Journal of the Korea Industrial Information Systems Research Vol.28 No.2, Apr. 2023 :11-19
http://dx.doi.org/10.9723/jksiis.2023.28.2.011 ISSN:1229-3741

문서 처리 자동화를 위한 인보이스 이미지의 구조 인식 방법.+

(Structure Recognition Method of Invoice Document Image for Document Processing Automation)

이 동 석[1], 권 순 각[2]*
(Dong-seok Lee and Soon-kak Kwon)

요 약 본 논문은 인보이스 문서 이미지에 문서 처리 자동화를 적용하기 위한 문서 구조 인식 방법과 문서 구조 인식 결과를 토대로 스프레드문서 형태로 출력하는 방법을 제안한다. 딥러닝 OCR 엔진을 통해 문서 내 단어 블록들과 해당 블록들의 문자 인식 결과를 얻는다. 단어 블록의 위치 정보들을 통해 같은 행과 같은 열에 존재하는 단어 블록들을 검출한다. 단어 블록들의 배치 정보를 통해 문서 영역을 분할한다. 문서의 구역 정보를 통해 얻어진 문서 구조를 토대로 스프레드시트의 알맞은 위치에 문자 인식 결과를 입력한다. 실험 결과 제안된 방법을 통한 항목 배치는 평균 92.30%의 정확도를 보인다.

핵심주제어: 문서 구조 인식, 문서 처리 자동화, 광학 문자 인식

Abstract In this paper, we propose the methods of invoice document structure recognition and of making a spreadsheet electronic document. The texts and block location information of word blocks are recognized by an optical character recognition engine through deep learning. The word blocks on the same row and same column are found through their coordinates. The document area is divided through arrangement information of the word blocks. The character recognition result is inputted in the spreadsheet based on the document structure. In simulation result, the item placement through the proposed method shows an average accuracy of 92.30%.

Keywords: Document structure detection, Document processing automation, Optical character recognition

1. 서 론

현재 디지털 시대로 돌입함에 있어 많은 업무 처리가 디지털 문서 형태로 진행되고 있지만, 아직도 인쇄 문서 형태로 업무를 처리하는 업무 형태도 적지 않다. 인쇄 문서를 처리하기 위해 사람이 직접 인쇄된 문서를 그대로 시스템에 기

* Corresponding Author: skkwon@deu.ac.kr
+ 본 논문은 2022년도 BB21+ 사업으로 지원되었으며, 또한 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업의 연구결과로 수행되었음(IITP-2023-2020-0-01791).
Manuscript received March 27, 2023 / revised April 24, 2023 / accepted April 25, 2023
1) 동의대학교 인공지능그랜드ICT연구센터, 제1저자
2) 동의대학교 컴퓨터소프트웨어공학과, 교신저자

입한다. 이 때 타이핑 실수로 특정 항목의 내용을 다른 항목에 기입하든지, 오타가 발생할 수도 있다. 특히 문서 처리 작업은 반복적이고 지속적인 작업이라는 특징이 있어 해당 휴먼에러가 일어날 확률이 높다. 타이핑 실수로 인한 오기입의 경우에는 특정 항목에서는 무결성 검사를 통해 해당 항목에 어울리지 않는 내용을 사전에 검사할 수도 있지만, 이는 모든 항목에 적용할 수 없다는 문제가 있다. 특히 금액란에서의 타이핑 실수는 중대한 업무 처리의 장애를 일으킬 뿐만 아니라 회사에 손실을 가져올 수 있다. 문서 처리 과정을 자동화하는 것은 이러한 위험성을 크게 줄일 수 있지만, 해당 기술에 대한 직원들의 낮은 신뢰도 등의 요인들로 인해 문서 처리 분야에서 자동화 전환이 느리다.

이미지 내 문자들을 인식하는 OCR(광학 문자 인식: Optical character recognition)은 다층의 인공신경망을 통해 문제를 해결하는 딥러닝의 적용을 통해 정확도가 대폭 증가하였다(He et al., 2018; Shi et al., 2019; Feng et al., 2021). 하지만 문서 처리의 자동화를 위해서는 이미지 내 각 항목들간 관계를 인식하는 것도 중요하다. 일례로 문서 처리에서 자주 보이는 표의 경우에는 각각의 항목들의 위치 관계도 크게 중요하지만, OCR만을 통해서는 이러한 위치관계를 처리할 수 없고, 별도의 후처리 과정이 더 필요하다. 단일 문서 형태만을 처리하는 업무 형태에서는 각각의 항목의 영역들을 지정함으로써 문서 자동화를 쉽게 구현할 수 있다. 하지만 다양한 문서 형태를 처리해야 하는 경우에는 해당 방법을 사용하기 매우 어렵다. 이를 해결하기 위해 컨볼루션 레이어(CNN: Convolution Nerual Network)가 포함된 인공신경망을 통해 표 등의 문서 내 구조를 인식하는 연구가 여러 이루어졌지만 CNN의 한계로 인해 다양한 형태의 문서에 적용하는 데 한계가 있다.

