論文96-33B-10-13

문서영상의 에지 정보를 이용한 효과적인 블록분할 및 유형분류

(An Efficient Block Segmentation and Classification of a Document Image Using Edge Information)

朴 昶 俊 *, 全 俊 亨 *, 崔 興 文 *

(Chang-Joon Park, Joon-Hyung Jeon, and Heung-Moon Choi)

요 약

본 논문에서는 문서영상의 에지 정보를 이용한 효과적인 블록분할 및 그 유형분류 알고리즘을 제안하였다. 제안한 알고리즘에서는 문서영상의 에지화소들의 세기와 방향의 분포로부터 블록특징을 추출함으로써 배경잡음과 밝기변화에 둔감하면서도 효과적인 유형분류가 가능하도록 하였다. 추출된 특징들을 역전파(backpropagation) 신경회로망에 입력시켜 문서 영상을 큰 문자, 작은 문자, 수식, 표, 순서도와 같은 5개의 문자블록과 그래프, 사진과 같은 2개의 비문자 블록을 포함한 7가지 블록으로 유형분류 하였다. 한편, 블록 분할할 때 실제 문서의 단 간격 및 줄 간격을 도입하여 연속부 길이 한정 알고리즘(constrained run length algorithm: CRLA)을 적용함으로써 적은 메모리로도 효율적인 블록분할이 가능하도록 하였다. 여러 형태의 다양한 명암도 문서영상에 대하여 실험 해 본 결과 그 내용 변화가 다양한 그래프블록을 제외하고는 모든 블록들을 정확하게 분류해 내었다.

Abstract

This paper presents an efficient block segmentation and classification using the edge information of the document image. We extract four prominent features from the edge gradient and orientation, all of which, and thereby the block classifications, are insensitive to the background noise and the brightness variation of the image. Using these four features, we can efficiently classify a document image into the seven categories of blocks of small-size letters, large-size letters, tables, equations, flow-charts, graphs, and photographs, the first five of which are text blocks which are character-recognizable, and the last two are non-character blocks. By introducing the column interval and text line intervals of the document in the determination of the run length of CRLA(constrained run length algorithm), we can obtain an efficient block segmentation with reduced memory size. The simulation results show that the proposed algorithm can rigidly segment and classify the blocks of the documents into the above mentioned seven categories and classification performance is high enough for all the categories except for the graphs with too much variations.

I. 서 론

최근, 컴퓨터를 이용한 다중매체(multimedia) 정보처리 및 데이터베이스시스템의 급속한 발전과 그에 따른 다양한 요구에 따라 기존에 나와 있던 문서를 컴퓨터 처리에 적합한 자료구조로 변환하는 문서인식시스템에 대한 연구들이 활발히 진행되고 있다[1-10]. 일반적으로 하나의 문서는 서로 다른 여러 유형의 블록들로 구성되며 각 유형은 그의 성격이 서로 다르므로 문서를 컴퓨터에 저장시키거나 컴퓨터간 통신을 할 때 문서인식시스템에서 각 블록들을 잘 구별하여 유형 분류한 후, 각 특성 블록별로 최적 알고리즘을 적용하여

* 正會員, 慶北大學校 電子電氣工學部

(The School of Electronics and Electrical Engineering, Kyungpook National University)

接受日字:1996年5月31日, 수정완료일:1996年9月16日

인식, 압축 코딩하는 것이 효과적이다.

단순한 문자인식만을 위한 문서인식시스템의 경우에는 문서를 문자블록과 비문자블록으로만 구분하여도 되나, 문서내의 특정 블록만을 선별하여 수정하거나 출력하는 것이 가능해야 되는 문서 데이터베이스시스템의 경우에는 문서를 필요한 만큼의 여러 블록으로 상세 분류하는 것이 요구된다. 문서인식시스템의 분류한도를 객관성 있게 단정하기는 어려우나, 일반적인 논문이나 교재의 경우 큰 문자, 작은 문자, 수식, 표, 순서도, 그래프 및 사진과 같은 7가지 블록들이 주로 포함되어 있으므로 이를 컴퓨터 처리하기 위한 문서인식시스템에서는 최소한 이들 7가지 유형의 블록은 분류해 낼 수 있어야 하며, 따라서 이들 각 블록을 잘 구별해 낼 수 있는 효과적인 블록 특징추출과 유형분류에 관한 연구가 필요하다.

유형분류를 위한 블록특징은 이진화 영상이나 명암도 영상으로부터 추출된다. 이진화 영상으로부터 특징을 추출하는 방법에는 질감을 이용하는 방법[1], 블록 높이 및 평균 흑화소의 길이 등을 이용하는 방법[2] 및 체인코드를 이용하는 방법[3] 등이 있다. 이러한 방법에서는 낮은 명암도를 가지는 배경잡음은 이진화할 때 제거되므로 특징추출이 배경잡음에 둔감하나, 영상의 밝기변화에 따라 최적의 이진화 임계값을 설정해야 하는 등의 문제점이 따르며, 흑화소수와 흑화소의 연속길이 등으로부터 특징을 추출하므로 논문지나 교재 등에 나오는 모든 블록들을 구분하기에는 부적합하다.

한편, 명암도 영상으로부터 특징 추출하는 방법에는 명암도의 분포를 이용하는 방법[4]과 SGLDM(spatial gray level dependency matrix)을 이용하는 방법 등[5]이 있다. 특히, Kim 등[5]은 명암도 영상으로부터 질감특징을 추출하여 큰 문자, 중간 문자, 작은 문자, 수식, 표, 순서도, 그래프, 사진 및 그림 등 9가지 블록으로 상세 분류하였으나, 특징추출에 많은 처리 시간이 요구된다. 이들과 같이 명암도 영상에서 특징을 추출하는 경우에는 최적의 이진화의 임계값을 설정해야 하는 등의 문제점은 없으나, 배경잡음 및 밝기변화에 대한 별도의 처리과정이 요구된다. 따라서, 이진화의 임계치 설정 등이 까다롭지 않으면서도 영상의 밝기변화나 문서 뒷면의 배경잡음에 둔감하고 최소한 논문지에 나오는 대부분의 블록 유형들을 모두 분류해낼 수 있는 특징추출과 유형분류방법이 요구된다.

