

카메라 문서 영상의 이진화 및 기울어짐 보정 방법

장대근*, 전병태**

A Method for Thresholding and Correction of Skew in Camera Document Images

Dae-Geun Jang*, Byung-Tae Chun**

요 약

카메라 영상은 조명에 매우 민감하다 그 결과 문자 인식을 매우 어렵게 만든다. 또한 카메라로 획득한 문서 영상을 대상으로 문자인식을 수행할 경우 비네팅현상이 발생하며 카메라와 피사체간의 위치와 각도의 불일치에 의해 기학적 왜곡이 발생한다. 비네팅 효과에 의한 문자분리의 어려움과 기하학적 왜곡으로 인한 문자모양의 변형으로 인해 실제로 사용이 가능한 카메라 문자인식 기술의 개발이 쉽지 않았다. 본 논문에서는 양선형 변환을 이용하여 카메라로 획득한 문서영상의 기하학적 왜곡을 보정함으로써 문자인식 성능을 향상시킬 수 있는 방법을 제안한다. 기존에는 카메라의 위치정보 없이 기하학적 왜곡을 보정하기 어려웠으나 제안한 방법은 이진화한 문서영상을 대상으로 문자와 문자줄을 추출하여 상하로는 문자줄의 배열방향을, 좌우로는 문자획의 기울어진 각도를 측정함으로써 문서영상 자체만으로 기하학적 왜곡보정이 가능하다.

Abstract

Camera image is very sensitive to illumination that result in difficulties for recognizing character. Also Camera captured document images have not only skew but also vignetting effect and geometric distortion. Vignetting effect make it difficult to separate characters from the document images. Geometric distortion, occurred by the mismatch of angle and center position between the document image and the camera, make the shape of characters to be distorted, so that the character recognition is more difficult than the case of using scanner. In this paper, we propose a method that can increase the performance of character recognition by correcting the geometric distortion of document images using a linear approximation which changes the quadrilateral region to the rectangle one. The proposed method also determine the quadrilateral transform region automatically, using the alignment of character lines and the skewed angles of characters located in the edges of each character line. Proposed method, therefore, can correct the geometric distortion without getting positional information from camera.

▶ Keyword : 카메라 문서 영상(camera document image), 기울어짐 보정(correction of skew)

• 제1저자 : 장대근
• 접수일 : 2005.06.04, 심사완료일 : 2005.06.28
* 특허청, ** 국립 환경대학교 웹정보공학과

1. 서론

입력 장치로서 카메라의 보급 확산으로 최근 카메라 문서 영상에 관한 많은 연구가 진행되고 있다. 특히, PDA 및 휴대폰과 같은 이동 장비에 부착된 카메라 보급의 확산은 카메라 문서 영상 인식 연구를 가속화 시키는 결과를 초래했다. 이것은 영상 정보를 즉시 획득하고, 현장의 정보를 디지털 정보를 변환하는 기능을 수행 해 줄 수 있다[1-6].

카메라를 이용하여 문서영상을 획득할 경우 영상의 가운데 부분에서 가장자리로 갈수록 어두워지는 비네팅현상때문에 배경으로부터 문자의 분리가 어려우며 카메라와 피사체 간 위치와 촬영각도의 차이에 의하여 문자모양이 기하학적으로 변형되어 문자인식에서 오인식이 발생하는 문제점이 있다. 비네팅현상의 경우 렌즈 제조기술이 향상되면서 그 영향이 감소하고 있으나 카메라와 피사체간 위치와 각도의 차이에 의한 기하학적 왜곡을 보정하려면 카메라와 피사체간 상대적 위치정보를 알아야 보정이 가능하지만 이러한 정보를 카메라로부터 얻는 것은 현실적으로 어려움이 있다.

본 논문에서는 (그림 1)과 같은 과정을 거쳐서 임의의 사변형 모양을 직사각형으로 변환하는 양선형 변환 방법을 이용하여 디지털 카메라, 휴대폰 카메라로부터 획득한 문자 영상의 기하학적 왜곡을 보정함으로써 카메라를 이용한 문자인식 성능을 향상시키는 방법을 제안한다.