이전 연구(Lee et al., 2022)에서는 문서 내 다양한 형태를 가지는 표의 구조를 인식하는 연구를 수행하였다. 본 논문에서는 표에 국한되지 않고 다양한 형태의 인보이스 문서 전체에 대해 문서 구조를 인식하는 방법을 제안한다. 먼저 딥러닝 OCR 엔진을 통해 이미지 형태의 인보이스 문서를 인식한다. 그 후 각 항목들의 가로, 세로 위치들을 통해 같은 행 및 열에 위치한 항목들을 찾은 후, 이를 바탕으로 문서 영역을 분할한다. 검출된 위치 정보를 통해 각각의 항목들의 스프레드시트 문서 내 행과 열 인덱스를 지정하고, 해당 항목의 내용을 기입한다.

2. 기존 딥러닝을 통한 문서 구조 인식

CNN 기반의 가상신경망을 통해 문서 구조를 인식하는 연구가 다음과 같이 수행되었다. CascadeTabNet(Prasad et al., 2020)은 Cascade Mask R-CNN(Cai et al., 2018)를 통해 표와 표 구조를 인식한다. Cascade Mask R-CNN는 영역 제안 네트워크(RPN, Region Proposal Network)에서 검출된 관심 영역이 실제 객체인지 여부를 판가름하는 기준인 IoU(Intersection over union)를 서서히 높여가면서 학습을 하는 기법이다. Smock et al.(2022)은 약 백만개의 표가 포함된 문서 영상 데이터셋을 통해 네트워크를 학습하여 표 구조를 인식한다. 이 때 표를 인식하기 위해 Faster R-CNN(Shaoqing et al., 2017) 또는 DETR(Carion et al., 2020)의 딥러닝 객체 인식 네트워크를 사용한다. PubTabNet(Zhong et al., 2020)은 디코더-인코더 구조를 통해 영상의 이미지를 지정된 표 형태로 재구성하는 네트워크이다. 하지만 CNN 기반의 문서 구조 인식 방법은 CNN의 구조적인 문제로 인해 구조 인식에 한계가 있다. CNN은 정사각형 형태의 영역 내의 국소적인 화소 패턴 특징을 추출한다. 객체 검출에서는 각 객체의 고유한 패턴을 추출하여 인식함으로써 높은 정확도로 객체를 검출할 수 있다. 하지만 문서 구조 인식에서는 이미지의 형태적인 패턴보다 각 항목 객체들의 위치 관계에 집중해야 한다. 하지만 CNN은 이러한 위치 관계를 검출하는 데 어려움을 겪는다. 또한 다양한 형태의 문서 구조는 CNN을 통한 구조 인식을 어렵게 만드는 또 하나의 요인이다. CNN 기반의 인공신경망으로 Fig. 1 (a)와 같이 직선으로 항목이 구분되는 일반적인 표는 인식이 잘되지만, Fig. 1 (b)와 같

이 항목이 직선이 아닌 공백으로 구분되어지고, 각 항목간의 폭이나 너비가 균일하지 못한 표에 대해서는 CNN을 통해 표나 항목을 인식하기 매우 어렵다. 본 논문에서는 이러한 문제를 해결하기 위해 이전에 인식된 항목들의 영역 및 위치를 기반으로 하여 표 구조 형태를 인식하여 스캔된 문서 이미지를 디지털 문서화하는 방법을 제안하고자 한다.

Id	Name	Email	Investments
231	Albert Master	albert.master@gmail.com	Bonds
210	Alfred Alan	aalan@gmail.com	Stocks
256	Alison Smart	asmart@biztalk.com	Residential Property
211	Ally Emery	allye@easymail.com	Stocks
248	Andrew Phips	andyp@mycorp.com	Stocks
234	Andy Mitchel	andym@hotmail.com	Stocks
226	Angus Robins	arobins@robins.com	Bonds
241	Ann Melan	ann_melan@iinet.com	Residential Property
225	Ben Bessel	benb@hotmail.com	Stocks
235	Bensen Romanolf	benr@albert.net	Bonds

(a)

POS	Quantity	Unit	Article No. Description	Price / 100 Unit EUR	Amount EUR
			Delivery-No:	/ Delivery Date:	
1	*	Piece	***.**.***	*,**	*,**
2	*	Piece	***.**.***	*,**	*,**
3	*	Piece	***.**.***	*,**	*,**
4	*	Piece	***.**.***	*,**	*,**
5	*	Piece	***.**.***	*,**	*,**
				Carry forward	*,**

(b)

Fig. 1 Table structure types: (a) table with dividing lines and (b) table without dividing line

3. 문서 처리 자동화를 위한 인보이스 문서 이미지 인식

본 논문에서는 이미지 형태의 인보이스 문서에서 문자를 인식하고, 문서 구조를 위치 관계를 통해 인식하여 이를 스프레드시트 문서 형태로 출력하는 방법을 제안한다. 먼저 문서 내 각 항목들의 위치와 텍스트를 딥러닝 OCR를 통해 인식하고, 같은 행과 열에 속한 항목들을 인식한다. 그 후 각각의 항목들이 스프레드시트 문서에 위치할 위치를 결정하고, 인식된 항목들을 입력하여 출력한다. Fig. 2는 제안하는 방법의 흐름도를 보인다.