본 논문에서는 문서영상의 에지 정보를 이용한 효과적인 블록분할 및 유형분류 알고리즘을 제안하였다. 제안한 알고리즘에서는 유형분류를 위한 블록 특징추출 단계에서 에지화소들의 세기와 방향의 분포로부터 에지 세기의 평균, 블록의 크기로 정규화된 에지화소수, 주된 수평-수직 방향 에지화소 함유율 및 주된 사선 방향 에지화소 함유율의 네 가지를 특징으로 추출하였다. 이와 같이 에지를 추출함으로써 문자, 선분 등과 같이 진한 부분과 배경잡음과 같이 흐린 부분으로 명암도 값이 이분화된 화소 중, 진한 에지화소들로부터 특징이 추출되어 문서의 배경잡음에 둔감하도록 하였다. 또한, 에지화소의 방향 분포로부터 블록내의 문자와 선분의 구성비를 구하기 때문에 문서의 밝기 변화에도 둔감하도록 하였다. 이러한 특징들을 역전파(backpropagation) 신경회로망에 입력시켜 문서 영상을 큰 문자, 작은 문자, 수식, 표, 순서도와 같은 5개의 문자블록과 그래프, 사진과 같은 2개의 비문자블록을 포함한 7가지 블록으로 유형분류 하였다. 한편, 블록분할할 때 실제 문서의 단 간격과 줄 간격을 도입하여 연속부 길이 한정 알고리즘(constrained run length algorithm: CRLA)[1, 2, 5]을 수평방향으로 적용한 결과에 다시 수직방향 적용하도록 함으로써 적용 결과보관을 위한 별도의 메모리를 요구하지 않으며, 적용 횟수도 줄어 적은 메모리로도 효율적인 블록분할이 가능하도록 하였다. 또한, 입력된 문서영상으로부터 먼저 에지를 추출한 후 블록분할하기 때문에 이진화의 임계치 설정도 용이하다. 제안한 블록분할 및 특징추출 알고리즘의 타당성을 확인하기 위하여 여러 형태의 다양한 명암도 문서영상에 대하여 시뮬레이션하고 그 결과를 검토 고찰하였다.

Ⅱ. 에지 정보를 이용한 블록분할 및 유형분류

제안한 문서유형분류시스템의 흐름도는 그림 1과 같다.

즉, 명암도 영상에 대해 Prewitt 연산자를 적용하여 에지화소의 세기와 방향을 추출한 후, 임계치 기법을 적용함으로써 배경잡음이 제거되며, 에지를 추출한 후 블록분할을 하기 때문에 블록 분할할 때 필요로 하는 이진화의 임계값 설정을 용이하도록 하였다. 또한, 개선된 방법으로 연속부 길이 한정 알고리즘을 적용하여 메모리 요구량이 적으면서도 효율적인 블록분할이 가

능하도록 하였다. 한편, 블록특징추출단계에서는 임계치 기법이 적용된 에지화소의 세기와 방향 분포로부터 배경잡음 및 밝기변화에 둔감하면서도 효율적인 유형분류가 가능한 네 가지 특징을 추출한 후 역전파 신경회로망을 이용하여 유형분류 하였다.

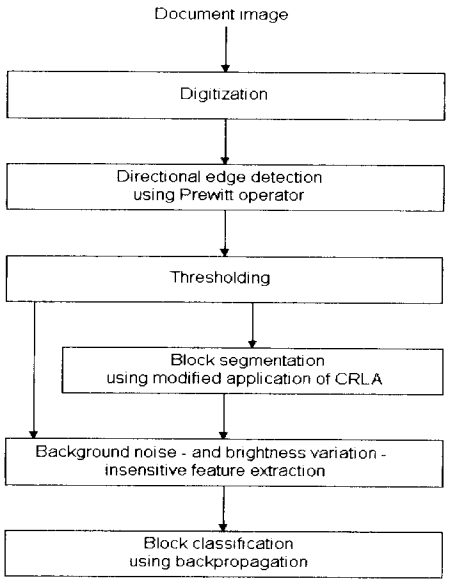


그림 1. 제안한 문서유형분류시스템의 흐름도
Fig. 1. Flow-chart for the proposed document analysis system.

1. Prewitt연산자를 이용한 에지 정보 추출

제안한 알고리즘에서는 Prewitt연산자[11]를 이용하여 블록분할과 특징추출에 사용되는 에지 정보를 추출하였다. Prewitt연산자는 일련의 마스크 연산을 통해 에지화소의 세기와 방향을 결정하게 되며, 이때 사용된 방향 성분과 각 방향 마스크는 그림 2와 같다.

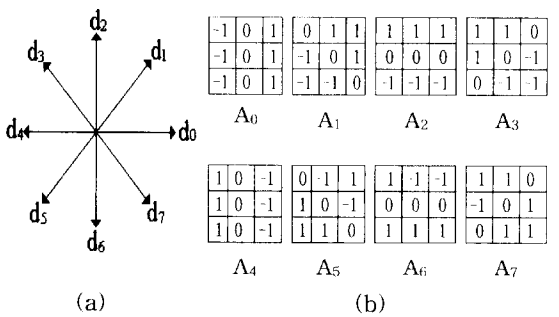


그림 2. (a) 8 가지 에지 방향 및 (b) 각 방향 Prewitt 경사 마스크
Fig. 2. (a) The eight edge directions and (b) their Prewitt gradient masks.

