

공간의존행렬과 신경망을 이용한 문서영상의 효과적인 블록분할과 유형분류

김 중 수[†] 이 정 환^{††} 최 흥 문^{†††}

요 약

본 논문에서는 공간의존행렬과 신경망을 이용하여 문서영상에서 인식대상 문자가 포함되어 있는 블록들을 좀 더 세분해 낼 수 있는 효과적인 방법을 제시하였다. 제안한 방법은 명암도 문서블록에서 공간의존행렬을 구하고 7가지 질감특징을 추출한 뒤 신경망을 이용해 문서블록을 9가지 유형으로 분류할 수 있도록 하였다. 특히 기존에는 비문자영역으로 분류되던 수식, 도표, 순서도 등 주로 문자가 포함되어 있는 블록들을 세분해 낼 수 있도록 하였다. 또한 신경망 학습알고리즘인 BP를 사용함으로써 기존의 선형분류시에 요구되던 유형별 임계값과 선형면결정상수를 찾는 어려움을 해소하였다. 명암도영상을 이진화하기 전에 먼저 Sobel연산을 적용함으로써 문서 뒷면에 의한 배경잡음의 영향을 줄일 수 있도록 하였고, 교차 문자르기후 분할함으로써 블록이 작은 조각으로 나누어지는 것을 방지하도록 하였다. 실험결과 제안한 방법에서는 문자가 포함되어 있는 블록은 큰 문자, 중간 문자, 작은 문자블록 및 수식, 순서도, 도표블록의 6가지로, 그리고 비문자블록은 인물사진, 그림, 그래프 등 3가지 유형으로 상세하게 분류할 수 있었으며 전체적인 분류성능도 우수함을 확인할 수 있었다.

An Efficient Block Segmentation and Classification Method for Document Image Analysis Using SGLDM and BP

Joong-Soo Kim[†] Jeong-Hwan Lee^{††} and Heung-Moon Choi^{†††}

ABSTRACT

We proposed an efficient block segmentation and classification method for the document analysis using SGLDM(spatial gray level dependence matrix) and BP(back propagation) neural network. Seven texture features are extracted directly from the SGLDM of each gray-level block image, and by using the nonlinear classifier of neural network BP, we can classify document blocks into 9 categories. The proposed method classifies the equation block, the table block and the flow chart block, which are mostly composed of the characters, out of the blocks that are conventionally classified as non-character blocks. By applying Sobel operator on the gray-level document image before binarization, we can reduce the effect of the background noises, and by using the additional horizontal-vertical smoothing as well as the vertical-horizontal smoothing of images, we can obtain an effective block segmentation that does not lead to the segmentation into small pieces. The result of experiment shows that a document can be segmented and classified into the character blocks of large fonts, medium fonts, small fonts, the character recognizable candidates of tables, flow charts, equations, and the non-character blocks of photos, figures, and graphs.

1. 서 론

학문과 산업이 전산화되고 세분화됨에 따라 많은 전문서적과 문헌들이 출간되고 있다. 이들 출판물에서 신속하고 정확하게 정보를 얻기 위해서는 문서 자동인식 장치의 개발이 필요하다[1, 2]. 일반적으로 문서의 한 페이지는 문자, 그림, 도표, 수식 등의 서로 다른 유형의 블록들로 복합 구성되어 있

[†] 정 회 원 : 안동대학교 컴퓨터공학과 부교수

^{††} 정 회 원 : 안동대학교 컴퓨터공학과 조교수

^{†††} 정 회 원 : 경북대학교 전자공학과 교수

논문접수: 1995년8월16일, 심사완료: 1995년10월30일

으며 이런 문서를 인식하기 위해서는 문서영상을 블록으로 먼저 분할하고 분할된 블록의 특징에 따라 각각의 유형으로 분류하여 각 유형에 적합한 세부적인 인식 및 처리방법을 적용할 필요가 있다 [1, 2, 3, 4]. 따라서 문서인식 및 처리 효율을 높이기 위하여 문서영상블록의 더 자세한 유형분류가 요구된다. 특히 단순한 문자블록 뿐만아니라 도표, 순서도, 수식 등 주로 인식가능 문자로 구성된 블록은 최대한 상세하게 분류해 낼 수 있어야 한다.

블록 유형분류는 블록분할, 특징추출 및 유형분류의 세 과정으로 구성되어 있다. 블록분할 방법에는 투영에 의한 방법[2, 3], 연결길이(runlength)를 이용한 문지르기(smoothing) 방법[2, 4] 및 레이블링에 의한 방법[5] 등이 있다. 투영에 의한 방법은 수직 및 수평방향으로 여러 회 반복 투영하여 블록을 나누므로 다른 유형에 포함된 문자 부분을 분리하기 어렵다. 레이블링에 의한 방법은 블록에 대한 많은 정보를 추출할 수 있으나 4회에 걸친 순차주사(scan)가 필요하므로 처리시간이 오래 걸린다. 연결길이를 이용한 문지르기 방법은 수직방향 문지르기한 영상과 수평방향 문지르기한 영상을 논리 곱(AND)연산하여 블록으로 분할한다. 이 방법은 기억장치를 많이 필요로 한다.

블록의 특징추출 및 유형분류 방법에는 이진화된 영상으로 부터 질감특징을 이용하여 블록을 분류하는 방법[2]과 블록높이 및 평균 흑화소의 길이 등을 특징으로 추출하여 블록분류하는 방법[3, 4, 5, 6, 7] 등이 있다. 질감특징을 이용하여 분류하는 방법에서는 이진 문서영상에서 3가지 특징을 추출하여 각 블록을 큰 문자, 중간 문자, 작은 문자, 사진, 그래픽블록으로 선형분류한다. 이 방법은 특징추출시 흑백화소의 길이에 따라 적절한 임계값과 효과적인 분류를 위해 선형면결정상수(decision surface parameter)를 찾아야 한다. 블록높이 및 평균 흑화소의 길이 등의 특징으로 분류하는 방법은 블록유형에 따른 각각의 임계값을 먼저 대략적으로 설정하여야 선형분류가 가능하다. 논문 [5]에서는 유형을 문자, 직선성분, 세션이미지, 그림, 잡영성분블록으로 나누었고 나머지 논문들에서는 주로 문자, 표제, 그래픽블록으로 구분하였다.