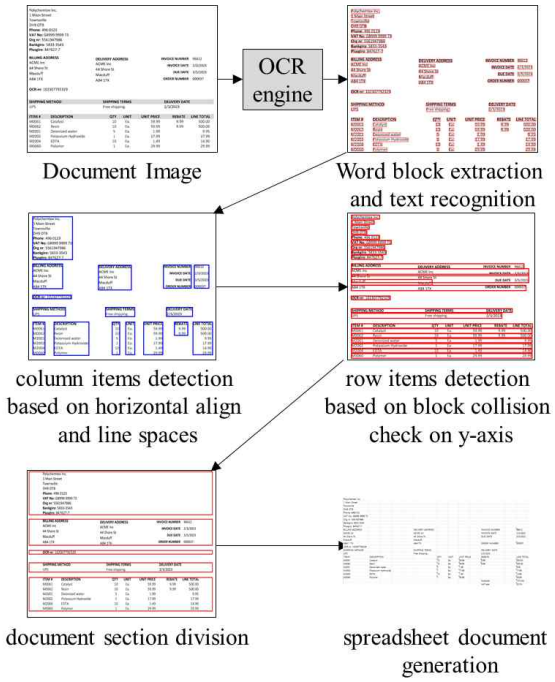


Fig. 2 Flow of proposed method

3.1 인보이스 문서 이미지에서 단어 블록 추출 및 문자 인식

먼저 단어 블록 및 블록 내 문자들을 딥러닝 OCR엔진을 통해 인식한다. 본 논문에서는 단어 블록을 검출하기 위한 네트워크로 DBNet(Liao et al., 2020) 를 사용하고, 단어 블록들 내 문자들을 인식하기 위한 네트워크로 VGG16을 백본으로 하는 CRNN(Convolutional to recurrent neural network) 구조의 네트워크(Shi et al., 2016)를 사용한다. Fig. 3은 OCR엔진을 통해 단어 블록들을 추출한 결과이다.

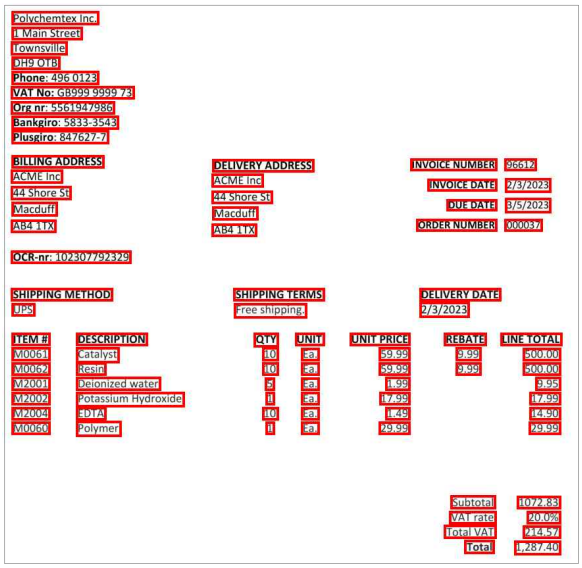


Fig. 3 Word block detection result

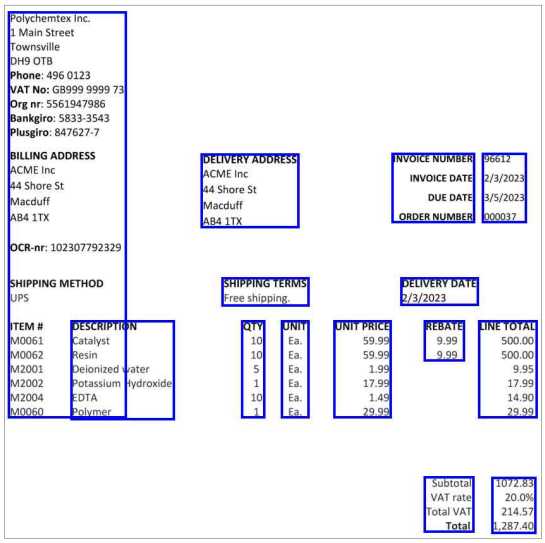
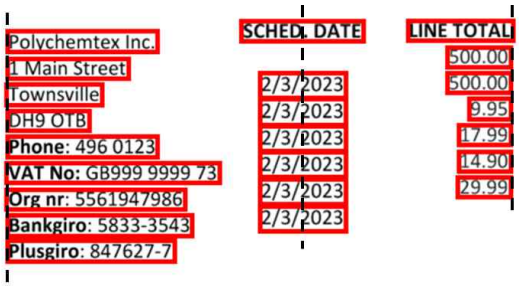


Fig. 5 Detection results of word blocks on same column

3.2 열 항목 검출

검출된 단어 블록에서 같은 열에 위치하는 블록들을 검출한다. 같은 열에 위치하는 단어 블록들을 검출하는 기준으로는 단어 블록이 차지하는 영역의 X좌표를 기준으로 한다. Fig. 4와 같이 두 단어 블록들의 맨 왼쪽 끝, 오른쪽 끝, 또는 그 가로 중앙중 하나의 X좌표가 일치한다면 두 개의 단어 블록은 같은 열에 속한다고 판단하고 같은 항목에 추가한다. Fig. 5는 해당 방법을 통해 같은 열에 위치하는 단어 블록들을 검출한 결과이다.