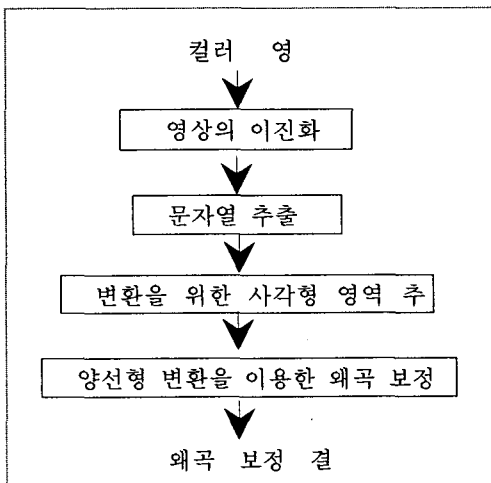


그림 1. 왜곡 보정 흐름도
Fig.1 Flow of correction for distortion

II. 영상 이진화 및 문자열 추출 방법

2.1 컬러 영상 이진화

본 논문에서는 카메라 문서의 이진화를 위하여 영역 중심 지역적 이진화 (region-oriented local thresholding) 방법을 제안하고자 한다.

2.1.1 컬러 영상의 그레이 영상 변환

컬러 영상은 보통 RGB 형태를 가지고 있으므로 식(1)과 같은 변환을 거쳐서 그레이 영상으로 변환한다[7][8].

$$Gray\ level = (R+G+B)/3 \dots\dots\dots (1)$$

2.1.2 영역 중심의 지역적 이진화 (region-oriented local thresholding)

본 논문에서는 문서 영상을 이진화하기 위하여 영역 중심의 지역적 이진화 방법을 기반으로 하는 DMM(difference of maximum and minimum intensity value) 방법을 제안하고자 한다.

카메라 문서 영상은 서브 윈도우(sub-window)에 있는 화소들의 최대값과 최소값의 차를 이용하여 이진화를 수행한다. 이 방법은 영역 중심의 이진화에 기반 하며, 좋은 이진화 수행 결과와 빠른 처리 시간 결과를 보여준다.

이진화 값 T_{mm} 은 $r \times r$ 서브 윈도우에서 자동적으로 구해진다. 카메라 문서 영상의 이진화를 위한 식은 식(2)와 같다.

$$T_{mm} = t \times (G_{high} - G_{low})$$

$$0.4 \leq t \leq 0.7 \dots\dots\dots (2)$$

여기서 G_{high} , G_{low} 는 $r \times r$ 서브 윈도우에 있는 명암도 값의 최대, 최소값이다.

2.2 문자 및 문자줄 추출

문서영상에서 영역분할 방법은 기본이 되는 화소단위에 서 시작하여 유사성을 갖는 부분을 점차적으로 크고 의미를

부여할 수 있는 단위로 단계적으로 병합하는 상향식과 문서의 전체적인 영역에서 시작하여 문서를 점점 작은 영역으로 분할하는 하향식 방법이 있다.

제한한 방법은 상향식 방법인 X. Li의 방법(9-11)을 사용하여 자소들을 분할하며 이 방법은 연속된 화소의 집합인 연결요소를 추출하고 이들을 연결하여 영역을 생성하는 방법으로 4방향 또는 8방향으로 탐색하여 연결요소를 생성하지 않고 중단위로 연결요소들을 추출한 후 인접공간 연결요소들을 병합하여 영역을 생성함으로써 계산량을 감소시켜 처리시간을 단축시킨 상향식 영역분할 방법이다.

(그림 2)는 이진화된 문자 영역, (그림 3)은 문자 영역 추출의 예를 보여주고 있다.

