 σ_y 는 가우시안 함수의 표준편차를 나
타내며 모 함수의 주파수 대역을 정하는 상수로 사용
된다. 그리고 W 는 변조 주파수를 나타내며 모 함수
의 중심 주파수를 정하는 상수로 사용된다. Gabor
필터는 다음과 같이 (2)식의 모 함수 $g(x,y)$ 를 스케일
링하고 회전하여 얻어진다.

$$g_{m,n}(x,y) = a^{-2m} g(x',y') \quad (3)$$

여기서 $a(a>1)$ 는 스케일링을 위한 상수, $m(m = 0, \dots, M-1)$ 은 스케일 인자, 그리고 $n(n = 0, \dots, N-1)$ 은 회전 인자를 각각 나타낸다. 그리고 $x' = a^{-m}(x \cos \theta_n + y \sin \theta_n)$, $y' = a^{-m}(-x \sin \theta_n + y \cos \theta_n)$, $\theta_n = n\pi/N$ 이다. (2)식과 (3)식에서 보면 Gabor 필터 $g_{m,n}(x,y)$ 의 주파수 대역은 가우시안 함수의 표준편차 σ_x 와 σ_y , 그리고 스케일 인자 m 에 의하여 결정되고, 중심 주파수는 변조 주파수 W 에 의하여 결정되며, 방향성은 회전 인자 n 에 의하여 결정된다는 것을 알 수 있다. 영상 I 에 대한 Gabor 변환은 (3)식의 Gabor 필터 $g_{m,n}(x,y)$ 를 I 에 적용하여 얻어지며, Gabor 특징은 Gabor 변환 영상 $GI = \{G^{(m,n)}, m = 0, \dots, M-1, n = 0, \dots, N-1\}$ 에 크기 연산자(magnitude operator)를 적용하여 얻은 영상 G 로부터 추출된다.

2.3 MDLC 특징

공간 영역의 MDLC를 표현하기 위하여 영상 I 의
화소 $p = (x,y)$ 에서의 밝기 값을 I_p 로 하고, 화소 p 를
중심으로 하는 국부 영역을 R_p 로 하자. 그러면 R_p
내에서의 국부 평균과 분산은 다음과 같이 표현된다.

$$\mu_p = E[I_q | q \in R_p] \quad (4)$$

$$\sigma_p^2 = E[(I_q - \mu_p)^2 | q \in R_p] \quad (5)$$

표 1. GLCM을 이용한 8개의 co-occurrence 특징(8)

Co-Occurrence Features	Description
Energy	$\sum_{(i,j)} \hat{P}^2(i,j)$
Entropy	$-\sum_{(i,j)} \hat{P}(i,j) \log \hat{P}(i,j)$
Inertia	$\sum_{(i,j)} (i-j)^2 \hat{P}(i,j)$
Contrast	$\sum_{(i,j)} \hat{P}(i,j) i-j ^k, k \in Z$
Local homogeneity	$\sum_{(i,j)} \frac{1}{1+(i-j)^2} \hat{P}(i,j)$
Cluster shade	$\sum_{(i,j)} (i-M_x + j-M_y)^3 \hat{P}(i,j)$
Cluster prominence	$\sum_{(i,j)} (i-M_x + j-M_y)^4 \hat{P}(i,j)$
Information measure of correlation	$\frac{-\sum_{(i,j)} \hat{P}(i,j) \log \hat{P}(i,j) - H_{xy}}{\max(H_x, H_y)}$

where $M_x = \sum_i \sum_j i \hat{P}(i,j)$ and $M_y = \sum_i \sum_j j \hat{P}(i,j)$
 $H_{xy} = -\sum_{(i,j)} \hat{P}(i,j) \log(\sum_j \hat{P}(i,j) \cdot \sum_i \hat{P}(i,j))$
 $H_x = -\sum_i \{\sum_j \hat{P}(i,j) \cdot \log \sum_j \hat{P}(i,j)\}$
 $H_y = -\sum_j \{\sum_i \hat{P}(i,j) \cdot \log \sum_i \hat{P}(i,j)\}$

여기서 $E[\cdot | \cdot]$ 는 조건 기대치 연산자
(conditional expectation operator)를 나타낸다. (4)
식과 (5)식의 두 통계치를 이용하여 화소 p 와 p 에서
거리차가 r 이고 방향은 단위 방향 벡터(unit direc-
tion vector) d 의 방향인 화소와의 방향성 국부 상관
(DLC: directional local correlation)을 다음과 같이
표현할 수 있다.