(a) (b) (c)

Fig. 4 Detection of word blocks on a column: (a) left aligned, (b) center aligned, and (c) right aligned

같은 열을 검출하는 방법을 통한 항목 검출에서 Fig. 6 (a)처럼 다수의 다른 항목들이 같은 항목으로 검출되는 경우가 발생한다. 이 문제는 같은 항목에 속한 문자간 간격은 유사하고, 다른 항목에 속한 단어 블록와의 간격은 크다는 것을 이용하여 항목을 분리함으로써 해결할 수 있다. 같은 열로 검출된 단어 블록에서 인접한 두 블록 간 Y좌표의 차이에 대한 히스토그램을 계산한다. 같은 줄 간격을 가지는 항목이더라도 글꼴에 따라 다른 픽셀 거리를 가질 수 있다. 예를 들어 'g', 'p'와 같은 글자는 'a', 'b'와 같은 글자에 비해 더 아래 위치까지 화소가 위치한다. 이를 고려하여 히스토그램을 측정할 때, 3픽셀 간격으로 빈도를 측정하고, 해당 범위 내 빈도들은 모두 같은 범주로 간주한다. 그 후 최대 빈도를 가지는 간격을 기준으로 단어 블록들을 분류한다. 예를 들어 Fig. 6 (a)에 대해 인접한 항목의 X좌표 간 차이를 히스토그램을 계산한 결과는 Fig. 6 (c)와 같이 26 픽셀 간격에서 최대 빈도가 나온다. 그 후 Y좌표가 26픽셀 이상의 차이가 나는 항목들을 다른 항목으로 분리하면 Fig. 6 (b)와 같이 각 항목들이 정확하게 별개의 항목으로 검출된다.

3.3 행 항목 검출

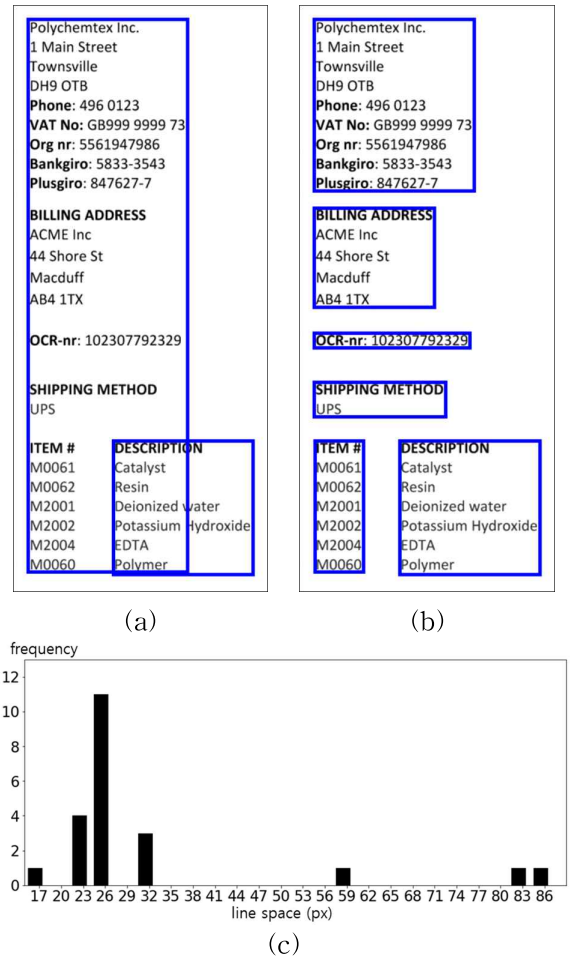


Fig. 6 Detection of column items: (a) detection based on horizontal aligned, (b) detection based on line space, and (c) histogram of line spaces

행 항목을 검출하기 위해 단어 블록 상단의 Y좌표를 기준으로 할 수 있다. 하지만 다른 항목에 속한 단어 블록 간 줄 간격이 다를 수 있다. 이는 해당 문서를 디지털화할 때 문서의 구조 파악을 힘들게 한다. Fig. 7 (a)은 각각의 행 내의 문자 줄 간격이 달라 동일한 열 내의 항목이 제대로 검출이 되지 않는 것을 보인다. 예를 들어 첫 번째 열과 두 번째 열의 'ACME Inc.' 단어 블록들은 동일한 행 항목이지만, 다른 Y좌표를 가지므로 각기 다른 행으로 검출된다. 그 결과 해당 영역은 총 12개의 행을 가지는 항목으로 검출된다.