**대용량 디지털 비디오 처리 및 활용에 대
대용량 영상을 수작업에 의한 복원은 불
자동화된 기술이 필요
방송된 영상의 재 사용에 필요
시청자에게 시청의 편의성 제공
현재 직선 언어를 다른 직선 언어로 대**

그림 2. 카메라 문서의 이진화
Fig.2 Thresholding in camera document image

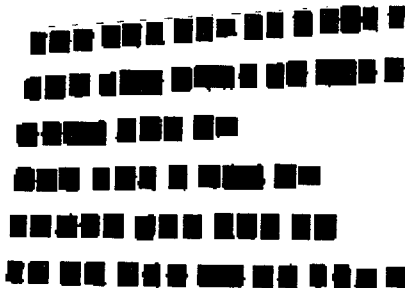


그림 3. 문자 영역 추출
Fig.3 Extracting character regions

카메라 문서영상에서의 문자줄은 기하학적 왜곡에 의하여 기울어지고 휘어지므로 이웃하는 문자들 간의 연결관계를 검사함으로써 문자줄을 확인할 수 있다.

그 과정은 분할한 가장 왼쪽 상단에 위치한 자소를 시작으로 가장 인접한 오른쪽 자소를 찾고 이 과정을 더 이상의 인접 자소가 없을 때 까지 기준위치를 오른쪽 영역으로 옮겨가며 반복함으로써 하나의 문자줄을 추출한다. 나머지 자소들을 대상으로 같은 방법을 적용함으로써 문서영상내의 모든 문자줄들을 찾아낸다.

(그림 4)은 두 자소간 교차여부를 판단하는 예로 수식 3을 만족하는 경우 두 자소는 교차 관계에 있게 된다.

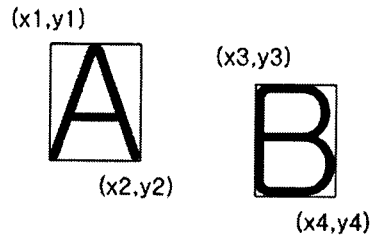


그림 4. 자소간 교차여부 검사
Fig.4 Verify crossing between characters

$$\begin{aligned}
 &(\text{Min}Y \geq \text{Max}Y) \cap (\text{Min}X < \text{Max}X) \dots\dots\dots (3) \\
 &\text{Min}Y = \min\{\max(y1, y2), \max(y3, y4)\} \\
 &\text{Max}Y = \max\{\min(y1, y2), \min(y3, y4)\} \\
 &\text{Min}X = \min\{\max(x1, x2), \max(x3, x4)\} \\
 &\text{Max}X = \max\{\min(x1, x2), \min(x3, x4)\}
 \end{aligned}$$

III. 기하학적 왜곡 보정

3.1 문서의 기울어진 영역 자동 추출

사변형을 구성하는 4개의 꼭지점 좌표는 가로방향 2개의 직선과 세로방향 2개의 직선이 서로 만나는 점을 찾음으로써 가능하다. 즉 변환영역을 둘러싸는 4개의 직선을 구함으로써 4개의 꼭지점을 찾는다.

변환영역을 둘러싸는 4개의 직선 중 2개의 가로방향 직선은 변환영역의 상, 하 2개의 최외각 문자줄과 그 구성문자들을 이용하여 구한다.

(그림 5)에서 두 문자가 하나의 문자줄을 구성하는 최외각 문자들이고 문자줄이 최상위 문자줄일 경우 직선의 기울

기는 $(x1,y1)$ 점과 $(x4,y3)$ 점간의 기울기 θ_u 가 되고 최하위 문자줄일 경우 기울기는 $(x1,y2)$ 점과 $(x4,y4)$ 점간의 기울기 θ_l 이 된다.

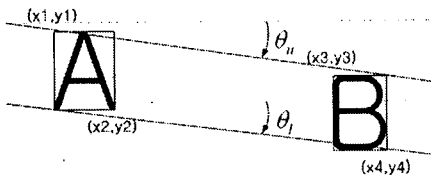


그림 5. 수평 방향 직선의 기울기 결정
Fig.5. Decision of slant for horizontal direction

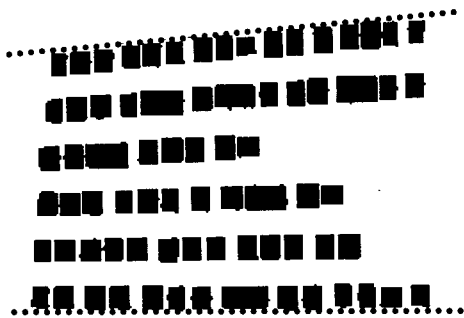


그림 6. 수평 방향 직선 결정
Fig.6 Decision of line for horizontal direction

직선은 기울기 θ_u 와 θ_l 을 갖는 각각의 직선 l_u 와 l_l 이 (그림 6)와 같이 변환영역 최외각을 지나도록 결정한다.