$$\rho_p^{(r,d)} = \frac{E[(I_{q+rd} - \mu_{p+rd})(I_q - \mu_p) | q \in R_p]}{\sigma_{p+rd} \sigma_p} \quad (6)$$

여기서 r 은 화소 p 와 $p+rd$ 사이의 Manhattan 거
리를 나타낸다. 그리고 μ_{p+rd} 와 σ_{p+rd} 는 국부 영역
 R_{p+rd} 내에서의 국부 평균과 표준 편차를 각각 나타
낸다. 국부 영역으로는 3x3 크기의 창(window)이 주
로 사용되고, 단위 방향 벡터로는 그림 2(a)와 같은
8방향 벡터가 주로 사용된다.

MDLC는 (6)식의 DLC를 이용하여 다음과 같이
표현된다[11].

$$\{\rho_p^{(r,d)}\}, r \in \{2^l | l = 0, \dots, L-1\}, \quad (7)$$

$$d \in \{d_n | n = 0, \dots, N-1\}$$

여기서 l 과 n 은 거리차와 방향 인자를 각각 나타
낸다. 그림 2(b)에서는 거리차의 수 $L=3$ 과 방향 벡
터의 수 $N=8$ 일 때 화소 p 를 중심으로 하는 여러

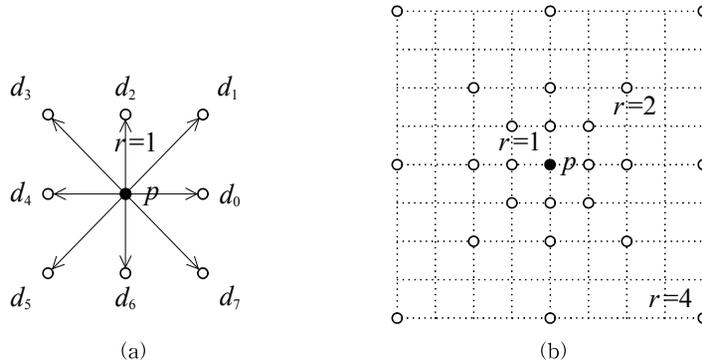


그림 2. 단위 벡터와 여러 거리차에 따른 방향성 화소의 구성 예: (a) 8방향 단위 벡터, (b) $L=3, N=8$ 인 경우의 여러 거리차에 따른 방향성 화소의 구성

거리차에 따른 방향성 화소들을 보여준다. 식 (6)과 (7)에서 보면 MDLC는 거리차와 방향에 따른 정규화된 국부 상관을 나타낸다는 것을 알 수 있다. MDLC 특징은 MDLC 영상 $M = \{\rho_p^{(r,d)}\}$ 로부터 추출된다.

2.4 GLCM과 Co-Occurrence 특징

밝기 값이 각각 i 와 j 이면서 거리가 r 이고 각도가 θ 인 임의의 두 화소 간의 결합 확률 질량 함수(joint probability mass function)를 $P(i, j; r, \theta)$ 라 하자. 그러면 GLCM은 결합 확률 질량 함수의 추정치 $\hat{P}(i, j; r, \theta)$ 를 일컬으며, GLCP(GLC probability)라고도 한다 [14]. 평균 GLCM $\hat{P}(i, j)$ 는 다음과 같이 추정된 결합 확률 질량 함수 $\hat{P}(i, j; r, \theta)$ 를 모든 가능한 (r, θ) 에 대하여 평균하여 얻어진다.

$$\hat{P}(i, j) = E[\hat{P}(i, j; r, \theta)] \tag{8}$$

[14]에서는 GLCM으로부터 얻게 되는 엔트로피

(entropy), 분산(variance), 모멘트(moment), 상관(correlation) 등의 28가지 질감 특징을 제안하였다. 그리고 Holmes 등[17]도 GLCM을 이용하여 질감 특징을 추출하였다. GLCM을 이용하여 얻어지는 특징은 co-occurrence 특징이라고도 한다. [8]에서는 표 1에서 보여주는 에너지, 엔트로피, 관성(inertia), 대비(contrast), 국부 균질성(local homogeneity), 클러스터 색조(cluster shade), 클러스터 중요성(cluster prominence), 상관의 정보 척도(information measure of correlation)의 8개의 통계치를 co-occurrence 특징으로 사용하였다. 본 논문에서는 표 1의 co-occurrence 특징을 도입한다.