이러한 문제점을 해결하기 위해 Fig. 7 (b)와 같이 Y축을 기준으로 두 단어 블록이 겹치는지를 측정하여 같은 열 항목을 검출한다. 이 때 한 단어 블록에서 다른 열에 속한 두 개 이상의 항목들이 겹치는 경우도 있다. 이러한 경우에는 최상단에 있는 단어 블록을 같은 열 항목으로 검출한다. Fig. 7 (c)는 제안된 방법을 통해 실제와 같이 5개의 행을 가지는 항목으로 검출되는 것을 보인다.

3.4 문자 구역 분할

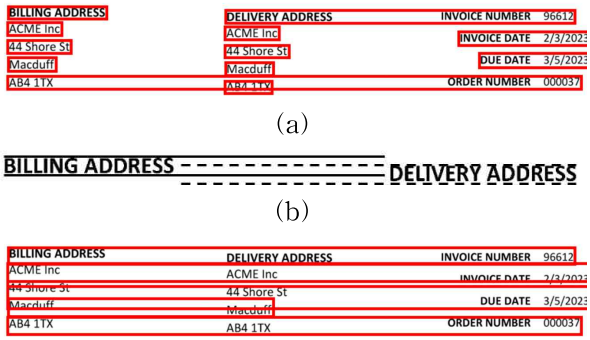


Fig. 7 Row item detection: (a) detection based on y-coordinate, (b) y- coordinate collision between two word blocks, and (c) detection based on y-coordinate collision

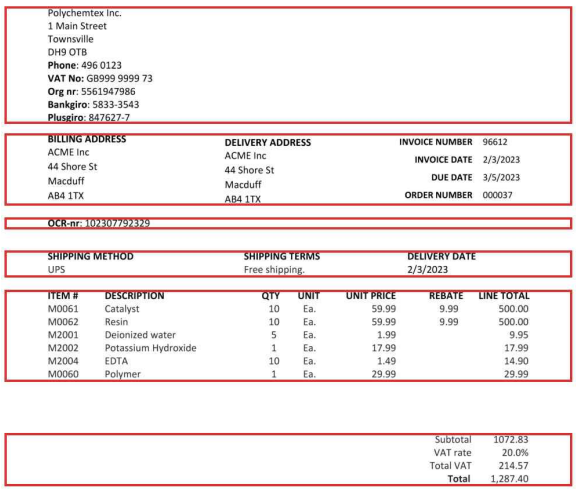


Fig. 8 Document division result

같은 행 및 열에 속한 항목들의 정보를 바탕으로 문서의 구역을 가로로 분할한다. 각 구역 내에는 하나 이상의 온전한 열 항목들이 속하도록 구분되어진다. Fig. 8은 문서의 구역을 분할한 결과를 보인다.

3.5 스프레드시트 문서 생성

인식된 단어 블록들의 위치정보와 문자 인식 결과를 통해 스프레드시트 문서를 생성한다. 이 때 스프레드시트는 정수의 행과 열의 인덱스를 가지는 위치에 텍스트를 입력할 수 있다. 이 때 행의 인덱스는 위에서 아래로 순차적으로 번호를 부여한다. 열의 정보는 해당 단어 블록이 속한 열 항목의 좌상단 X좌표를 이용한다. 이 때 스프레드시트 문서가 C개의 열을 사용한다고 한다면, 좌상단 점의 X좌표가 x인 열 항목에 속한 단어 블록의 열 인덱스 c는 다음과 같이 정할 수 있다. 식 (1)에서 W는 문서 이미지의 폭을 의미하고, $\lfloor\ \rfloor$ 연산자는 소수점을 버리는 연산이다. 이 때 만약 c 인덱스에 이미 할당된 열이 있다면, 다음 열 인덱스로 정한다. Fig. 9는 제안된 방법을 통해 생성된 스프레드시트 문서이다.

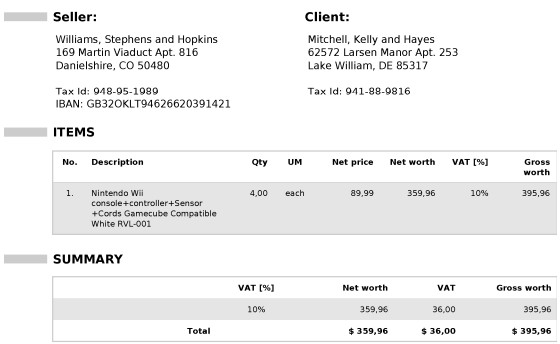
$$c = \lfloor \left(\frac{x \times (C+1)}{W} \right) \rfloor \qquad (1)$$