대부분의 문서는 가로방향으로 쓰여져 있으므로 변환영역을 둘러싸는 2개의 세로방향 직선은 각 문자줄의 가장자리에 위치한 문자들의 기울어진 정도를 측정하여 결정한다. 그 방법은 먼저 각 문자줄의 좌,우 가장자리에 위치한 3개의 문자들을 대상으로 (그림 7)과 같이 문자의 바닥선 중간점 지나는 y축을 기준으로 $-20^\circ \leq \theta \leq 20^\circ$ 범위 내에서 1° 씩 회전한다.

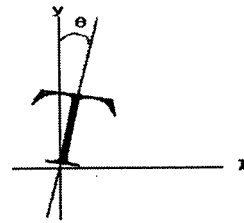


그림 7. 문자의 회전각 측정
Fig.7 Estimation of rotation angle of character

$|\theta| > 20^\circ$ 인 경우는 기하학적 왜곡보정 전, 후의 문자인식률이 거의 차이가 나지 않을 정도로 왜곡이 심한 상태이므로 재촬영이 효과적이다.

그리고 1° 씩 회전할 때마다 수식 4을 이용하여 세로방향 흑화소의 최대 길이를 측정한다. 수식 4에서 y_1, y_2, \dots, y_n 은 (그림 8)과 같이 수직선을 $0 \leq k \leq x_e$ 까지 변화시키는 동안 각 수직선에서의 연속된 흑화소의 길이를 나타낸다.

$$y_{max} = \max(y_1, y_2, \dots, y_i, y_{i+1}, \dots, y_n) \dots\dots (4)$$

(그림 8)과 같이 세로방향 흑화소 길이 중 최대 값을대 상 문자 중 그 문자의 높이(h_{char})에 해당하는 세로방향 획이 있는 경우 θ 가 변화함에 따라 특정 θ 에서 수식 5를 만족하는 문자가 존재하면 θ 가 찾고자하는 문자의 기울어진 각도이며 수식 5를 만족하는 모든 θ 들의 평균 값(θ_{av})을 문자의 기울기로 설정한다.

$$\frac{3}{4} h_{char} \leq y_{max} \leq h_{char} \dots\dots\dots (5)$$

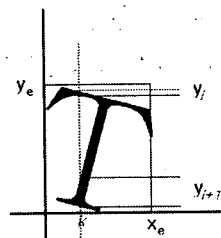


그림 8. 세로 방향 흑화소 길이 측정
Fig.8 Estimation of length for black pixels

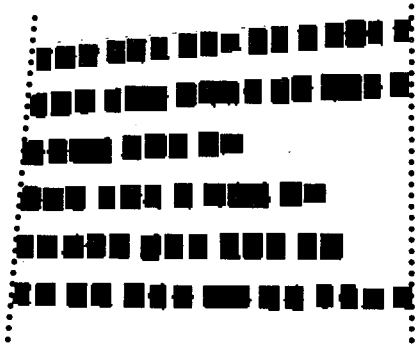


그림 9. 세로 방향 직선 결정
Fig.9 Decision of straight line of vertical direction

2개의 세로방향 직선은 변환영역의 좌, 우 양쪽에서 구한 각각의 기울기를 이용하여 (그림 9)과 같이 직선이 영역의 최외각을 지나도록 굿는다. 인 경우는 측정의 오차를 감안하여 로 수정한다.

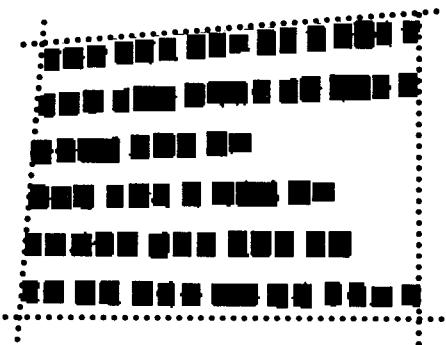


그림 10. 기울어진 영역 추출 결과
Fig.10 Result in extraction of slant region

최종 기울어진 영역 추출 결과는 (그림 10)과 같다.