3. 제안된 질감 특징을 이용한 언어 인식

그림 3은 제안된 질감 특징을 이용한 문서 영상에서의 언어 인식의 블록도를 보여준다. 언어 인식을 위한 시험 영상 I 가 들어오면 먼저 Gabor 변환을 수

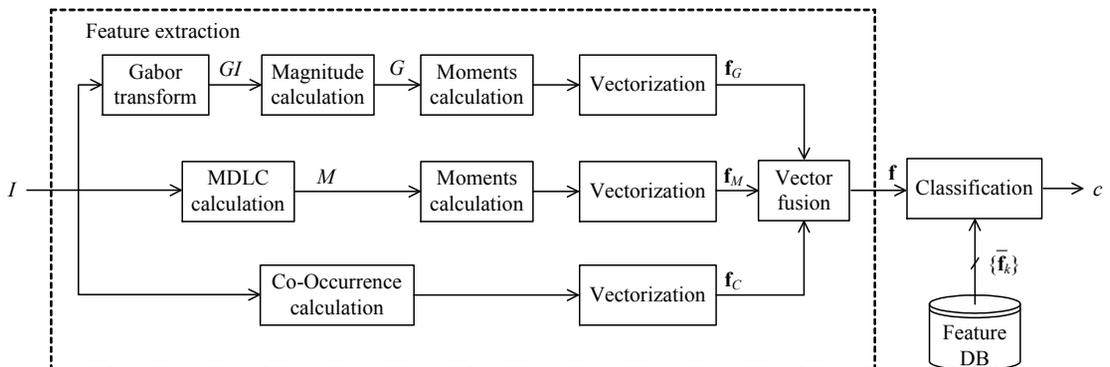


그림 3. 제안된 언어 인식의 블록도

행하여 Gabor 영상 GI 를 얻고, GI 에 크기 연산자를 적용하여 얻은 영상 G 로부터 통계치를 계산하여 그 결과를 특징 벡터 \mathbf{f}_G 로 만든다. 다음으로 영상 I 에 대하여 MDLC 연산자를 적용하여 MDLC 영상 M 을 얻고 이들로부터 통계치를 계산하여 그 결과를 특징 벡터 \mathbf{f}_M 로 만든다. 이어서 I 로부터 co-occurrence 특징을 계산하여 특징 벡터 \mathbf{f}_C 로 만든다. 이들 특징 벡터 $\mathbf{f}_G, \mathbf{f}_M, \mathbf{f}_C$ 들은 벡터 융합을 거쳐 최종 특징 벡터 \mathbf{f} 를 생성한다. 분류기에서는 특징 DB에 저장되어 있는 대표 특징 벡터 $\{\bar{\mathbf{f}}_k | k=1, \dots, K\}$ 중에서 \mathbf{f} 와 가장 유사한 벡터 $\bar{\mathbf{f}}_c$ 를 찾아 그 인자 c 를 출력한다.

3.1 특징 벡터의 생성

특징 벡터 \mathbf{f}_G 를 생성하기 위하여 먼저 m 번째 스케일 n 번째 회전 방향에서의 Gabor 크기 영상 $|G^{(m,n)}|$ 으로부터 전역 평균 $\mu_{G^{(m,n)}}$ 과 전역 표준 편차 $\sigma_{G^{(m,n)}}$ 을 계산하고 그 결과를 다음과 같이 벡터화하여 $\mathbf{f}_{G^{(m,n)}}$ 을 얻는다.

$$\mathbf{f}_{G^{(m,n)}} = [\mu_{G^{(m,n)}}, \sigma_{G^{(m,n)}}] \quad (9)$$

이어서 다음과 같이 모든 m 과 n 에 대하여 얻어지는 벡터를 모아서 \mathbf{f}_G 를 생성한다.

$$\mathbf{f}_G = [\mathbf{f}_{G^{(m,n)}} | m=0, \dots, M-1, n=0, \dots, N-1] \quad (10)$$

특징 벡터 \mathbf{f}_M 을 생성하기 위해서는 거리차가 r 이고 방향이 d 벡터 방향인 MDLC 영상 $M^{(r,d)}$ 로부터 전역 평균 $\mu_{M^{(r,d)}}$ 와 전역 표준 편차 $\sigma_{M^{(r,d)}}$ 를 계산하고 그 결과를 다음과 같이 벡터화하여 $\mathbf{f}_{M^{(r,d)}}$ 를 얻는다.