	A	B	C	D	E	F	G
1	Polychemtexlnc.						
2	1Main Street						
3	Townsville						
4	DH9 OTB						
5	Phone: 4960123						
6	VAT No: GB9999999 73						
7	Org nr: 5561947986						
8	Bankgiro: 5833-3543						
9	Plusgiro: 847627-7						
10							
11	BILLING ADDRESS	DELIVERY ADDRESS				INVOICE NUMBER	96612
12	ACME Inc	ACMEInc				INVOICE DATE	2/3/2023
13	44 Shore St	44 Shore St				DUEDATE	3/5/2023
14	Macduff	Macduff					
15	AB41TX	AB41TX				ORDER NUMBER	000037
16							
17	OCR-nr: 102307792329						
18							
19	SHIPPING METHOD	SHIPPINGTERMS				DELIVERY DATE	
20	UPS	Free shipping.				2/3/2023	
21							
22	ITEM#	DESCRIPTION	QTY	UNIT	UNIT PRIC	REBATE	LINETOTAL
23	M0061	Catalyst	10	Ea.	59.99	9.99	500.00
24	M0062	Resin	10	Ea.	59.99	9.99	500.00
25	M2001	Deionized water	5	Ea.	1.99		9.95
26	M2002	Potassium Hydroxide	1	Ea.	17.99		17.99
27	M2004	EDTA	10	Ea.	1.49		14.90
28	M0060	Polymer	1	Ea.	29.99		29.99
29						Subtotal	1072.83
30						VAT rate	20.0%
31						Total VAT	214.57
32						Total	1,287.40
33							
34	Polychemtex Inc.	Phone: 4960123				Bankgiro: 5833-3543	
35	1Main Street	Fax: 4960124				Plusgiro: : 847627-7	
36	Townsville	nlo@polycnentex.om					
37	DH9 OTB						

Fig. 9 Generated spreadsheet document through proposed method

4. 실험결과

본 논문에서는 실험 데이터로 인보이스 전자 문서 데이터셋(Kozłowski et al., 2021)의 모든 80장의 문서 이미지를 사용한다. 또한 520장의 문서 이미지를 가지는 RVL-CDIP 데이터셋(Harley et al., 2015)에서 활용하고자 하는 인보이스 문서와 유사한 형식을 가진 80개의 이미지를 사용한다. 실험에 쓰인 전자 문서 데이터셋과 RVL-CDIP 데이터셋에서 총 단어 블록은 각각 약 5500개, 6000개이다. 인보이스 전자 문서



Seller:

Williams, Stephens and Hopkins
169 Martin Viaduct Apt. 816
Danielshire, CO 50480

Tax Id: 948-95-1989
IBAN: GB32OKLT94626620391421

Client:

Mitchell, Kelly and Hayes
62572 Larsen Manor Apt. 253
Lake William, DE 85317

Tax Id: 941-88-9816

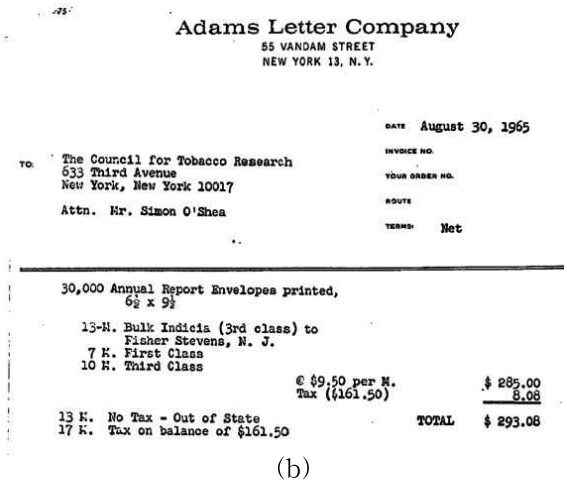
ITEMS

No.	Description	Qty	UM	Net price	Net worth	VAT [%]	Gross worth
1.	Nintendo Wii console+controller+Sensor +Cords Gamecube Compatible White RVL-001	4,00	each	89,99	359,96	10%	395,96

SUMMARY

	VAT [%]	Net worth	VAT	Gross worth
	10%	359,96	36,00	395,96
Total		$ 359,96	$ 36,00	$ 395,96

(a)



Adams Letter Company
55 VANDAM STREET
NEW YORK 13, N.Y.

DATE August 30, 1965
INVOICE NO.
YOUR ORDER NO.
ROUTE
TERMS Net

TO: The Council for Tobacco Research
633 Third Avenue
New York, New York 10017
Attn. Mr. Simon O'Shea

30,000 Annual Report Envelopes printed,
6¼ x 9½
13-M. Bulk Indicia (3rd class) to
Fisher Stevens, N. J.
7 M. First Class
10 M. Third Class
@ $9.50 per M. $ 285.00
Tax (@161.50) 8.08
13 M. No Tax - Out of State TOTAL $ 293.08
17 M. Tax on balance of $161.50

(b)

Fig. 10 Invoice document image dataset for simulation: (a) samples of electronic invoices and (b) RVL-CDIP dataset

는 전자 문서 형태의 인보이스 문서를 그대로 이미지화한 것으로, 문서의 글씨가 또렷하고, 이미지의 왜곡이나 잡음이 없다. 반면 RVL-CDIP 데이터셋은 회전되어 스캔된 이미지가 있고, 잡음이 포함된 데이터셋이다. Fig. 10은 두 데이터셋의 문서 형태를 보인다.