일반적으로 연결길이를 이용한 문지르기 방법은 수평방향 문지르기한 영상과 수직방향 문지르기한

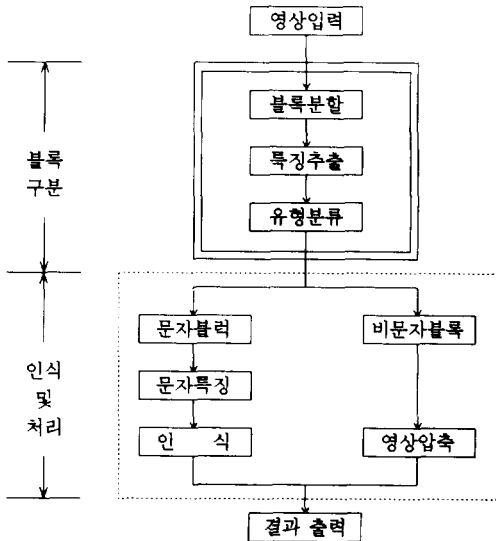
영상으로 블록을 분할함으로써 불필요하게 작은 여러 개의 블록으로 나누어지는 문제점이 있다. 위의 두 가지 기존의 특징추출 및 유형분류 방법은 이진화된 영상을 사용하기 때문에 이진화시 적절한 명암도 보정과 최적의 임계값을 찾아야 하며, 비문자 블록을 좀 더 자세하게 유형분류하여 그 중에서도 문자들이 주로 포함되어 있는 도표나 순서도 등을 세분해 내기에는 정보량도 부족하다.

본 논문에서는 명암도영상의 공간의존행렬(SG-LDM:spatial gray level dependence matrix)에서 7가지 질감특징을 추출하고 신경망 학습알고리즘인 BP(back propagation)를 사용하여 문서영상을 좀 더 효과적으로 분할 및 분류할 수 있는 방법을 제시하였으며, 기존의 연결길이를 이용한 문지르기 방법을 수정한 2중 교차 문지르기의 블록분할 방법도 제안하였다. 제안한 블록 유형분류 방법에서는 명암도 문서영상의 각 블록에서 공간의존행렬을 구하여 7가지의 질감특징을 추출하고 정규화한 뒤 BP의 입력으로 사용하여 유형분류하도록 하였다. 문자블록은 큰 문자, 중간 문자, 작은 문자 3가지 유형으로, 비문자블록은 다시 수식, 도표, 순서도 등 문자인식 대상블록과 기타 인물사진, 그림, 그래프 등의 순수 비문자블록 유형으로 상세하게 분류하였다. 명암도 문서영상에 먼저 Sobel연산을 취하여 이진화함으로써 뒷장으로부터의 배경잡음들을 제거하고 윤곽이 강조된 영상을 얻을 수 있도록 하였으며, 이진화된 영상을 수직 및 수평방향 문지르기와 수평 및 수직방향 문지르기의 2중 교차 문지르기를 한 후 분할함으로써 블록이 불필요하게 여러 개의 조각으로 분할되는 것을 막았다. 제안된 블록분할 및 유형분류 방법의 타당성과 분류율을 확인실험하였고 결과검토했다.

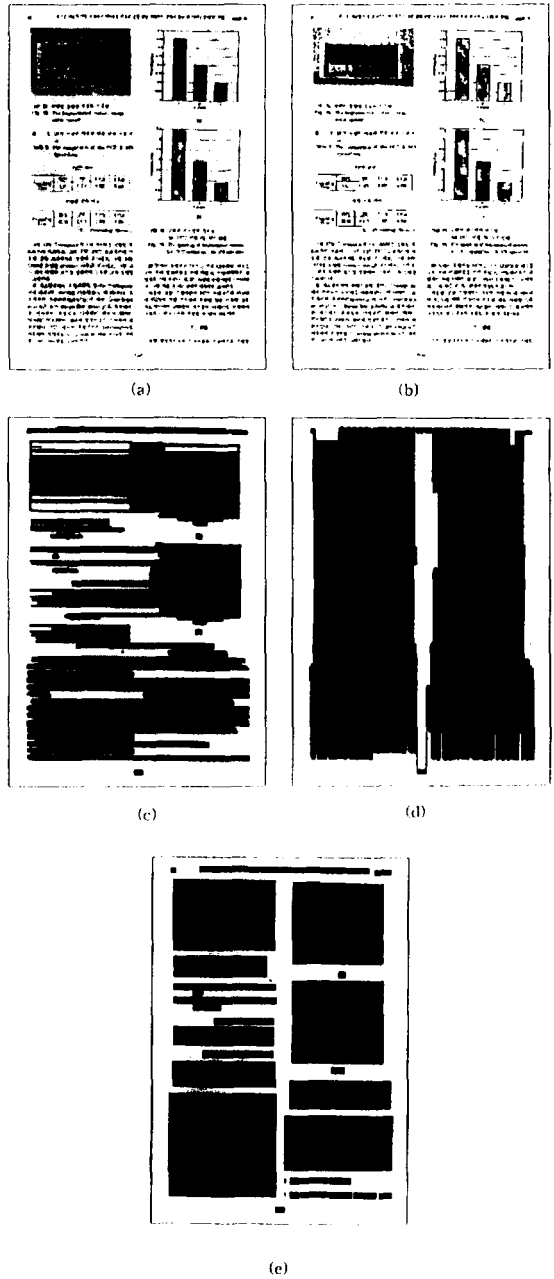
2. 문서블록의 분류과정

현재 문서인식에서는 문서의 각 블록을 문자블록, 비문자블록으로 분류하여 문자블록은 다시 세 부분할하여 특징추출을 거쳐 내용에 포함된 문자들을 인식하고, 비문자블록은 대부분 영상압축하여 처리한다. 이 단계에서 각 블록에 대한 유형분류는 인식률에 중요한 영향을 미치므로 이에 대한 깊은 연구가 필요하다[15]. 본 연구에서는 오분류률 줄