$$\mathbf{f}_{M^{(r,d)}} = [\mu_{M^{(r,d)}}, \sigma_{M^{(r,d)}}] \quad (11)$$

그리고는 다음과 같이 모든 r 과 d 에 대하여 얻어지는 벡터를 모아서 \mathbf{f}_M 을 생성한다.

$$\mathbf{f}_M = [\mathbf{f}_{M^{(r,d)}} | \forall (r, d)], \quad r \in \{2^l | l=0, \dots, L-1\}, \\ d \in \{d_n | n=0, \dots, N-1\} \quad (12)$$

특징 벡터 \mathbf{f}_C 는 8개의 co-occurrence 특징을 모아서 생성된다.

3.2 벡터 융합

특징 벡터 $\mathbf{f}_G, \mathbf{f}_M, \mathbf{f}_C$ 는 그 값의 분포 범위가 서로 다를 수 있으므로 이들을 융합하되 그 값의 분포 범

위가 같도록 각각 정규화(normalization)하면 효율적인 언어 인식이 되도록 할 수 있다. 이를 위하여 먼저 특징 벡터 $\mathbf{f}_G, \mathbf{f}_M, \mathbf{f}_C$ 를 다음과 같이 모아서 융합된 특징 벡터 $\bar{\mathbf{f}}$ 를 얻는다.

$$\bar{\mathbf{f}} = [\mathbf{f}_G, \mathbf{f}_M, \mathbf{f}_C] \quad (13)$$

Gabor 크기 영상의 전역 통계치로 주어지는 \mathbf{f}_G 는 Gabor 크기 영상이 스케일과 방향별로 그 값의 분포 범위가 거의 같이 주어지므로 \mathbf{f}_G 의 원소들은 서로 비슷한 분포 범위를 갖는다. \mathbf{f}_M 도 MDLC 영상이 방향과 거리차에 따라 그 값의 분포 범위가 거의 같으므로 그 원소들이 거의 비슷한 분포 범위를 갖는다. 그러므로 \mathbf{f}_G 와 \mathbf{f}_M 은 각각 모든 원소에 대하여 동일한 값으로 정규화한다. 서로 다른 성격의 통계치로 주어지는 \mathbf{f}_C 는 원소들이 서로 다른 분포 범위를 가질 수 있다. 따라서 \mathbf{f}_C 는 각 원소에 대하여 서로 다른 값으로 정규화한다.

이러한 점을 고려하여 융합된 특징 벡터 $\bar{\mathbf{f}}$ 는 다음과 같이 특징 벡터 $\mathbf{f}_G, \mathbf{f}_M, \mathbf{f}_C$ 의 표준 편차 벡터 $\sigma_G, \sigma_M, \sigma_C$ 로 정규화한다.

$$\mathbf{f} = \Sigma^{-1} \bar{\mathbf{f}} \quad (14)$$

여기서 $\Sigma = \text{diag}[\sigma_G, \sigma_M, \sigma_C]$ 이다. 이때 σ_G 는 학습 단계에서 k 번째 언어에 대한 Gabor 특징 벡터 \mathbf{f}_{G_k} 의 표준 편차 벡터 σ_{G_k} 를 k 에 대하여 평균하여 얻은 벡터 $E[\sigma_{G_k} | k]$ 의 각 원소를 스칼라 평균한 값 σ_G 로 이루어진 벡터이다. 즉 $\sigma_G = \sigma_G \mathbf{1}$ 이다. 그리고 σ_M 도 MDLC 특징 벡터 \mathbf{f}_{M_k} 에 대하여 σ_{G_k} 와 같은 방식으로 주어지는 벡터이다. σ_C 는 σ_G, σ_M 과 달리 모든 원소가 다른 값을 갖으며, 학습단계에서 k 번째 언어에 대한 co-occurrence 특징 벡터 \mathbf{f}_{C_k} 의 표준 편차 벡터 σ_{C_k} 를 k 에 대하여 평균하여 얻은 벡터 $E[\sigma_{C_k} | k]$ 이다.