3.2 양선형 변환 방법

제안한 방법은 사변형 모양으로 왜곡된 변환영역을 직사각형의 영역으로 변환하는 양선형변환을 사용한다. 이방법은 대각선 방향의 값들이 보존되지 않고 비선형적 왜곡을 선형적으로 근사화 할 수 밖에 없는 단점이 있으나 (그림 11)과 같이 사변형을 구성하는 4개의 꼭지점 좌표만 찾으면 변환이 가능하며 3차원이나 비선형 보정방법에 비해 계산량이 적은 장점이 있다.

(그림 11)에서 $q_0(u_0, v_0) \sim q_3(u_3, v_3)$ 는 변환전 사변형의 좌표이고 $p_0(x_0, y_0) \sim p_3(x_3, y_3)$ 는 변환 후 직사각형의 좌표이다. 변환은 수식 (6) ~ (12)을 이용하여 변환후의 좌표 (x, y) 에 해당하는 변환전의 좌표 (u, v) 를 구한다.

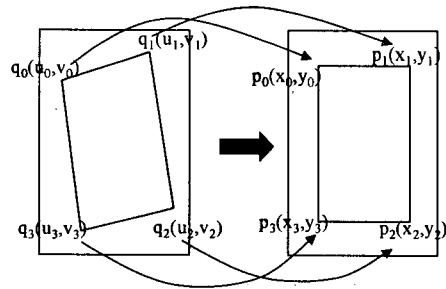


그림 11. 양선형 변환
Fig.11. Bilinear transform

$$u = u_0 + (u_3 - u_0) \times dy \dots\dots\dots (6)$$

$$v = v_0 + (v_3 - v_0) \times dx \dots\dots\dots (7)$$

$$u_1 = u_0 + (u_1 - u_0) \times dx \dots\dots\dots (8)$$

$$u_2 = u_1 + (u_2 - u_1) \times dx \dots\dots\dots (9)$$

$$v_1 = v_0 + (v_1 - v_0) \times dy \dots\dots\dots (10)$$

$$v_2 = v_1 + (v_2 - v_1) \times dy \dots\dots\dots (11)$$

$$dx = \frac{x}{x_1 - x_0}$$

$$dy = \frac{y}{y_3 - y_0} \dots\dots\dots (12)$$

IV. 실험 결과 및 고찰

본 실험은 2.8GHz 성능을 갖는 PC에서 실험을 했으며, Visual C++로 프로그램을 작성하였다. 실험에 사용된 영상은 200만 화소 디지털 카메라 영상 82장, 200만 화소

휴대폰 영상 79장, 100만 화소급PDA 영상 85장의 실험 영상을 상용했다.

카메라 문자영상 기하학적 왜곡보정은 PSNR(power signal to noise ratio)과 문자인식을 2가지를 측정하여 그 성능을 시험한다. PSNR측정에서는 문서영상을 그래픽 도구를 이용하여 인위적으로 기하학적 왜곡을 가한 후 왜곡 보정을 수행하여 복원된 영상과 원영상과의 PSNR을 비교하여 그 결과를 <표 1>에 나타내었다. 또한 왜곡보정 전, 후의 문자 인식률을 실험하여 그 결과를 <표 2>에 나타내었다.

그 결과를 보면 3종류의 시험영상 모두 보정 후 PSNR과 문자 인식률이 향상되어 원영상에 근접함을 확인할 수 있다.