여 문서의 인식성능을 더욱 높일 수 있도록 분류유형의 종류를 더욱 세분하였다. 기존의 문자블록은 큰 문자, 중간 문자, 작은 문자로 분류하고, 기존에 비문자블록으로 분류되던 도표, 순서도, 수식 등은 그 내용에 상당수의 문자가 있으므로 이를 문자인식 대상블록으로 분류하였으며, 비문자블록은 그림, 그래프 및 인물사진 등의 블록만을 포함시키도록 하였다. 작은 문자 외의 문자블록들은 정규화함으로써 문자로의 인식이 가능해질 뿐아니라 큰 문자가 도형이나 그림으로 잘못 분류되는 것도 막을 수 있도록 하였다. 문자인식 대상블록에서는 도표는 직선 속에 문자가 있고 순서도는 도형 속에 문자가 있으므로 각각의 직선이나 도형에 관한 적절한 처리를 하고 문자부분을 인식하게 되면 결과적으로 도표와 순서도의 인식도 이루어지게 되며, 수식은 별도의 인식방법의 개발이 필요하다. 그래프, 사진 및 그림블록은 보통 영상압축으로 처리된다. 일반적인 문서인식과정은 (그림 1)과 같으며, 본 논문에서는 문서인식 과정의 블록구분 중에서 블록분할 부분에 2중 교차 문지르기를 사용한 블록분할 방법을 이용하였고, 블록의 특징추출 및 유형분류 부분에 평암도영상을 이용한 질감특징 추출방법과 학습알고리즘 BP를 이용함으로써 블록의 유형을 효과적으로 분류하도록 하였다.



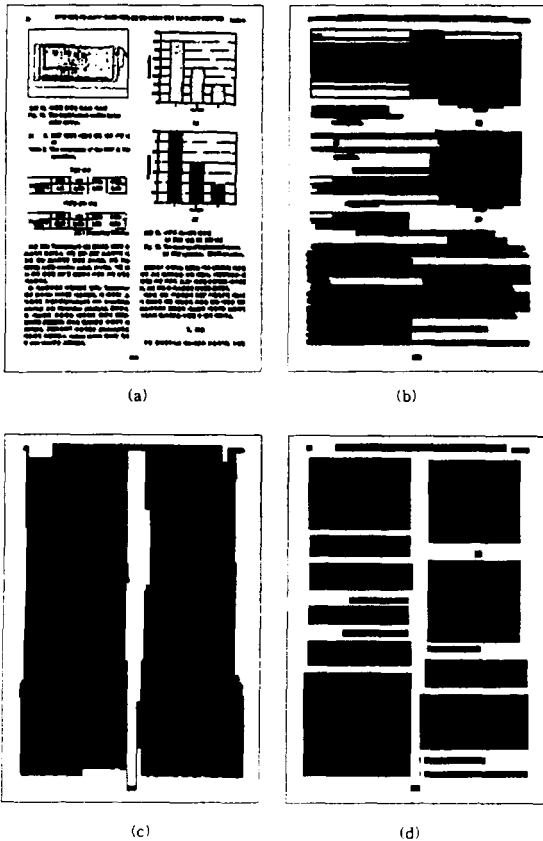
(그림 1) 문서 자동인식 시스템의 구성
(Fig. 1) Configuration of automatic document recognition system



(그림 2) 기존방법의 블록분할. (a) 원영상 (b) 이진화된 영상 (c) 수평 문지르기된 영상 (d) 수직 문지르기된 영상 (e) 기존방법을 적용한 분할영상
(Fig. 2) Block segmentation of document image. (a) original image (b) binary image (c) horizontally smoothed image (d) vertically smoothed image (e) segmented image by conventional method.

2.1 문서의 블록화

블록분할 과정에서는 한 페이지의 문서 속에 있는 서로 다른 특징을 갖는 영역들을 각각의 블록으로 분리한다. 기존의 투영에 의한 방법, 연결길이를 이용한 방법과 레이블링에 의한 방법 등에서는 모두 이진영상을 사용하여 블록분할을 한다. 이진화시 결정되는 영상의 질은 매우 중요하므로 이진영상을 블록분할에서 특징추출 및 유형분류에 이르기까지 계속 사용하기 위하여서는 명암도 보정과 최적의 임계값을 잘 찾아야 한다.



(그림 3) 제안한 방법에 따른 블록분할. (a) Sobel 처리 후 이진화된 영상 (b) 수평-수직 문지르기된 영상 (c) 수직-수평 문지르기된 영상 (d) 분할 결과

(Fig. 3) Block segmentation by the proposed method. (a) binary image after Sobel operation (b) horizontally & vertically smoothed image (c) vertically & horizontally smoothed image (d) segmented result.

Wahl 등[4]이 제안한 블록분할 방법에서는 먼저 문서영상(그림 2(a))을 이진화한 후 이 이진영상(그림 2(b))으로 수평방향 문지르기를 한 영상(그림 2(c))과 수직방향 문지르기를 한 영상(그림 2(d))로 논리곱연산을 하여 사각형으로 분할한다. 여기서는 (그림 2(d))의 윗부분과 우측하단에서처럼 작은 틈이 남게 되며 결과적으로 (그림 2(e))의 윗줄과 우측아래 부분에서 블록이 여러 개의 작은 조각으로 세분되었다. 이 방법에서는 이진화시 이웃블록이 영역분리 임계값보다 떨어질 경우 블록이 불필요하게 작은 조각으로 나누어 질 수 있으므로 이를 교정하는 과정이 요구된다. 따라서 이와 같은 문제점을 보완하기 위하여 제안된 블록분할 방법에서는 명암도 문서영상을 입력한 후 3x3 마스크의 Sobel연산을 수행한 후 영상을 이진화한다. 여기서 이진화 임계값은 흑화소의 비율이 약 10% 정도에서 선택된다[16]. 본 연구에서는 블록분할만을 위해 이진영상을 사용한다. 이는 (그림 2(a))에서 보는 바와 같이 문서의 뒤쪽 내용이 배경잡음으로 나타나는데 Sobel연산을 수행한 뒤 이진화하면 (그림 3(a))에서와 같이 배경잡음이 제거되고 윤곽이 강조된 이진영상이 구해진다. 이 영상으로 수평 및 수직방향 문지르기를 한 영상(그림 3(b))와 수직 및 수평방향 문지르기를 한 영상(그림 3(c))를 함께 논리곱연산을 수행하여 각 블록을 사각형으로 분할하였다. 이렇게 대칭적으로 두 번 반복 문지르기를 수행함으로써 블록이 필요 이상 작게 여러 개로 나누어지는 것을 방지하여 좋은 결과를 얻을 수 있도록 하였다(그림 3(d)).