3.3 WPCA를 이용한 분류기

DB 내의 k 번째 언어에 해당하는 대표 특징 벡터로서 $J \times 1$ 차원을 갖는 $\bar{\mathbf{f}}_k$ 에 대한 WPCA는 $\bar{\mathbf{f}}_k$ 에 대한 PCA 과정과 화이트닝(whitening) 과정으로 구성된다[16]. 편의상 대표 특징 벡터 $\bar{\mathbf{f}}_k$ 의 공분산 행렬을 \mathbf{C}_k 로 표시하고, 내림차순으로 정렬된 \mathbf{C}_k 의 고유치와 고유 벡터를 각각 $\{\lambda_{k,i} | i=1, \dots, U\}$ 와 $\{\phi_{k,i} | i=1, \dots, U\}$

로 표시하자. 그러면 대표 특징 벡터 $\bar{\mathbf{f}}_k$ 에 대한 WPCA는 다음과 같이 주어진다.

$$\bar{\mathbf{y}}_k = \bar{\Lambda}_k^{-1/2} \Phi_k^T \bar{\mathbf{f}}_k \quad (15)$$

여기서 Φ_k 와 $\bar{\Lambda}_k$ 는 각각 다음과 같이 고유 벡터로 이루어진 행렬과 고유치로 이루어진 대각선 행렬을 나타낸다.

$$\Phi_k = [\phi_{k,1}, \phi_{k,2}, \dots, \phi_{k,U}] \quad (16)$$

$$\bar{\Lambda}_k = \text{diag}[\lambda_{k,1}, \lambda_{k,2}, \dots, \lambda_{k,U}, \delta, \dots, \delta] \quad (17)$$

이때 고유치는 안정화를 위하여 역치 δ 로 클리핑(clipping)되었다. 얼굴 인식에 사용되는 WPCA에서는 공분산 행렬을 모든 클래스에서 얻고 역치 보다 큰 고유치에 해당하는 고유 벡터만을 사용한다. 반면에 본 논문에서 사용하는 WPCA에서는 공분산 행렬을 각 클래스, 즉 각 언어에서 얻고 해당 언어의 모든 고유벡터를 사용한다.

본 논문에서 사용하는 WPCA를 이용한 분류기에서 시험 영상 I 의 특징 벡터 \mathbf{f} 와 k 번째 언어에 해당하는 대표 특징 벡터 $\bar{\mathbf{f}}_k$ 의 거리는 (15)식과 같은 방식으로 주어지는 \mathbf{f} 에 대한 WPCA \mathbf{y}_k^{ts} 와 (15)식의 $\bar{\mathbf{y}}_k$ 사이의 cosine 거리로서 다음과 같이 표현된다.

$$d(\mathbf{f}, \bar{\mathbf{f}}_k) = -\frac{\mathbf{y}_k^{ts} \cdot \bar{\mathbf{y}}_k}{\|\mathbf{y}_k^{ts}\| \|\bar{\mathbf{y}}_k\|} \quad (18)$$

여기서 \mathbf{y}_k^{ts} 는 다음과 같이 주어진다.

$$\mathbf{y}_k^{ts} = \bar{\Lambda}_k^{-1/2} \Phi_k^T \mathbf{f} \quad (19)$$

그리고 \cdot 는 벡터의 내적(inner product)을 나타내고, $\|\cdot\|$ 는 벡터의 크기를 나타낸다.

4. 실험 결과 및 검토

본 논문에서 제안된 언어 인식 방법의 성능을 평가하기 위하여 영어, 프랑스어, 그리스어, 베트남어, 러시아어, 몽골어, 히브리어, 페르시아어, 에티오피아어, 힌디어, 말라야람어, 태국어, 중국어, 일본어, 한국어의 15개국 언어에 대하여 문서 영상 DB를 구축하고 시뮬레이션을 수행하였다. 문서 영상 DB를 구축하기 위하여 먼저 각 언어별로 문서를 스캔하여 모 영상(mother image)을 획득하였다. 이때 각 모 영상에는 두 가지의 폰트를 가지며 각 폰트 별로 정상적인 영상, 1.5도와 3.0도 만큼 회전된 영상, 그리고 0.8:1로 스케일링 된 영상의 총 8종류의 영상을 가지도록 하였다. 각 모 영상은 128×128 크기의 영상으로 50장씩 나누어져 각 언어 별로 총 400장의 영상을 갖도록 하였다. 이때 각 언어별 400장의 영상 중에서 절반인 200장은 시험 영상으로, 나머지 절반인 200장은 학습 영상으로 사용되 실험 조건에 따라서 학습 영상의 수를 24~56장으로 변화를 주었다. 그림 4는 실험에 사용된 15개국 언어 영상의 샘플을 보여준다.