그림에서 에지화소의 방향(d_0, d_1, ... , d_7)은 8가지 마스크(A_0, A_1, ... , A_7)의 상승적분값 중 최대값을 가지는 마스크에 따라 결정되며 그 때의 최대값이 그 에지화소의 세기가 된다. 한편, 문자의 에지화소 방향은 주로 수평-수직방향과 사선방향이 혼재하는 반면, 비문자의 경우 주로 수평-수직방향 또는 사선방향으로만 에지화소의 방향이 나타나므로, 수평-수직과 사선 방향만 확인하여도 문자와 비문자의 구분은 가능하다. 따라서, 본 연구에서는 추출된 d_0에서 d_7까지의 8 방향을 수평-수직 방향을 의미하는 짝수 번과 사선 방향을 의미하는 홀수 번의 두 가지 방향으로만 구분하였다. 이와 같이 에지 정보가 추출된 영상에 대해 임계치 기법을 적용하므로 임계치값 설정이 용이할 뿐 아니라 상당 부분의 배경잡음도 제거 할 수 있다.

2. 메모리 요구량이 적은 블록분할

문자인식의 전처리 단계인 문자열 추출 등에 사용되는 기존의 연속부 길이 한정 알고리즘[1, 2, 6]을 그대로 블록분할에 적용시키려면 수평 및 수직 방향으로 각각 연속부 길이를 크게 하여 연속부 길이 한정 알고리즘을 적용한 후, 각각의 결과를 메모리에 저장시킨다. 저장된 메모리 내의 결과들에 대해 논리적 AND 연산을 수행한 결과에 대해 다시 수평 방향으로 연속부 길이를 작게 하여 연속부 길이한정 알고리즘을 적용시키게 된다. 이와 같이 기존의 연속부 길이 한정 알고리즘을 그대로 적용하여 블록 분할할 경우, 문자블록인 경우 문단 단위가 아니라 문자열 단위의 필요 이상 작은 블록으로 나누어지게 되므로 효율적인 블록 분할이 어렵게 된다. 또한, 수평 및 수직 방향으로 연속부 길이 한정 알고리즘을 적용한 결과를 보관할 별도의 메모리가 필요할 뿐 아니라 최소 3회의 연속부 길이 한정 알고리즘을 적용해야 되는 등의 문제점이 생기게 된다.

이를 해결하기 위해서, 본 논문에서는 먼저, 수직 방향 투영[11]을 이용하여 각 단 사이의 간격을 구하고, 수평 방향 투영[11]을 이용하여 줄 간격을 구한다. 그리고, 단 사이의 간격보다 작은 연속부 길이로 수평 방향 연속부 길이 한정 알고리즘을 적용한 결과영상에 대해 다시 줄 간격 보다 큰 연속부 길이로 수직 방향 연속부 길이 한정 알고리즘을 적용한다. 이러한 과정으로 블록 분할함으로써, 문자열 단위의 작은 블록으로 나뉘어지는 문제를 해결하였다. 또한, 수평 및 수직 방향 연속부 길이 한정 알고리즘을 적용한 결과를 보관

할 별도의 메모리가 필요 없으므로 메모리 요구량을 기존의 연속부 길이 한정 알고리즘 적용 방법에 비해 2배 정도 줄였을 뿐 아니라, 연속부 길이 한정 알고리즘의 적용 횟수도 2회로 줄임으로써 속도도 향상시킬 수 있도록 하였다.

3. 배경잡음 및 밝기변화에 둔감한 블록특징 추출

제안한 알고리즘에서는 문서의 밝기변화 및 배경잡음에 둔감하면서도 효과적인 유형분류가 가능하도록 하기 위해 임계치 기법이 적용된 에지화소의 세기와 방향 분포로부터 에지 세기의 평균, 블록의 크기로 정규화된 에지화소수, 주된 수평-수직 방향 에지화소 함유율 및 주된 사선 방향 에지화소 함유율의 네 가지를 특징으로 추출하였다.

먼저, 에지 세기의 평균(edge gradient average)을 의미하는 첫 번째 특징 F_1을

$$F1 = \frac{\sum_{i=T_h}^{G_{max}} H(i) \cdot i}{\sum_{i=T_h}^{G_{max}} H(i)} \quad (1)$$

와 같이 정의하였다. 식에서 $H(i)$는 블록내의 에지화소 세기의 누적분포[11]를, i는 누적분포의 인덱스를, T_h는 에지화소 세기에 적용한 임계치를 각각 나타내며, G_{max}는 에지화소 세기의 최대값을 의미한다. 일반적으로 배경과 물체간의 명암도 차이가 상대적으로 작은 사진블록의 경우 에지 세기가 약하게 나타나는 반면, 문자, 도표, 수식, 그래프 등과 같이 배경과 물체간의 명암도 차이가 상대적으로 큰 블록에서는 에지화소 세기가 강하게 나타난다. 따라서, 사진블록의 특징 F_1은 다른 블록들의 특징 F_1보다 작은 값을 가지므로 특징 F_1으로 사진 블록을 다른 블록들과 구별할 수 있다. 한편, 블록의 크기로 정규화된 에지화소수(normalized number of edge pixels)를 의미하는 두 번째 특징 F_2를

$$F2 = \frac{\sum_{i=Th}^{G_{max}} H(i)}{Nx \cdot Ny} \quad (2)$$

와 같이 정의하였다. 식에서 N_x , N_y는 각각 블록의 가로 크기, 세로 크기를 의미한다. 따라서, 에지화소수가 다른 블록들에 비해 상대적으로 많은 작은 문자, 큰 문자, 표와 같은 블록의 경우, 특징 F_2는 상대적으로 큰 값을 가지게 되며, 배경이 대부분을 차지하므로 에지화소수가 다른 블록들에 대해 상대적으로 적은 그래프블록의 경우 특징 F_2는 상대적으로 작은 값을 가지게 된다. 또한, 수식과 순서도블록의 F_2 특징은 작은 문자, 큰 문자, 표블록의 F_2 특징과 그래프블록의 F_2 특징의 사이의 값을 가지게 된다. 따라서, F_2에 의해 그래프블록을 다른 블록과 구별 할 수 있다. 1990년부터 1994년까지 발간된 대한전자공학회 논문지와 International Conference on Document Analysis and Recognition '93 으로부터 발췌한 영상에서 각 유형별로 20개씩의 블록에 대한 F_1 및 F_2 특징값 분포를 표 1에 나타내었다.