2.2 블록에서의 특징추출

문서영상을 블록으로 분할한 후 각 블록이 어떤 유형에 속하는 가를 알기 위해 특징을 추출하여야 한다. 블록의 특징추출 방법에는 이진화된 영상에서 흑백 쌍 연결길이(black-white pair run length) 및 흑백흑 조합 연결길이(black-white-black combination run length)행렬을 이용하여 질감특징을 추출하는 방법[2]과 블록 좌표, 가로 및 세로의 비율, 총화소수, 흑화소의 비율 등을 추출하여 이용하는 방법[3, 4, 5, 6, 7] 등이 있다. 후자의 방법에서는 작은 문자블록에서는 작은 문자

를 대표할 수 있는 크기를 임계값으로, 또 직선은 세로와 가로의 특정한 비율을 임계값 등으로 사용하게 되므로 유형에 따른 대표값을 먼저 정하여야 한다. 전자의 방법에서 Wang 등[2]은 입력문서 한 페이지 전체를 이진화하여 이 영상을 특정한 형태의 반복적인 규칙을 갖는 질감영상으로 간주하고 흑백 및 흑백흑 화소의 연결길이가 거의 같고, 흑백 및 흑백흑 화소의 분포가 비슷한 것을 각각 한 값으로 여겨 이 빈도수들을 구하여 흑백 쌍 연결길이 및 흑백흑 조합 연결길이 행렬의 요소값으로 사용하여 3가지 질감특징을 추출하였다. 이 방법에서는 특징추출시 흑백화소의 길이에 따른 적절한 임계값을 찾아야 한다. 또한 위의 두 가지 방법에서는 공히 이진화시 영상에 대한 명암도 보정과 최적의 임계값을 찾는 것은 어려울 뿐 아니라 이진영상을 사용하므로 블록들을 자세하게 유형을 분류하기에는 정보량이 부족하다. 따라서 제안한 방법에서는 명암도영상에서 직접 질감특징을 추출하여 사용하였다.

명암도영상에서 질감특징을 추출하는 방법[10, 11]에는 1차 통계특성을 구하는 방법, 2차 통계특성을 구하는 방법, 전력스펙트럼(power spectrum)밀도함수를 이용하는 방법, 영상을 국부적으로 변환하는 방법 등이 있으며 본 논문에서는 앞의 여러 가지 방법 중에서 2차 통계특성을 구하는 방법을 사용하였다. 이는 먼저 영상내의 서로 다른 두 화소 사이의 2차 확률밀도를 구하고 여기서 통계특성을 구하는 것으로 여기에는 GLDM(gray level dependence matrix), GLRLM(gray level run length matrix) 그리고 공간의존행렬을 사용하는 방법 등이 있다. 이들 중 공간의존행렬이 명암도영상에서 질감특성을 추출하는 수단으로 가장 우수하다고 알려져 있으므로[11, 12] 본 연구에서는 이를 문서블록의 유형분류를 위한 특징추출에 적용하였다.

Haralick[11]은 θ 방향에 대하여 d 만큼 떨어진 두 화소 사이의 공간의존행렬을 구하는 방법을 제안하였으며 다음과 같이 표현할 수 있다. 사각형 모양의 영상이 있을 때, N_c 를 수평방향의 최대 영상 크기, N_r 을 수직방향의 최대 영상크기, N_g 를 최대 양자화 레벨이라 하면, 수평방향의 공간영역 N_c , 수직방향의 공간영역 L_r 및 화소값의 레벨 G 는

$$\begin{aligned} L_c &= \{1, 2, \dots, N_c\} \\ L_r &= \{1, 2, \dots, N_r\} \\ G &= \{1, 2, \dots, N_g\} \end{aligned} \quad (1)$$

와 같이 표현할 수 있다. 위와 같이 정의하면, $L_c \times L_r$ 위치에 G 를 할당하는 함수 I 는

$$I : L_c \times L_r \rightarrow G \quad (2)$$

와 같다. 그리고 공간의존행렬은 다음과 같이 상대적 빈도수 $\Phi(d, \theta)$ 로 나타낼 수 있다.

$$\Phi(d, \theta) = [f(i, j | d, \theta)] \quad (3)$$

여기서 $f(i, j | d, \theta)$ 는 공간의존행렬의 한 원소를 의미이고 θ 방향으로 d 만큼 떨어진 두 화소사이의 화소값이 i 에서 j 로 변하는 빈도수를 의미한다. 공간의존행렬을 구한 후, 사용한 전체 화소수로 공간의존행렬의 각 원소를 나누면 확률밀도함수를 구할 수 있다. (그림 4)는 공간의존행렬을 구하는 하나의 예이다. (그림 4(a))는 크기가 4×4 이고 최대 명암도레벨이 4인 원영상의 명암도레벨이고, (그림 4(b))는 공간의존행렬의 구조를 나타낸 것으로서 여기서 #는 빈도수를 나타낸다. (그림 4(c))는 두 화소의 방향이 $0^\circ, 45^\circ, 90^\circ, 135^\circ$ 인 경우에 각각 구한 공간의존행렬을 나타낸다. 여기에서 두 화소사이의 거리는 모두 1로 하였다.

	gray label	0	1	2	3
0 0 1 1	0	#(0,0)	#(0,1)	#(0,2)	#(0,3)
0 0 1 1	1	#(1,0)	#(1,1)	#(1,2)	#(1,3)
0 2 2 2	2	#(2,0)	#(2,1)	#(2,2)	#(2,3)
2 2 2 3	3	#(3,0)	#(3,1)	#(3,2)	#(3,3)

(a) (b)

4 2 1 0	6 0 2 0	2 1 3 0	4 1 0 0
$\theta_H=2$	$\theta_V=0$	$\theta_{10}=1$	$\theta_{90}=1$
2 4 0 0	4 2 0	1 2 1 0	2 2 0
1 0 6 1	2 2 2 2	3 1 0 2	0 2 4 1
0 0 1 2	0 0 2 0	0 0 2 0	0 0 1 0
	0°	90°	135°
			45°

(c)

(그림 4) 공간의존행렬의 예. (a) 4×4 영상 (b) 공간의존행렬 구조 (c) 4 방향에 대한 공간의존행렬
(Fig. 4) Example of SGLDM. (a) 4×4 image (b) SGLDM structure (c) the 4 directional SGLDMS.