실험에서는 제안된 언어 인식 방법을 Gabor 특징, co-occurrence 특징[8], WLC 특징[8], 그리고 MDLC 특징에 의한 방법과 성능 비교를 하였다. 또한 Gabor와 co-occurrence 특징의 융합에 의한 방법과 Gabor와 MDLC 특징의 융합에 의한 방법[11]과도 성능 비교를 하였다. 실험에서 사용된 co-occurrence 특징을 위한 GLCM의 거리는 [8]에서와 같이 $r=1,2$ 이고 각도는 $\theta=0, 45, 90, 135^\circ$ 이다. Gabor 필터를 위한 (2)식과 (3)식의 상수는 $\sigma_x=0.9809$, $\sigma_y=1.2873$, $W=0.4$, $a=2.8284$, $M=3$, $N=4$ 로 하였다. MDLC를 위한 거리차의 수는 $L=3$ 으로, 방향 벡터의 수는 그림 2(a)에서 보여준 것의 절반인 $N=4$ 로, 국부 창의 크기는 3×3 으로 하였다. WPCA에서의 고유치에 대한 역치

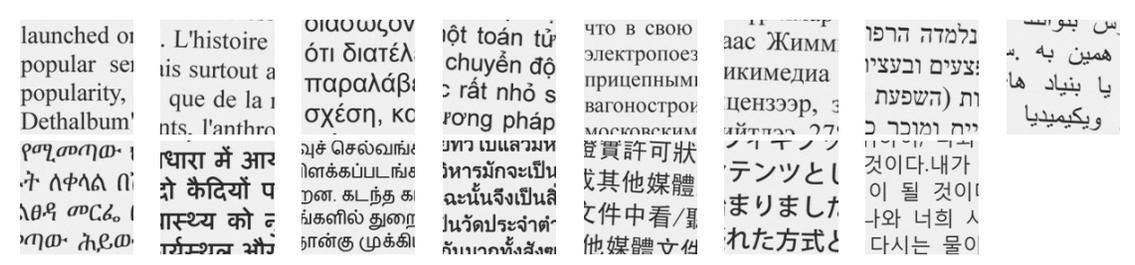


그림 4. 실험에 사용된 15개국 언어 영상의 샘플: 왼쪽 위부터 오른쪽 아래 순서로 영어, 프랑스어, 그리스어, 베트남어, 러시아어, 몽골어, 히브리어, 페르시아어, 에티오피아어, 힌디어, 말라야람어, 태국어, 중국어, 일본어, 한국어

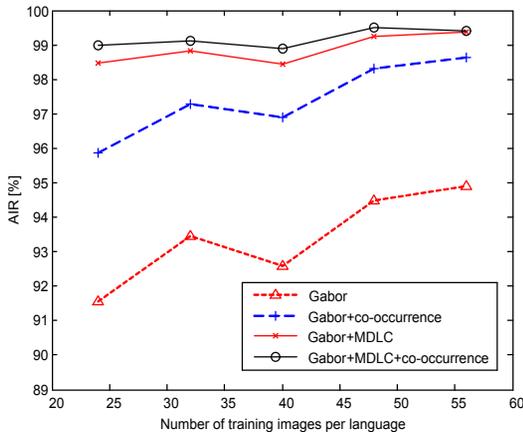


그림 5. 각 언어별 학습 영상의 수에 따른 Gabor 특징에 의한 방법과 특징 융합에 의한 3가지 방법들의 AIR 성능

δ 는 모든 학습 영상에 대한 고유치를 내림차순으로 정렬한 다음 하위 $\epsilon\%$ 를 역치화 할 수 있는 값으로 결정하였다. 언어 인식 성능의 평가 척도로는 시험 영상의 수에 대한 정확하게 인식된 영상의 수의 비를 나타내는 평균 인식률(AIR: averaged identification rate)을 사용하였다.