표 1. 기하학적 왜곡보정 전·후의 PSNR 비교
Table 1. Compare PSNR after and before correcting the geometric distortion

항목 PSNR	기하학적 왜곡 보정 전		
	Digital camera 영상	cellular phone 영상	PDA영상
평균(dB)	7.2	6.5	6.1
합계평균	6.6		

항목 PSNR	기하학적 왜곡 보정 후		
	Digital camera 영상	cellular phone 영상	PDA영상
평균(dB)	12.4	11.7	11.2
합계평균	11.8		

표 2. 왜곡보정 전·후의 문자인식률 비교
Table 2. Compare ratio of character recognition after and before correcting the geometric distortion

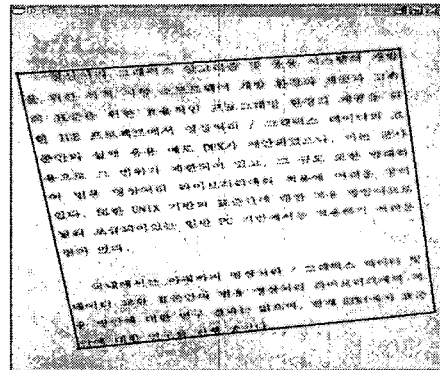
항목 문자인식률	기하학적 왜곡 보정 전		
	Digital camera 영상	cellular phone 영상	PDA영상
평균(dB)	90.3	89.1	89.3
합계평균	89.6		

항목 문자인식률	기하학적 왜곡 보정 후		
	Digital camera 영상	cellular phone 영상	PDA영상
평균(dB)	93.5	95.0	93.8
합계평균	94.1		

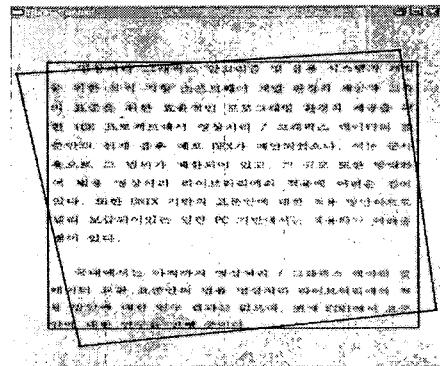
(그림 12)은 200만화소 휴대폰 카메라로 획득하였으며 기하학적으로 왜곡된 예로 제한한 방법을 이용하여 왜곡을

보정한 예이다. (그림 (a))에서는 설정된 사변형의 변환영역을, (b)에서는 (a)에서 설정한 영역과 변환후의 직사각형 영역을 함께 표시하였다.

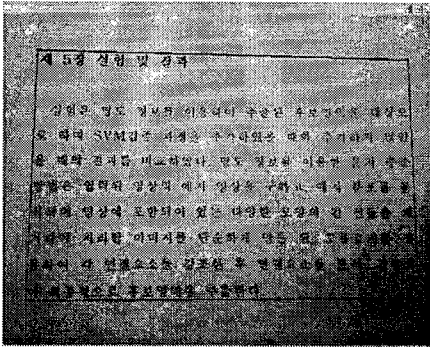
(그림 13)은 200만화소 휴대폰 카메라로 획득하였으며 기하학적으로 왜곡된 예로 제한한 방법을 이용하여 왜곡을 보정한 예이다. (그림 (a))에서는 설정된 사변형의 변환영역을, (b)에서는 (a)에서 설정한 영역과 변환후의 직사각형 영역을 함께 표시하였다.



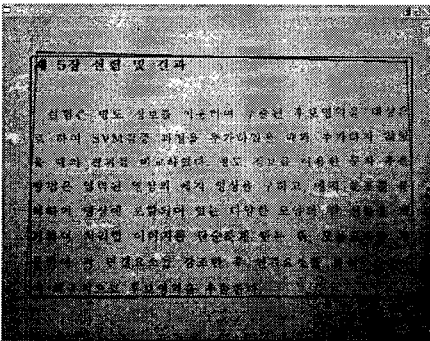
(a) 변환전 문서영상
(a) Before transform in document



(b) 변환 후 결과영상
(b) After transform in document
그림 12. 휴대폰 카메라로 획득한 영상
Fig.12 Image from cellular phone camera



(a) 변환전 문서영상
(a) Before transform in document



(b) 변환 후 결과영상
(b) After transform in document
그림 13. 디지털 카메라로 획득한 영상
Fig.13 Image from digital camera

V. 결론

카메라를 이용하여 문서영상을 획득할 경우 문자들이 기울어지거나 휘어지는 기하학적 왜곡은 3차원적이고 비선형적으로 나타난다. 그러나 이러한 왜곡을 보정하기 위하여 3차원이나 비선형 보정방법을 사용할 경우 왜곡된 문서영상으로부터 계산에 필요한 파라미터들을 정확히 알아내기가 어려우며 계산량이 많아지는 단점이 있다. 따라서 제안한 방법에서는 2차원의 선형보정방법인 양선형 변환을 이용하여 3차원적이고 비선형적인 왜곡현

상을 2차원의 선형으로 근사화하여 보정한다.