그리고 각각의 방향에 대하여 정규화된 공간의존행렬 $p(i,j | d,\theta)$ 는

$$p(i,j | d,\theta) = \frac{f(i,j | d,\theta)}{[\sum_{x=0}^{Ng} \sum_{y=0}^{Ng} f(x,y | d,\theta)]} \quad (4)$$

이 되며 확률을 의미한다. 이는 공간의존행렬의 각 원소의 확률로서 이를 모두 합하면 1이 된다.

$f(i,j | d,\theta)$ 를 4방향에 대하여 표시하면 다음과 같다.

$$f(i,j | d,0^\circ) = \#\{((k,l),(m,n)) \in (Lr \times Lc) \times (Lr \times Lc) \mid k-m=0, |l-n|=d, I(k,l)=i, I(m,n)=j\} \quad (5)$$

$$f(i,j | d,45^\circ) = \#\{((k,l),(m,n)) \in (Lr \times Lc) \times (Lr \times Lc) \mid (k-m=d, l-n=-d) \text{ or } (k-m=-d, l-n=d), I(k,l)=i, I(m,n)=j\} \quad (6)$$

$$f(i,j | d,90^\circ) = \#\{((k,l),(m,n)) \in (Lr \times Lc) \times (Lr \times Lc) \mid (|k-m|=d, l-n=0, I(k,l)=i, I(m,n)=j)\} \quad (7)$$

$$f(i,j | d,135^\circ) = \#\{((k,l),(m,n)) \in (Lr \times Lc) \times (Lr \times Lc) \mid (k-m=d, l-n=d) \text{ or } (k-m=-d, l-n=-d), I(k,l)=i, I(m,n)=j\} \quad (8)$$

그리고 각 방향에 대하여 공간의존행렬로부터 각 방향에 얻을 수 있는 질감특징은 다음과 같다 [8].

$$\mu(\Phi(d,)) = \mu = \sum_{i=0}^{Ng} i \sum_{j=0}^{Ng} p(i,j | d,\theta) \quad (9)$$

$$\sigma(\Phi(d,\theta)) = \sigma^2 = \sum_{i=0}^{Ng} (i-\mu)^2 \sum_{j=0}^{Ng} p(i,j | d,\theta) \quad (10)$$

$$C(\Phi(d,\theta)) = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (i-\mu)(j-\mu)p(i,j | d,\theta)/\sigma^2 \quad (11)$$

$$E(\Phi(d,\theta)) = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} p^2(i,j | d,\theta) \quad (12)$$

$$H(\Phi(d,\theta)) = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} -p(i,j | d,\theta) \log p(i,j | d,\theta) \quad (13)$$

$$T(\Phi(d,\theta)) = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (i-j)^2 p(i,j | d,\theta) \quad (14)$$

$$I(\Phi(d,\theta)) = \sum_{i=0}^{Ng} \sum_{j=0}^{Ng} (1+(i,j)^{-1}) p(i,j | d,\theta) \quad (15)$$

여기서 식 (9)는 θ 방향으로 d 만큼 떨어진 화소들 사이의 상호발생 빈도수의 평균(average), 식 (10)은 분산(variance), 식 (11)은 상관계수(coefficient), 식 (12)는 에너지(energy), 식 (13)은 엔트로피(entropy), 식 (14)는 명암대비(contrast), 식 (15)는 영상의 국부적인 균일성(local homogeneity)를 표시하는 척도이다.

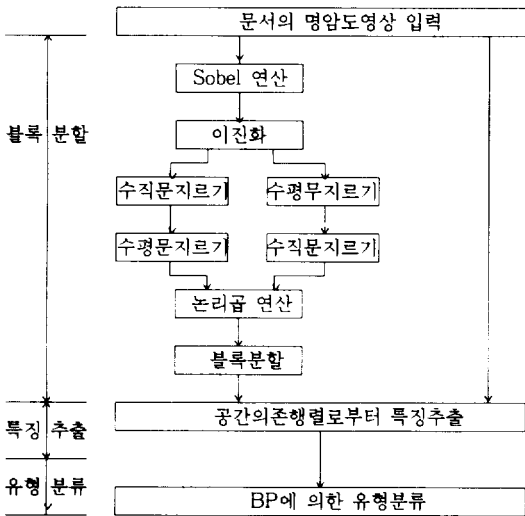
각 블록에 해당하는 명암도영상에서 4방향에 대한 공간의존행렬을 구할 수 있는데, 그 중에서 수평방향인 0° 의 공간의존행렬을 이용하여 7가지 질감특징을 추출하였다. 하나의 블록마다 7가지 특징이 추출되므로 모든 블록의 특징에서 같은 종류들끼리 묶어 0~1 사이로 정규화하면 한 블록을 나타내는 특징 7개도 모두 0~1 사이의 값으로 된다. 이 값을 BP의 입력으로 사용하였다.