Fig. 11은 단순하게 각 단어 블록의 좌상단 x, y좌표들을 기준으로 스프레드시트 문서 내 인덱스를 결정했을 때와 제안된 방법을 적용했을 때를 비교한 것이다. 단순히 단어 블록의 좌상단 좌표를 기준으로 각 항목의 스프레드시트 인덱스를 결정했을 경우, 문서의 각 항목간의 관계가 고려되지 않아 같은 행, 열 항목이 생성된 스프레드시트 문서 다른 행, 열 인덱스에 위치한다. 하지만 제안된 방법을 적용했을 때는 같은 행과 열에 대한 검출이 정확하게 되어 정확한 스프레드시트 문서가 생성된 것을 보인다.

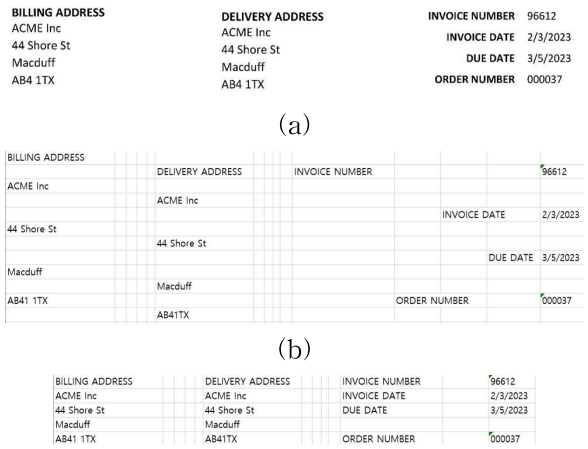


Fig. 11 Result of spreadsheet document generation: (a) original document image, (b) based on top-left point coordinates of word blocks, and (c) by applying proposed method.

제안된 방법을 통해 생성된 스프레드시트 문서에서 생성된 스프레드시트가 원래 문서 구조를 유지하는지를 실험한다. 단어 블록 배치의 정확도도 단어 블록 단위로 측정한다. 단어 블록마다 상대적인 단어들 간의 위치 관계를 비교하여 원본에서의 위치 관계가 유지되었는지를 확인한다. 원본 이미지에서 특정 단어 블록의 우측에 다른 단어 블록이 있는데, 인식 결과에서는 하단에 배치되는 경우 등을 잘못된 배치로 판단한다. Table 1은 제안된 방법을 통해 스프레드시트 문서를 생성할 때의 항목 배치 정확도를 측정한 결과를 보인다. 항목 배치 정확도는 각각의 문자 항목들에 대해 인접한 오른쪽과 아래쪽의 단어 블록이 생성된 스프레드시트 문서에서 각각 같은 행 및 열에 위치하였는지의 대한 여부를 측정하여 이를 전체 단어 블록의 개수로 나눈 것이다. 각 단어 블록의 좌상단 좌표를 기준으로 항목의 스프레드시트 인덱스를 결정했을 경우에는 전자 문서에서는 43.7%, RVL-CDIP 데이터셋에서는 20.2%로 매우 낮게 나타났다. 하지만 제안된 방법을 적용했을 때는 전자 문서에서는 98.7%, RVL-CDIP의 경우에는 85.9%의 정확도로, 그 성능이 확연히 개선된 것을 보인다. RVL-CDIP의 경우에는 문서가 정위치에서 스캔되지 않고, 약간의 회전이 포함되어 있어 정확한 항목 배치가 되지 않는 경우가 Fig. 12와 같이 일부 발생하였다.

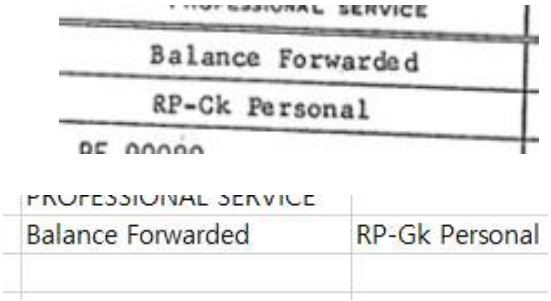


Fig. 12 Wrong word block placement due to rotated document image

Table 1 Accuracies of word block placement

spreadsheet index determination method	electronic invoices	RVL-CDIP
based on top-left point	43.7%	20.2%
proposed method	98.7%	85.9%