표 1. 각 블록별 특징 F1 및 F2의 특징값 분포

Table 1. The distribution of F1 and F2 for each block.

Categories of Blocks	F_1	F_2
Large Letters	115 - 135	0.21 - 0.27
Small Letters	115 - 135	0.25 - 0.29
Tables	115 - 135	0.18 - 0.26
Flow Charts	115 - 135	0.10 - 0.15
Equations	115 - 135	0.11 - 0.15
Graphs	110 - 135	0.01 - 0.09
Photographs	65 - 85	0.01 - 0.10

특징 F_1에 의해 사진블록이, 특징 F_2에 의해 그래프블록이 다른 블록들로부터 구분됨을 표를 통해 알 수 있다. 그러나, 특징 F_2는 단위 면적당 서로 비슷한 개수의 에지 성분을 가지는 작은 문자, 큰 문자, 표블록을 각각 구분할 수 없으며, 순서도와 수식블록도 서로 구분할 수 없다. 따라서, 특징 F_2에 의해 서로 구분이 되지 않는 이와 같은 블록들을 구분할 수 있도록 하기 위해 세 번째 특징 F_3인 주된 수평-수직 방향 에지화소 함유율(axial dominant edge pixel contents)을

$$F3 = \frac{\sum_{i=2}^{N_x-3}\sum_{j=2}^{N_y-3}\sum_{k=i-2}^{i+2}\sum_{l=j-2}^{j+2}[(E_{ij}-O_{kl})\neq 0]}{\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}[E_{ij}]} \quad (3)$$

와 같이 정의하였다. 식에서 E_{ij}는 위치 (i,j)에서의 임계치 이상의 세기를 가지는 수평-수직 방향 에지화소를 의미하며, O_{kl}은 위치 (k,l)에서의 임계치 이상의 세기를 가지는 사선 방향 에지화소를 의미한다. 식 (3)의 분모항은 블록내의 수평-수직 방향의 에지화소수를 나

타내며, 분자항은 수평-수직 방향 에지화소 중 사선 방향의 에지화소에 인접한 부분을 제거하고 남은 수평-수직 방향 에지화소수를 나타낸다. 일반적으로, 수평-수직 방향과 사선 방향의 에지화소들이 서로 가까운 거리에 혼재해 있는 문자의 경우에는 분자항이 거의 0에 가까운 값을 가지게 된다. 또한, 표블록과 같이 수직 및 수평 직선이 많이 존재하는 경우에는 분자항의 값이 분모항과 비슷한 값을 가지게 된다. 따라서, 특징 F_3는 블록 내에서 문자와 수평 및 수직 직선의 구성비를 나타내게 된다. 특징 F_3에 의해 문자만으로 구성된 작은 문자블록과, 수평 및 수직선이 많이 존재하는 표 블록 그리고 긴 직선 성분이 조금 존재하는 큰 문자블록들이 서로 구분될 수 있으며, 직선이 조금 존재하는 수식블록과 직선이 많이 존재하는 순서도블록도 서로 구분될 수 있다. 이러한 세 번째 특징 F_3의 추출 결과를 그림 3에 나타내었다.

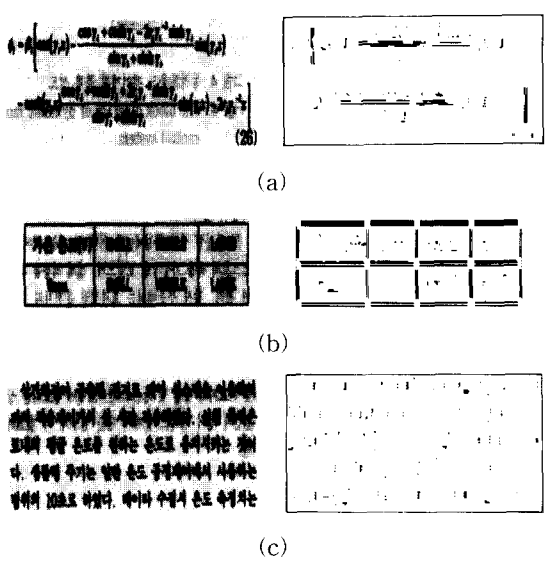


그림 3. (a) 수식 (b) 표 (c) 문자블록과 각각의 블록 특징 F_3

Fig. 3. F_3 for each block of (a) the equation, (b) the table, and (c) the text.

그림에서 보는 바와 같이 수식의 경우 수식 기호에 해당하는 긴 직선 부분들만 남아 있으며, 표의 경우 문자가 제거되고 표의 뼈대 구조만이 남고 문자 블록의 경우는 문자가 대부분 제거됨을 알 수 있다. 즉, 특징 F_3 추출 후에는 블록 내에서 문자 부분이 제거된다. 따라서, 블록 내에서 문자와 직선의 구성비를 나타내는 특징 F_3를 이용함으로써, 작은 문자, 큰 문자, 표블록을 서로 구분할 수 있으며, 또한 수식과 순서도블록을 서로 구분할 수 있다. 따라서, 전술한 특징 F_1, F_2 및 F_3를 함께 이용하면 사진, 작은 문자, 큰 문자, 수식, 표, 순서도 및 그래프블록을 서로 구분할 수 있다. 그러나, 그래프의 경우 그 형태가 매우 다양하여 그 블록내의 정보량, 즉, 에지화소의 개수 등의 변화가 그래프의 종류나 표현 방법에 따라 상당한 차이가 있으므로, 특징 F_2 및 F_3를 이용하여 그래프블록을 다른 블록들로부터 구분하기는 어려운 경우가 있다. 특히, 특별한 종류의 그래프 즉, 모눈종이 그래프와 같이 직선을 배경으로 하는 그래프의 경우 전술한 특징들을 이용하면 표로 오분류할 수 있다. 따라서, 본 연구에서는 이와 같이 특별한 종류의 그래프 분류를 위해 주된 사선 방향 에지화소 함유율(diagonal dominant edge pixel contents)을 나타내는 네 번째 특징 F_4를

$$F4 = \frac{\sum_{i=2}^{N_x-3}\sum_{j=2}^{N_y-3}\sum_{k=i-2}^{i+2}\sum_{l=j-2}^{j+2}[(O_{ij}-E_{kl})\neq 0]}{\sum_{i=0}^{N_x-1}\sum_{j=0}^{N_y-1}[O_{ij}]} \quad (4)$$