2.3 블록의 유형분류

블록의 유형분류는 큰 문자, 작은 문자, 그림, 도표, 순서도 등 문서에 포함된 여러 가지 블록을 구성내용에 따라 블록을 분류하는 것이다. Wang 등 [2]은 이진영상의 질감특징을 이용하여 선형분류하는 방법을 제안하였는데 이 방법에서는 분류를 하기 위해서 수 차례의 실험을 통해 영상형태에 따른 최적의 선형면결정상수를 찾아야 하며, 블록의 유형은 큰 문자, 중간 문자, 작은 문자, 사진, 그래픽블록 등 5가지로 분류하였다. 블록 좌표, 가로 및 세로의 비율, 총화소수, 흑화소의 비율 등의 특징을 이용하여 분류하는 방법[3, 4, 5, 6, 7]은 각 유형에 맞는 적절한 임계값을 먼저 설정하여야 선형분류가 가능하며 논문 [5]는 유형을 문자, 직선 성분, 사각형성분, 그림, 잡형성분블록 등 5가지로 나누었고, 나머지 논문은 주로 문자, 표제, 그래픽블록 등 3가지로 구분하였다.

본 논문에서는 위의 선형분류에 따른 문제점을 해결하기 위해 비선형분류기인 BP를 사용하여 문자블록은 큰 문자, 중간 문자, 작은 문자로, 문자인식 대상블록에는 도표, 순서도, 수식으로, 또 비문자블록에는 그림, 그래프 및 인물사진 등의 9가지 블록으로 상세하게 분류하였다. BP의 입력층의 노드는 각 블록에서 추출되는 7가지 특징이 한 블록의 특성을 나타내므로 7개가 되며, 1개의 은닉층은

8개의 노드를 사용하였고, 출력층은 블록유형의 종류에 따라 노드의 수가 결정된다. 이상과 같이 제안한 방법은 먼저 명암도영상을 입력하여 Sobel 연산을 한 뒤 이진화한 영상에 2중 교차 문지르기를 하여 만들어진 2개의 영상을 논리곱연산하여 블록으로 분할한다. 여기서 구해진 블록의 위치정보를 이용하여 명암도영상의 각 블록으로부터 공간의존행렬을 구하여 7가지 특징을 구한 뒤 정규화를 거쳐 BP에 적용하여 각 블록을 여러가지 유형으로 분류한다. 이를 순서도로 나타내면 (그림 5)와 같다.



(그림 5) 제안한 문서영상 블록 유형분류 순서도
(Fig. 5) Flowchart of the proposed block classification for document image analysis

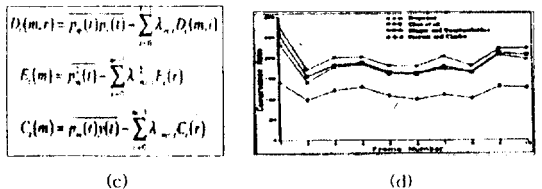
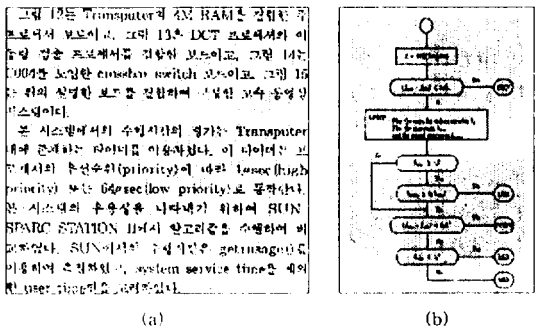
3. 실험 및 결과 고찰

제안된 방법의 블록분할 및 분류 방법을 실험하여 Wang 등[2]의 방법에 대한 결과와 비교 고찰하였다. 문서는 스캐너를 이용하여 입력하였으며, 그 해상도는 100 DPI(dot per inch)이며 명암도는 256이다. 프로그램은 C-언어로 작성하여 이를 SUN SPARC-10에서 수행하였다.

입력된 명암도영상에 3×3 마스크의 Sobel 연산을 취한 다음 이진화를 수행하였고 기존의 문지르기 기법을 개선한 2중 교차 문지르기 방법을 적용한 후 블록으로 분할하였다. 이 때 이진화 임계값

은 흑화소 비율이 약 10% 정도에서 선택하여 사용하였다. 블록분류를 위해 사용된 특징은 수평방향의 공간의존행렬에서 추출된 7가지이며, 이 때 사용된 마스크의 크기는 16×16이다. 추출된 특징값들을 0~1 사이로 정규화하여 BP의 입력으로 사용하였다. 신경회로망 학습시 사용된 매개 변수는 alpha, beta 및 eta를 각각 0.7, 0.9, 0.001로 사용하였고, 수렴을 위한 개개의 오차한계는 0.001로 하였다. 각 실험에서 학습의 반복회수 4000회부터 5000회 사이에서 수렴되었다. 은닉층은 8개의 노드를 갖는 단일층을 사용하였다. 이는 실험을 통해 결정된 값이다.

블록분류 성능비교에는 전자공학회 논문지로부터 6가지 유형의 52개 블록을 사용하였다. BP의



미선형 시스템의 디지털 Feedback 선형화 기법

Processing Time	SUN	MF	2P3	4P3
1.6	0.33	2.22	0.13	

Processing Time	SUN	MF	2P3	4P3
35.38	2.42	1.33	0.45	

(그림 6) 전자공학 논문지로부터 분할된 블록 예 (a) 작은 문자 (b) 순서도 (c) 수식 (d) 그래프 (e) 큰 문자 (f) 도표
(Fig. 6) The example blocks segmented from KITE (a) small font letter (b) flow chart (c) equation (d) graph (e) large font letter and (f) table

학습에는 실험 대상블록의 반을, 분류는 그 나머지를 사용하였다. 전자공학회 논문지에 대한 실험에서는 작은문자, 큰문자, 순서도, 도표, 수식, 그래프블록 등 6가지로 분류되었다.

(그림 6)은 본실험에 사용된 전자공학회 논문지로부터 분할된 작은 문자, 도표, 그래프, 수식, 큰 문자 및 순서도블록 등의 예이다.

Wang 등의 방법을 적용한 결과는 <표 1>과 같다. 이 때 사용된 특징은 F1, F2, F3이며 객관적인 비교를 위해 분류기는 같은 BP를 사용하였다. 여기서 F3을 추출할 때 사용되는 임계값 T1과 T2에 따라 분류결과에 큰 차이가 발생된다. <표 1>은 Wang 등이 사용한 T1 및 T2 값으로 각각이 50, 15 일 때의 결과로서 26개의 블록 중에서 14개를 분류하였다. <표 2>는 실험을 통해 가장 분류율이 높은 경우로 T1 및 T2 값은 50, 20이며, 26개 중 19개를 분류하였다.