5. 결 론

본 논문에서는 다양한 형식의 인보이스 문서에 대해서 해당 문서 구조를 인식할 수 있는 방법을 제안하였다. 먼저 딥러닝 OCR 엔진을 통해 문서 이미지에서 단어 블록들의 영역 위치와 문자들을 인식하였다. 단어 블록의 위치를 기반으로 하여 같은 행과 같은 열에 위치한 단어 블록들을 검출한다. 단어 블록의 행과 열의 정보를 통해 문서 영역을 분할한 후, 분할된 각 구역에 대해 문자 블럭들의 스프레드시트에서의 행과 열 인덱스를 결정하였다. 제안된 방법은 올바르게 스캔된 문서에 대해서는 정확하게 문서 구조를 인식하지만, 회전되거나 왜곡되어 스캔된 문서에 대해서는 정확도가 상대적으로 떨어진다. 이러한 문제는 문서 이미지의 전처리를 통해 해당 문제를 해결할 수 있다. 제안된 방법을 통해 다양한 문서 구조를 가지는 인보이스 문서들을 인식할 뿐만 아니라 다른 종류의 문서들에 대해서도 문서 구조 인식을 통한 문서 처리 자동화를 수행할 수 있다.

References

Cai, Z. and Vasconcelos, N. (2018). Cascade R-CNN: Delving Into High Quality Object Detection, *Preceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, Salt Lake City, UT, USA, pp. 6154-6162, 2018.

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., and Zagoruyko, S. (2020). End-to-End Object Detection with Transformers, *Proceedings of the European Conference on Computer Vision*, Aug. 23-28, pp. 213-229.

Feng, H., Wang, Y., Zhou, W., Deng, J., and Li, H. (2021). DocTr: Document Image Transformer for Geometric Unwarping and Illumination Correction, *Proceeding of ACM International Conference on Multimedia*, Oct. 20-24, Chengdu, China, pp. 273-281.

Harley, A. W., Ufkes, A., and Derpanis, K. G. (2015). Evaluation of Deep Convolutional Nets for Document Image Classification and Retrieval. *Proceedings of the International Conference on Document Analysis and Recognition*, Aug. 23-26, Tunis, Tunisia, pp. 991-995.

He, T., Tian, Z., Huang, W., Shen, C., Qiao Y., and Sun, C. (2018). An End-to-End TextSpotter with Explicit Alignment and Attention, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 18-23, Salt Lake City, UT, USA, pp. 5020-5029.

Kozłowski, M. and Weichbroth, P. (2021). Samples of Electronic Invoices, *Mendeley Data*. https://doi.org/10.17632/tnj49gpmtz.2.

Lee, D. S. and Kwon, S. K. (2022). Structure Recognition Method in Various Table Types for Document Processing Automation. *Journal of Korea Multimedia Society*, 25(5), 695-702. https://doi.org/10.9717/kmms.2022.25.5.69

Liao, M., Wan, Z., Yao, C., Chen, K., and Bai, X. (2020). Real-time Scene Text Detection with Differentiable Binarization. *Proceedings of the AAAI conference on artificial intelligence*, Feb. 7-12, New York, NY, USA, pp. 11474-11481

Prasad, D., Gadpal, A., Kapadni, K., Visave, M., and Sultanpure, K. (2020). CascadeTabNet: An Approach for End to End Table Detection and Structure Recognition from Image-based Documents, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, June 14-19, Seattle, Wa, USA, pp. 2439-2447.

Shaoqing, R., Kaiming, H., Girshick, R., and Sun, J. (2017). Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks, *IEEE Transection on*

Pattern Analysis and Machine Intelligence, 39(6), 1137-1149. https://doi.org/10.1109/TPAMI.2016.2577031.

Shi, B., Bai, X., and Yao, C. (2016). An End-to-end Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(11), 2298-2304. https://doi.org/10.1109/TPAMI.2016.2646371.

Shi, B., Yang, M., Wang. X., Lyu, P., Yao, C., and Bai, X. (2019). ASTER: An Attentional Scene Text Recognizer with Flexible Rectification, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(9), 2035-2048. https://doi.org/10.1109/TPAMI.2018.2848939.

Smock, B., Pesala R., and Abraham, R. (2022). PubTables-1M: Towards Comprehensive Table Extraction from Unstructured Documents, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, June 19-20, New Orleans, LA, USA, pp. 4624-4632.

Zhong, X., Bavani, E. S., and Yepes, A. J. (2020). Image-Based Table Recognition: Data, Model, and Evaluation, *Proceedings of the European Conference on Computer Vison*, Aug. 23-28, pp. 564-580.

이 동 석 (Dong-seok Lee)

- 정회원
- 동의대학교 컴퓨터소프트웨어공학과 공학석사
- 동의대학교 컴퓨터소프트웨어공학과 공학박사
- 동의대학교 인공지능그랜드ICT연구센터 연구교수
- 관심분야 : 멀티미디어 신호처리, 영상 인식

권 순 각 (Soon-kak Kwon)

- 정회원
- 경북대학교 전자공학과 공학사
- KAIST 전기및전자공학과 공학석사
- KAIST 전기및전자공학과 공학박사
- 동의대학교 컴퓨터소프트웨어공학과 교수
- 관심분야 : 영상처리, 영상딥러닝