그림 5는 각 언어별 학습 영상의 수에 따른 Gabor 특징에 의한 방법과 특징의 융합에 의한 3가지 방법의 AIR 성능을 보여준다. 그림 5에서 Gabor+co-occurrence는 Gabor와 co-occurrence 특징의 융합에 의한 방법, Gabor+MDLC는 Gabor와 MDLC 특징의 융합에 의한 방법, 그리고 Gabor+MDLC+co-occurrence는 제안된 Gabor, MDLC, co-occurrence 특징의 융합에 의한 방법을 나타낸다. 이때 WPCA 분류기에서 고유치에 대한 역치 δ 를 위한 ϵ 는 학습 영상의 수에 따라 0.01~1.1로 선택되었다. 그림 5에서 보면 Gabor, Gabor+co-occurrence, 그리고 Gabor+MDLC는 각각 91.57%~94.90%, 95.87%~98.63%, 98.47%~99.37%의 AIR 성능을 보인다. 반면에 제안된 Gabor+MDLC+co-occurrence는 99.00%~99.50%의 성능을 보인다. 이들 결과로부터 제안된 방법은

Gabor 특징에 의한 방법에 비하여 학습 영상의 수에 따라 4.50%~7.43%의 성능 향상을 가져온다는 것을 알 수 있다. 그리고 Gabor+co-occurrence에 비하여서는 0.77%~3.13%의 성능 향상을 가져오고, Gabor+MDLC에는 0.03%~0.53%의 성능 향상을 가져온다는 것을 알 수 있다.

표 2는 그림 5의 결과에서 제안된 방법의 성능이 가장 우수한 성능을 보인 학습 영상의 수가 48인 경우의 각 특징에 따른 AIR 성능을 보여준다. 표 2에서 보면 co-occurrence 특징은 64.53%, MDLC 특징은 90.53%, WLC 특징은 92.30%, 그리고 Gabor 특징은 94.47%의 AIR 성능을 각각 보인다. 융합된 Gabor+co-occurrence 특징은 98.30%, Gabor+MDLC 특징은 99.23%의 성능을 보인다. 반면에 제안된 Gabor+MDLC+co-occurrence 특징은 99.50%의 AIR 성능을 보여 co-occurrence 특징에는 34.97%, MDLC 특징에는 8.97%, WLC 특징에는 7.20%, Gabor 특징에는 5.03%, 그리고 Gabor+co-occurrence 특징에는 1.20%의 성능 향상을 가져온다는 것을 알 수 있다. 또한 Gabor+MDLC 특징에는 0.27%의 성능 향상을 보여 co-occurrence 특징이 Gabor+MDLC 특징에 융합됨으로써 그 만큼의 성능 향상을 가져왔음을 알 수 있다.

표 3은 각 언어별 학습 영상의 수가 48인 경우에 제안된 방법의 분류 행렬(confusion matrix)을 보여준다. 표 3에서 보면 제안된 방법이 영어, 프랑스어, 그리스어, 베트남어, 중국어, 일본어, 한국어에 대하여 0.5%~2.5%의 에러를 보이고, 러시아어, 몽골어, 히브리어, 페르시아어, 에티오피아어, 힌디어, 말라야람어, 태국어의 8개국어에 대하여는 에러가 없다는 것을 알 수 있다. 이상의 결과들로부터 제안된 방법은 특징 벡터의 차원이 비교적 낮은 56이면서도 매우 우수한 언어 인식 성능을 보인다는 것을 알 수 있다.

표 2. 각 언어 별 학습 영상의 수가 48인 경우의 각 특징별 AIR 성능

Type of texture features	Co-Occurrence	MDLC	WLC	Gabor	Gabor+co-occurrence	Gabor+MDLC	Gabor+MDLC+co-occurrence
AIR [%]	64.53	90.53	92.30	94.47	98.30	99.23	99.50
Feature dimension	8	24	96	24	32	48	56

표 3. 각 언어 별 학습 영상의 수가 48인 경우의 제안된 방법에 대한 분류 행렬, 여기서 AL은 실제 언어, IL은 인식된 언어를 나타낸다

IL \ AL	Eng	Fre	Gre	Vtn	Rus	Mgl	Heb	Per	Eth	Hin	Mal	Tha	Chn	Jap	Kor
Eng	97.5	0.0	0.0	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Fre	0.0	98.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Gre	0.0	0.0	99.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Vtn	0.5	0.0	0.0	99.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Rus	2.0	0.0	0.5	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Mgl	0.0	1.5	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Heb	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Per	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Eth	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0
Hin	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0
Mal	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0
Tha	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0
Chn	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.0	1.0	0.0
Jap	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	98.5	0.0
Kor	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	0.5	100
Error [%]	2.5	1.5	0.5	0.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0	1.5	0.0