<표 3>에는 제안한 방법에 의한 결과로서 잘못된 분류가 발생하지 않았다. <표 4>는 IEEE Medical Image 논문에서 127개 블록 9가지 유형에 대한 분류 결과이다. 블록의 유형분류는 작은문자, 중간문자, 큰문자, 순서도, 도표, 수식, 그래프, 그림 및 인물사진블록 등 9가지로, 잘못된 분류는 그림블록에서 1개가 발생하였다.

처리시간은 한 페이지당 블록분할에 약 0.5초가 소요되며, 공간의존행렬에 의한 블록의 특징추출에는 16 x 16 마스크 당 1/256초 정도가 걸렸다. 52개의 블록을 BP에서 학습시킬 때 3분 정도가 걸렸으며, 학습이 완료된 상태에서 하나의 블록을 분류하는 데는 C.P.U. 사용시간이 약 0.05초가 필요하다.

기존의 특징추출 및 유형분류 방법과 제안한 방법의 가장 큰 차이점은 기존의 방법에서는 이진영상을 사용하는 반면, 제안된 방법에서는 블록분할을 제외한 모든 경우에 명암도영상을 사용한다. 이진영상을 사용할 경우에는 이진화시 적절한 임계값 설정과 명암도 보정이 필요하며, 특징추출과 분류시에도 적절한 대표값 설정 등에 어려움이 있고, 블록의 유형은 3~5가지 정도로 분류된다. 유형의 분류에 있어서도 기존 방법에서는 선형분류기를 사용하였으나 제안한 방법에서는 비선형분류의 기능을 갖고있는 BP를 사용하였다.

실험결과 제안한 방법에서는 블록분할을 위한 이진화 이전에 먼저 Sobel연산을 취한 후 이진화함으로써 배경잡음들을 제거하고 윤곽이 강조된 영상을 얻을 수 있었으며, 이진화된 영상을 수직 및 수평방향으로 문지르기한 영상과 수평 및 수직방향 문지르기한 영상을 함께 논리곱연산하는 2중 교차문지르기한 후 분할함으로써 기존의 방법과는 달리 블록을 필요이상 여러 개의 작은 블록으로 분할하는 단점을 없앨 수 있었다. 그리고 명암도영상에서 추출된 다양한 질감특징을 사용하므로 정보량이 많기 때문에 문서영상의 블록을 6~9가지의 유형으로 상세분류가 가능하였으며 IEEE Medical Image 잡지의 실험에서 그림블록에서 1개가 발생하였다.

그러나 제안한 유형분류 방법은 명암도영상을 사용하므로 이진영상을 사용할 경우보다 메모리가 더 많이 필요하고 처리시간이 더 소요되는 단점은 있으나, 이는 장차 병렬화 등 시스템구현 방법에 따라 극복될 수 있으리라 본다.

<표 1> Wang의 방법에 따라 전자공학회 논문지를 블록 유형분류한 결과 (T1=50, T2=15)

<Table 1> The results of the Wang's block classification method for KITE when T1=50, T2=15 .

kinds of block	total blocks	learning blocks	test blocks	sucess rate
small font	8	4	4	1/4
large font	10	5	5	5/5
flow chart	6	3	3	1/3
table	8	4	4	1/4
equation	8	4	4	1/4
graph	12	6	6	5/6

<표 2> Wang의 방법에 따라 전자공학회 논문지를 블록 유형분류한 결과 (T1=50, T2=20)

<Table 2> The results of the Wang's block classification method for KITE when T1=50, T2=20 .

kinds of block	total blocks	learning blocks	test blocks	sucess rate
small font	8	4	4	2/4
large font	10	5	5	5/5
flow chart	6	3	3	3/3
table	8	4	4	2/4
equation	8	4	4	3/4
graph	12	6	6	6/6

〈표 3〉 제안한 방법에 따라 전자공학회 논문지를 블록 유형 분류한 결과

〈Table 3〉 The results of the proposed block classification for KITE

kinds of block	total blocks	learning blocks	test blocks	sucess rate
small font	8	4	4	2/4
large font	10	5	5	5/5
flow chart	6	3	3	3/3
table	8	4	4	4/4
equation	8	4	4	4/4
graph	12	6	6	6/6

〈표 4〉 제안한 방법에 따라 IEEE-MI 잡지를 블록 유형분류한 결과

〈Table 4〉 The results of the proposed block classification for IEEE-MI

kinds of block	total blocks	learning blocks	test blocks	sucess rate
small font	20	10	10	10/10
medium font	15	8	8	8/8
large font	10	5	5	5/5
flow chart	14	7	7	7/7
table	15	8	8	8/8
equation	16	8	8	8/8
graph	18	9	9	9/9
figure	12	6	6	5/6
photo	7	4	4	4/4

4. 결 론

본 논문에서는 명암도영상의 공간의존행렬에서 7가지 질감특징을 추출하고 신경망 학습알고리즘인 BP를 사용하여 문서영상을 좀 더 효과적으로 분할 및 분류할 수 있는 방법을 제시하였으며, 기존의 연결길이를 이용한 문지르기 방법을 수정한 2중 교차 문지르기의 블록분할 방법도 제안하였다. 제안한 블록 유형분류 방법에서는 명암도 문서영상의 각 블록에서 공간의존행렬을 구하여 7가지의 질감특징을 추출하고 정규화한 뒤 신경회로망 BP의 입력으로 사용하여 유형분류하도록 하였다. 문자블록은 큰 문자, 중간 문자, 작은 문자 3가지 유형으로, 비문자블록은 다시 수식, 도표, 순서도 등 문자인식 대상블록과 기타 인물사진, 그림, 그래프 등의 순수 비문자블록 유형으로 상세하게 분류하였다.