와 같이 정의하였다. 식에서 분모항은 블록내의 모든 사선 방향 에지화소수를 나타내고, 분자항은 사선 방향 에지화소 중 수평-수직 방향 에지화소에 인접한 에지화소들을 제거하고 남은 사선 방향 에지화소수를 의미한다. 즉, 특징 F_4는 블록 내에서 수직, 수평직선을 제외한 순수 사선 방향 직선을 구성하는 에지화소의 비율을 나타내게 된다.

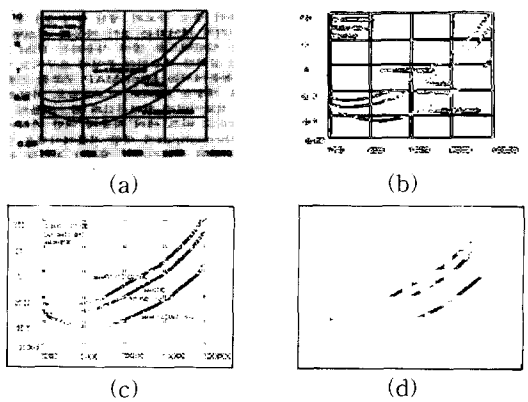


그림 4. 그래프 블록에 대한 특징 F_4 추출 (a) 그래프 블록 (b) 수평-수직 방향의 에지 성분 (c) 사선 방향의 에지 성분 (d) 추출된 특징 F_4

Fig. 4. Extraction of F_4 for a graph block (a) the graph, (b) it's axial edge components, (c) it's diagonal edge components, and (d) the extracted F_4.

따라서, 특징 F_4를 이용하면 모눈 눈금 그래프와 같은 특별한 종류의 그래프도 구분할 수 있다. 그래프블록에 대한 특징 F_4를 그림 4에 나타내었다. 그림 5에는 문자, 표, 수식블록에 대한 특징 F_4 추출 결과를 도시하였다.

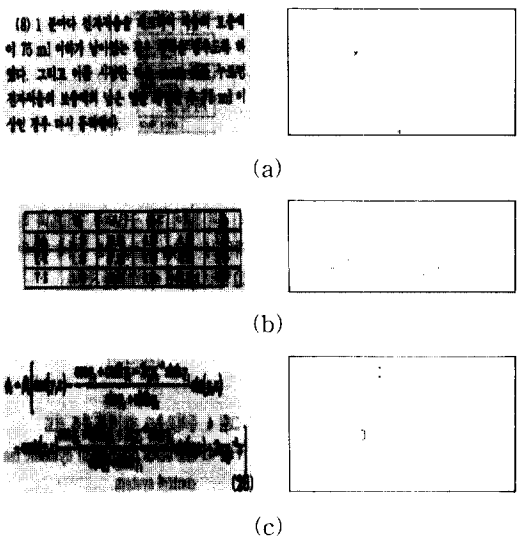


그림 5. (a) 문자 (b) 표 (c) 수식과 그들의 블록특징 F_4

Fig. 5. F_4 for each block of (a) the text, (b) the table, and (c) the equation.

강한 직선 성분을 배경으로 하는 특별한 종류의 그래프에서 특징 F_4를 추출하면, 배경인 직선 성분을 제외한 사선 방향 직선을 구성하는 에지화소들이 추출됨을 그림 4를 통해 알 수 있다. 그리고, 모든 사선 방향 에지화소가 수평-수직 방향 에지화소에 인접한 거리에서 혼재해 있는 문자, 표, 수식블록의 특징 F_4는 거의 0에 가까운 값을 나타내게 됨을 그림 5를 통해 알 수 있다. 따라서, 특징 F_4를 이용하면 특징 F_1, F_2 및 F_3를 이용하여도 구분하기 힘든 그래프 블록을 구별할 수 있다.

전술한 4가지 특징들은 임계치 기법이 적용된 후 문자, 선분 등과 같이 진한 부분의 에지화소로부터 추출되므로 배경잡음에 둔감하다. 또한, 에지화소 방향 분포, 즉, 문자, 수직, 수평 및 사선 방향 직선의 비율로부터 특징이 추출되므로, 영상의 밝기변화에 둔감한 성질을 가지게 된다. 이러한 4가지 특징에 대한 공간 분포도를 그림 6에 나타내었다.

그림 6에 보듯이 특징 F_1과 특징 F_4에 의해 에지화소의 세기가 약한 사진블록과 사선 방향 선분이 많은 그래프블록이 다른 블록들로부터 구분되며, 특징 F_2와 특징 F_3에 의해 다른 5가지의 블록들이 서로 구분될 수 있음을 알 수 있다.

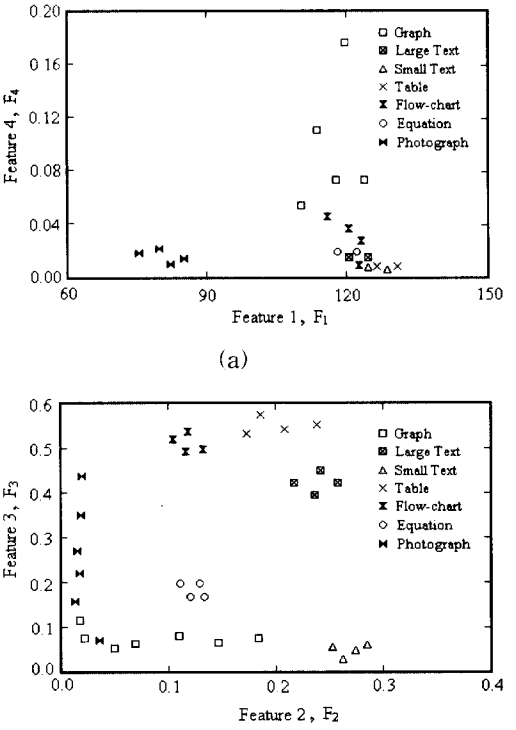


그림 6. 특징공간 분포 (a) F_1 및 F_4의 공간 분포 (b) F_2 및 F_3의 공간 분포

Fig. 6. Feature space for (a) the F_1 and F_4 and (b) the F_2 and F_3.