5. 결 론

본 논문에서는 Gabor 특징, MDLC 특징, co-occurrence 특징의 융합에 의한 문서 영상에서의 질감 특징 기반 언어 인식 방법을 제안하였다. 제안된 방법에서는 특정 주파수 성분과 방향성 정보와 관련된 질감 특징을 추출하는 Gabor 특징과 국부적인 밝기의 유사성과 관련된 특징을 추출하는 MDLC 특징에 영상 내의 화소 간의 통계적 특성 8가지를 표현하는 co-occurrence 특징을 융합한 것을 질감 특징으로 사용하였고 분류기로는 WPCA를 사용하였다. 실험 결과 제안된 방법은 15개국 언어로 이루어진 실험 문서 영상 DB에 대하여 학습 영상의 수에 따라 99.0%~99.50%의 AIR 성능을 보여 Gabor 특징에 비하여 4.50%~7.43%의 성능 향상을 보였고, Gabor 특징과 MDLC 특징의 융합에 의한 방법 보다는 0.03%~0.53%의 성능 향상을 보여주었다. 그리고 기존의 co-occurrence 특징, WLC 특징, 웨이블릿 영역의 BDIP와 BVLC 특징, 그리고 MDLC 특징에 의한 방법 보다는 월등한 성능 향상을 보였다.

참 고 문 헌

[1] D. Ghosh, T. Dube, and A.P. Shivaprasad, "Script Recognition—a Review," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32, No. 12, pp. 2142–2161, 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, 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, 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, 2008.

[5] G.S. Pearke and T.N. Tan, "Script and Language Identification from Document Images," *Proc. the IEEE Workshop Document Image*

Anal., pp. 10-17, 1997.

[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, 1998.

[7] W. Chan and G. Coghil, "Text Analysis using Local Energy," *Pattern Recognit.*, Vol. 34, No. 12, pp. 2523-2532, 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, 2005.

[9] P.S. Hiremath and S. Shivashankar, "Wavelet Based Co-occurrence Histogram Features for Texture with an Application to Script Identification in a Document Image," *Pattern Recognit. Lett.*, Vol. 29, No. 9, pp. 1182-1189, 2008.

[10] 장익훈 외, "Gabor 특징과 웨이브렛 영역의 BDIP와 BVLC 특징을 이용한 질감 특징 기반 언어 인식," 전자공학회논문지, 제48권, SP편, 제4호, pp. 72-82, 2011.

[11] I.H. Jang, N.C. Kim, and M.H. Park, "Texture-feature Based Language Identification using Gabor and MDLC Features," *Proc. the IEEE Int. Conf. Multimedia Expo*, 2011.

[12] B.S. Manjunath and W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 18, No. 8, pp. 837-842, 1996.

[13] 김원희 외, "Gabor 웨이블릿을 이용한 회전 변환에 무관한 질감 분류 기법," 한국멀티미디어학회논문지, 제10권, 제9호, pp. 1125-1134, 2007.

[14] 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, 1973.

[15] G.V. Wouwer, P. Scheunders, and D.V. Dyck, "Statistical Texture Characterization from Discrete Wavelet Representation," *IEEE Trans. Image Process.*, Vol. 8, No. 4, pp. 592-598, 1999.

[16] C. Liu, "The Bayes Decision Rule Induced Similarity Measures," *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 6, pp. 1086-1090, 2007.

[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, 1984.



장 익 훈

1986년 2월 경북대학교 전자공학과 공학사
 1988년 2월 경북대학교 전자공학과 공학석사
 1998년 8월 경북대학교 전자공학과 공학박사

1988년 2월~1994년 2월 국방과학연구소 연구원
 1998년 3월~현재 경운대학교 항공전자공학과 부교수
 관심분야 : 영상처리, 컴퓨터비전 등



김 지 흥

1986년 2월 경북대학교 전자공학과 공학사
 1988년 2월 경북대학교 전자공학과 공학석사
 1996년 8월 POSTECH 전자전기공학과 공학박사

1988년 2월~1997년 2월 한국전자통신연구원 선임연구원
 2008년 미국 조지아공대 방문교수
 2002년 3월~현재 동의대학교 영상정보공학과 교수
 관심분야 : 영상통신, 의료영상, 컴퓨터비전