실험결과 제안한 방법에서는 명암도영상에서 추출된 다양한 질감특징을 사용하므로 정보량이 많기 때문에 문서영상의 블록을 문자가 포함되어 있는 블록은 큰 문자, 중간 문자, 작은 문자블록 및 수식, 순서도, 도표블록의 6가지로, 그리고 비문자블록은 인물사진, 그림, 그래프 등 3가지 유형으로 상세하게 분류할 수 있었다. 명암도 문서영상에 먼저 Sobel연산을 취하여 이진화함으로써 뒷장으로부터의 배경잡음들을 제거하고 윤곽이 강조된 영상을 얻을 수 있었으며, 이진화된 영상을 수직 및 수평방향 문지르기와 수평 및 수직방향 문지르기의 2중 교차 문지르기를 한 후 분할함으로써 블록이 불필요하게 여러 개의 조각으로 분할되는 것을 방지할 수 있었다. 전체적인 분류성능은 문서영상의 블록을 6~9가지의 유형으로 상세분류가 가능하였으며 잘못된 분류는 IEEE Medical Image 잡지의 실험에서 그림블록에서 1개가 발생하였다. 따라서 기존의 이진영상을 사용한 방법보다 분류성능이 우수함을 확인할 수 있었다.

앞으로 공간의존행렬에서 구한 7가지 특징을 해석하여 그들 중에서 분류에 결정적인 역할을 하는 것들을 찾아 사용되는 특징 수를 줄이는 것과 블록 분할에서는 한 블록속에 다른 유형의 블록을 포함하고 있는 경우 이를 분리해내는 레이블링 방법과 제안한 방법의 적응적 적용 등에 관한 연구가 필요하다고 사료된다. 그리고 제안한 유형분류 방법은 명암도영상을 사용하므로 이진영상을 사용할 경우보다 메모리가 더 많이 필요하고 처리시간이 더 소요되는 단점은 있으나, 이는 장치 병렬화 등 시스템구현 방법에 따라 극복될 수 있으리라 본다.

참 고 문 헌

- [1] L. O'Gorman and R. Kasturi, "Document Image Analysis Systems," IEEE Computer, vol. 5, pp. 5-8, July 1992
- [2] D. Wang and S. N. Srihari, "Classification of Newspaper Image Blocks Using Texture Analysis," Computer Vision, Graphics, and Image Processing, vol. 47, pp. 327-352, Jan. 1989.
- [3] T. Akiyama and N. Hagita, "Automated

Entry System for Printed Documents," Pattern Recognition, vol. 23, no. 11, pp. 1141-1154, Feb. 1990.

[4] F. M. Wahl, K. Y. Wong, and R. G. Casey, "Block Segmentation and Text Extrac-tion in Mixed Text/Image Documents," Computer Vision, Graphics and Image Processing, vol. 22, pp. 375-390, Feb. 1982.

[5] 백영복, 임길택, 김우태, 진성일, "영역 레이블링 방식을 이용한 일반 문서의 Layout Understanding," 제 6회 영상처리 및 이해에 관한 워크샵 논문집, vol. 1, pp. 204-209, Jan. 1994.

[6] 이인동, 권오석, 김태균, "블록영상의 추출 알고리즘," 한국정보과학회 논문지, vol. 18, no. 2, pp. 218-226, Mar. 1991.

[7] T. Pavlidis and J. Zhou, "Page Segmentation and Classification," CVGIP: Graphical Models and Image Processing, vol. 54, no. 6, pp. 484-496, Nov. 1992.

[8] 이정환, "조건부 1차원 히스토그램을 이용한 텍스처 영역 추출 및 경계 부호화," 한국과학기술원 박사 학위 논문, Jun. 1993.

[9] L. V. Gool, P. Dewaele, and A. Oosterlink, "Texture Analysis Anno 1983," Computer Vision, Graphics and Image Processing, Vol. 29, pp. 336-357, July, 1985.

[10] W. K. Pratt, Digital Image Processing, John Wiley & Sons, Inc., New York, 1978.

[11] R. M. Haralick, "Statistical and Structural Approaches to Texture Classification," Proceeding of the IEEE. vol. 67, no. 5, pp. 786-804, May, 1979.

[12] R. W. Connors and C. A. Halow, "A Theoretical Comparison of Texture Algorithms," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 3, pp. 204-222, May 1980.

[13] P. D. Wasserman, Neural Computing,

Van Nostrand Reinhold, New York, 1989.

[14] A. Blum, Neural Networks in C++, John Wiley & Sons, Inc. New York, 1992.

[15] 이인동, 권오석, 김태균, "문서 인식을 위한 전처리 기술의 소개," 정보과학회 지, vol. 9, no. 1, pp. 14-21, Feb. 1991.

[16] J. Prewitt, "Object Enhancement and Extraction," in Picture Processing Psychopictorics, B. Lipkin and A. Rosenfeld (eds.), Academic Press, New York, pp. 75-149, 1970.



김 중 수

1982년 경북대학교 전자공학과 졸업
 1984년 경북대학교 대학원 전자공학과 전자계산전공 졸업(공학석사)
 1988년 경북대학교 대학원 전자공학과 전자계산전공 박사과정 수료

1987년~현재 안동대학교 컴퓨터공학과 부교수
 관심분야: 영상처리, 패턴인식, 신경회로망, GIS



이 정 환

1983년 경북대학교 공과대학 전자공학과 졸업
 1986년 한국과학기술원 전기 및 전자공학과 졸업(공학석사)
 1993년 한국과학기술원 전기 및 전자공학과 졸업(공학박사)
 1990년~현재 안동대학교 전자공학과 조교수

관심분야: 영상처리, 컴퓨터비전, 멀티미디어, 컴퓨터그래픽, 신호처리 등



최 홍 문

1966년 인하대학교 전기공학과(전자전공) 졸업
 1974년 인하대학교 대학원 전기공학과 전자전공 졸업(공학석사)
 1984년 인하대학교 대학원 전자공학과 전자전공 졸업(공학박사)

1982년~83년 뉴욕주립대학원 객원교수
 1974년~현재 경북대학교 전자공학과 교수
 관심분야: 논리회로설계, 병렬처리, 신경회로망 및 그 응용