4. 역전파 신경회로망을 이용한 유형분류

본 논문에서는 추출된 특징을 기반으로 각 블록을 유형분류하기 위해 비선형 분류 기능을 갖는 삼층 구조의 역전파 신경회로망[12]을 사용하였다. 그림 7은 입력층(input layer)과 은닉층(hidden layer) 및 출력층(output layer)의 삼층 구조를 가지는 신경회로망을 예시한 것이다.

이러한 다층 구조의 신경회로망은 다음과 같은 역전파 신경망 알고리즘을 이용하여 효과적으로 학습시킬 수 있다. 즉, 추출된 특징값의 입력값에 대해서 은닉층 j번째 뉴런과 출력층의 k번째 뉴런의 입력값 net_j, net_k와 출력값 O_j, O_k는 다음과 같이 각각 구해진다. 즉,

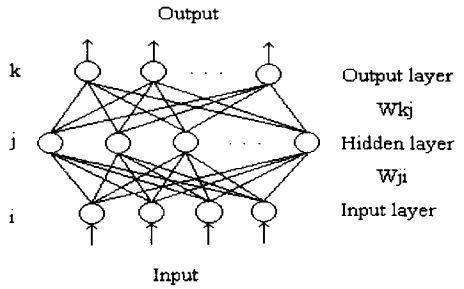


그림 7. 역전파 신경회로망의 구조
Fig. 7. Structure of the backpropagation.

$$net_j = \sum_{i=0}^{N_i-1} W_{ji}\, O_i \qquad (0 \le j \le N_h-1) \tag{5}$$

$$O_j = f(net_j) = 2 / (1+\exp(-net_j)) - 1 \tag{6}$$

$$net_k = \sum_{j=0}^{N_h-1} W_{kj}\, O_j \qquad (0 \le k \le N_o-1) \tag{7}$$

$$O_k = g(net_k) = net_k \tag{8}$$

이다. 식 (6)과 (8)의 N_i, N_h 및 N_o는 각각 입력층, 은닉층 및 출력층의 뉴런 수를 나타내고, 식 (7)과 (9)에서 $f()$와 $g()$는 본 논문에서 사용한 활성함수로서 각각 [-1, +1] 의 값을 갖는 sigmoid 함수와 입력값을 그대로 출력하는 선형함수를 나타낸다. 한편, 출력층과 은닉층 사이의 가중치 W_{kj} 및 은닉층과 입력층 사이의 가중치 W_{ji}는 다음과 같이 갱신된다. 즉,

$$\delta_k = T_k - O_k \tag{9}$$

$$W_{kj}(t+1) = W_{kj}(t) + \eta_2\, \delta_k\, O_j + \alpha_1 \triangle W_{kj}(t-1) \tag{10}$$

$$\delta_j = O_j(1-O_j) \sum_{k=0}^{N_o-1} W_{kj}\, \delta_k \tag{11}$$

$$W_{ji}(t+1) = W_{ji}(t) + \eta_1\, \delta_j\, O_i + \alpha_2 \triangle W_{ji}(t-1) \tag{12}$$

이다. 여기서 t는 반복 학습횟수이고, η_1과 η_2는 학습률, α_1과 α_2는 모멘텀, $\triangle W_{kj}$는 은닉층과 출력층 사이의 가중치의 갱신량, $\triangle W_{ji}$는 입력층과 은닉층 사이의 가중치의 갱신량을 각각 나타낸다.

Ⅲ. 시뮬레이션 결과 및 고찰

제안한 유형분류 방법의 타당성과 성능을 확인하기 위해 HP ScanJet Ⅲ-P를 이용해서 대한전자공학회 논문지에서 선택한 다양한 명암도 문서영상에 대해 시뮬레이션 하였다. 제안된 방법에서는 먼저 입력된 명암도 영상에 대해 Prewitt 연산자를 적용하여 에지화소의 세기와 방향을 추출한 후, 임계치 기법을 적용하여 약한 세기의 에지화소를 제거함으로써 문서 뒷면으로부터의 배경잡음을 제거하였다.

블록 분할할 때 연속부 길이 한정 알고리즘을 개선된 방법으로 적용하여, 기존의 방법에서 단점으로 지적되던 많은 메모리 요구량을 감소시켰고, 속도 또한 증가시켰으며, 필요 이상으로 작은 블록들로 나뉘어지던 현상도 제거하였다. 기존의 방법으로 블록 분할한 결과와 개선된 방법으로 블록 분할한 결과 비교는 그림 8에서와 같다.

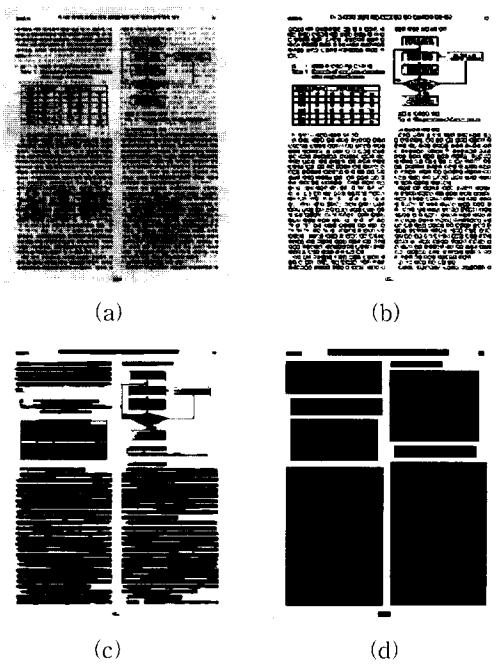


그림 8. 문서영상에 대해 블록 분할한 결과 (a) 원래의 영상 (b) 에지를 구한 영상 (c)기존의 방법으로 블록 분할한 영상 (d) 제안된 방법으로 블록 분할한 영상
Fig. 8. Block segmentation of a document image (a) an example of a document image, (b) the Prewitt operated image, (c) the result of the conventional segmentation, and (d) the result of the proposed segmentation.