제안한 방법은 카메라로 획득한 컬러 문서영상을 이진화한 영상을 이용하여 문자줄을 추출한다. 따라서 이진화가 잘되지 않을 경우 문자줄 추출이 어려워 제안한 방법을 적용할 수 없는 문제점이 있다. 또한 제안한 방법은 글자의 기울어진 정도를 측정할 수 있는 대상글자가 존재하지 않는 경우도 있을 수 있으므로 개선이 필요하다.

참고문헌

- [1] H.Fujisawa, H.Sako, Y.Okada and S.W.Lee, "Information capturing camera and development issues", Proc. of the 6th ICDAR, 2001, pp.205-208
- [2] M.Sawaguchi, K.Yamamoto and K.Kato, "A proposal of character recognition method for low resolution images by using cellular phone", Proc.of the 9th Korean-Japan joint workshop FCV, 2003, pp.216-221
- [3] M.Seeger and C.Dance, "Binarizing camera images for OCR", Proc. of the 6th ICDAR, 2001, pp.54-58
- [4] L.Fan and C.L.Tan, "Binarizing document image using coplanar prefilter", Proc.of the 6th ICDAR, 2001, pp.34-38
- [5] 한상훈, 이강호, "연속 영상 분석에 의한 다중 차량 검출 방법의 연구," 한국 컴퓨터 정보학회 논문지, 2003, 6 v.008, n.002, pp.37-43 1598-849X
- [6] 지정규, 김용균, "가변 탐색창을 이용한 실시간 객체 추적 시스템," 한국 컴퓨터 정보학회 논문지, 2002, 12 v.007, n.004, pp.52-58 1598-849X
- [7] N. Otsu, A Threshold Selection Method From Gray-level Histograms, IEEE Trans. Systems, Man, and Cybernetics, vol. SMC-9, No.1, pp. 62-66, 1979.
- [8] K. K. Kim, S. Y. Chi, and Y. K. Chung, Region-oriented local thresholding of camera document image, IASTED SPPRA, pp. 260-264, June, 2003.

- [9] X. Li, W. Gao, S. Y. Chi, K. A. Moon and H. J. Kim, An Efficient Method for Page Segmentation, Proc. ICICS, Vol.2, pp.957-961, 1997.
- [10] George Wolberg, Digital Image Warping, IEEE Computer Society Monograph, 1990.
- [11] K. K. Kim, J. Y. Lee, and J. H. Kim, Character segmentation of camera document image, IASTED SPPRA, Rhodes, Greece, pp. 256-259, June, 2003.

저 자 소 개



장 대 근

1990년 경북대학교 전자공학과(학사)
 1995년 경북대학교 전자공학과(석사)
 2003년 경북대학교 전자공학과(박사)
 1996년~2005년 한국전자통신연구원
 선임연구원
 2005년 5월~현재 : 특허청 통신
 사무관
 <관심분야> MPEG, 영상처리, 영상
 압축, 영상인식



전 병 태

1986년 한남대학교 전산과(학사)
 1989년 숭실대학교 전산과(석사)
 2001년 고려대학교 컴퓨터학과(박사)
 1992년 5월 IR52 장영실 상 수상
 (과기부 장관상)
 1989년~2004년 한국전자통신연구원
 선임연구원
 2004년 2월~현재 국립 한경대학교
 웹정보공학과 교수
 2003.8.~현재 정통부 산하
 프로그램심의조정위원회
 감정전문위원
 2004.7~현재 한국정보기술학회 이사
 2004.5~현재 한국전자통신연구원
 초빙연구원
 <관심분야> 멀티미디어 영상처리,
 모바일 통신