그림을 통해 개선된 방법으로 블록분할 한 경우에는 문단 단위로 자르는 효과가 있음을 확인할 수 있다. 특

징추출 단계에서는 임계치 기법을 적용한 영상에서 에지 세기의 평균, 블록의 크기로 정규화된 에지화소수, 주된 수평-수직 방향 에지화소 함유율 및 주된 사선 방향 에지화소 함유율을 특징으로 추출한다. 추출된 4가지 특징값을 0과 1사이로 정규화한 후 역전파 신경회로망에 입력시킴으로써 큰 문자, 작은 문자, 수식, 표, 순서도, 그래프 및 사진블록의 7가지 블록으로 분류하도록 하였다. 유형분류를 위해 사용된 역전파 신경회로망의 입력층, 은닉층, 출력층의 뉴런 수는 각각 4, 30, 7로 하였고 학습률은 0.3, 모멘텀은 0.5로 하였다. 제안한 알고리즘을 두 가지 문서영상 A 및 B에 대해 적용하여 블록분할 및 유형 분류한 결과는 그림 9에서와 같다.

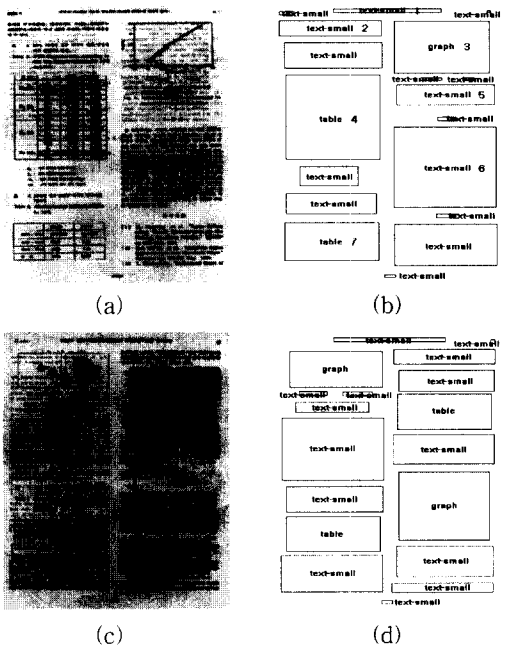


그림 9. 블록분할 및 유형분류 결과 (a) 입력영상 A와 (b) 그의 결과 및 (c) 다른 입력영상 B와 (d) 그의 결과

Fig. 9. Results of the block segmentation and classification (a) original image A, (b) its result, (c) the original image B, and (d) its result.

그림을 통해 제안한 방법은 문단 단위로 잘 분할하며 따라서 이를 기반으로 유형분류를 효과적으로 할 수 있음을 알 수 있다. 그림 9(b)에 나타낸 여러 블록들 중 일련번호가 붙은 7개의 블록들에 대한 특징값들은 표 2와 같다.

표 2. 각 블록에 대한 F_1~F_4 특징값

Table 2. The value of the F_1~F_4.

Block number	F1	F2	F3	F4
1 (text-small)	119	0.28	0.05	0.01
2 (text-small)	120	0.26	0.04	0.01
3 (graph)	118	0.08	0.12	0.19
4 (table)	119	0.23	0.53	0.01
5 (text-small)	117	0.27	0.05	0.01
6 (text-small)	119	0.29	0.04	0.01
7 (table)	120	0.20	0.59	0.01

본 실험에서는 각 블록 유형별로 그 유형에서 발생되는 모든 형태를 포함하는 학습블록 10개씩과 시험블록 10개씩, 총 140개의 블록을 대상으로 학습 및 시험하였고, 각 유형별 분류율에 대한 결과는 표 3과 같다.

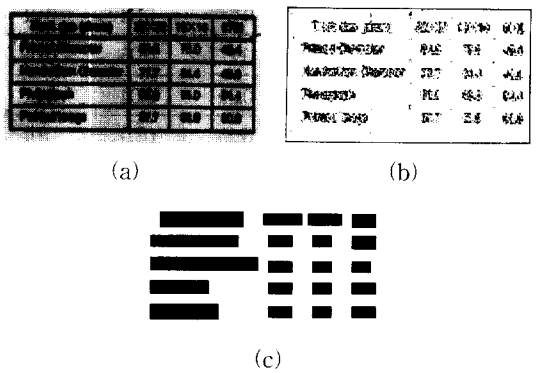
표 3. 제안한 방법의 유형분류 결과

Table 3. The result of the block classification by the proposed method.

Categories of Blocks	Number of Blocks (learning)	Number of Blocks (testing)	Recognition Ratio
Small Letters	10	10	20/20
Large Letters	10	10	20/20
Equations	10	10	20/20
Flow Charts	10	10	20/20
Tables	10	10	20/20
Graphs	10	10	18/20
Photographs	10	10	20/20

시뮬레이션 결과, 표 3에서 보는 바와 같이 그래프블록에 대한 일부 오분류를 제외하고는 나머지 모든 블록에 대해서는 정확한 분류가 가능하였다. 오분류가 발생하는 그래프블록은 주로 수평-수직방향 에지로 구성된 막대그래프와 매우 복잡하면서도 사선 방향의 선분이 없는 블록들이었다. 이와 같이, 특징 F_2 및 특징 F_4에 의해 주로 분류되는 그래프블록에서는 이들 특징값을 크게 하는 성분이 없는 유형의 그래프는 오분류됨을 알 수 있었다. 매우 복잡한 그래프는 비문자블록인 사진블록으로 분류되었고, 막대그래프는 약간의 문자와 수평, 수직 선분으로 구성되어 있으므로 표블록으로 오분류되었다. 따라서, 막대그래프블록이나 매우 다양한 유형을 가지는 그래프블록에 대한 오분류를 막을 특징추출 방법의 연구가 더 진행되어야 할 것으로 본다.

또한, 특징 F_3를 추출한 후에는 문자 부분만이 제거되는 현상이 있으므로 이 특징 F_3 추출과정에서 표나 순서도블록에서 문자와 선분을 분리하고 문자열만 용이하게 추출할 수 있음도 확인하였다. 이를 다시 확인하기 위해 세 번째 특징을 추출하는 과정을 그림 10에 표시하였다.



(a) (b)

(c)

그림 10. F_3 특징추출 과정에서의 문자열 추출 (a) 표블록 (b) 제거된 성분들 (c) 추출된 문자열

Fig. 10. Character string extraction during the F_3 extraction (a) the table block, (b) the removed element, and (c) the extracted character string position.

그림에서 보는 바와 같이 표나 순서도로부터 특징 F_3를 추출하는 과정에서 문자 부분들은 제거되므로 이를 따로 저장하여 문지르기 및 잡음제거를 한 후 블록분할을 하면 문자열의 위치도 추출할 수 있음을 알 수 있다.

Ⅳ. 결 론

본 논문에서는 문서영상의 에지 정보를 이용한 효과적인 블록분할 및 유형분류 알고리즘을 제안하였다. 제안한 알고리즘에서는 에지화소들의 세기와 방향 분포로부터 에지 세기의 평균, 블록의 크기로 정규화된 에지화소수, 주된 수평-수직 방향 에지화소 함유율 및 주된 사선 방향 에지화소 함유율의 네 가지를 특징으로 추출하여 블록내의 문자와 선분의 구성비를 나타내게 함으로써 배경잡음과 밝기변화에 둔감한 유형분류가 가능하도록 하였다. 이러한 특징들을 역전파 신경회로망에 입력시켜 문서 영상을 큰 문자, 작은 문자, 수식, 표, 순서도와 같은 5개의 문자블록과 그래프, 사진과 같은 2개의 비문자블록을 포함한 7가지 블록으로 상세하게 유형분류 하였다. 한편, 블록 분할할 때 실제 문서의 단 간격과 줄 간격을 도입하여 연속부 길이 한정 알고리즘을 적용함으로써 중간결과 보관을 위한 별도의 메모리를 요구하지 않으며, 적용 횟수도 줄어 적은 메모리로도 효율적인 블록분할도 가능하도록 하였다.

제안한 블록분할 및 특징추출 알고리즘의 타당성을 확인하기 위하여 여러 형태의 다양한 명암도 문서영상에 대하여 실험 해 본 결과 그 내용 변화가 다양한 그래프블록의 분류율이 90%로 낮을 뿐 나머지 모든 블록들은 정확하게 분류해 낼 수 있었다. 다만 다양한 유형을 가지는 그래프블록의 오분류를 줄이기 위한 특징추출 방법은 앞으로 좀 더 연구되어야 하겠다.

참 고 문 헌

[1] 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.

[2] F. M. Wahi, K. Y. Wong, and R. G. Casey, "Block segmentation and text extraction in mixed text/image documents," *Computer Graphics and Image Processing,* vol. 22, pp. 375-390, Feb. 1982.

[3] 함영국, 김인관, 정홍규, 박래홍, 이창범, 김상중, 윤병남, "텍스트와 그래픽으로 구성된 혼합문서 인식에 관한 연구," *대한전자공학회논문지,* vol. 31, no. 7, pp. 76-89, July 1994

[4] S. Imade, S. Tatsuta, and T. Wada, "Segmentation and classification for text/image documents using neural network," *Proc. of the Second International Conference on Document Analysis and Recognition*, Tsukuba, Japan, pp. 930-934, Oct. 1993.

[5] J. S. Kim, J. C. Shim, J. H. Lee, and H. M. Choi, "Classification of document image blocks based on textual features and BP," *ISPACS '94,* Seoul, Korea, pp. 104-108, Oct. 1994.

[6] T. Akiyama and N. Hagita, "Automated entry system for printed documents,"

Pattern Recognition, vol. 23, no. 11, pp. 1141-1154, 1990.

[7] Y. Hirayama, "A block segmentation method for document images with complicated column structures," *Proc. of the Second International Conference on Document Analysis and Recognition*, pp. 91-94, Tsukuba, Japan, Oct. 1993.

[8] L. A. Fletcher and R. Kasturi, "A robust algorithm for text string separation from mixed text/graphics images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 10, no. 6, pp. 910-918, Nov. 1988.

[9] 백영묵, 임길택, 김우태, 진성일, "영역 레이블링 방식을 이용한 일반 문서의 Layout Understanding," *제 6회 영상처리 및 이해에 관한 워크샵 논문집*, 경주, pp. 204-209, Jan. 1994

[10] D. X. Le and G. R. Thoma, "Document classification using connectionist models," *Proc. of IEEE International Conference on Neural Networks*, Orlando, Florida, vol. 5, pp. 3009-3014, June 1994.

[11] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Addison Wesley, New York, 1992.

[12] P. D. Wasserman, *Neural Computing*, Van Nostrand Reinhold, New York, 1989.

저 자 소 개

朴 昶 俊(正會員)

1971년 6월생. 1994년 2월 경북대학교 전자공학과 졸업(공학사). 1996년 2월 경북대학교 전자공학과 석사과정 졸업(공학석사). 1996년 ~ 1996년 10월 현재 경북대학교 전자공학과 박사과정 재학중. 주관심분야는 문서인식, 패턴인식 및 신경회로망 등임

崔 興 文(正會員) 第 33卷 B編 第 2號 參照

全 俊 亨(正會員) 第 33卷 B編 第 2號 參照