
카메라 획득 문서영상에서의 글자모양 왜곡보정

장대근* · 김의정**

Character Shape Distortion Correction of Camera Acquired Document Images

Dae-geun Jang* · Eui-jeong Kim**

요 약

스캐너로 획득한 문서영상을 대상으로 문자인식을 하는 경우와 달리 카메라로 획득한 문서영상을 대상으로 문자인식을 수행할 경우 카메라 렌즈의 특성과 카메라와 피사체간의 위치와 각도의 불일치에 의해 글자모양의 왜곡이 발생한다. 따라서 이러한 왜곡들로 인해 실제로 사용이 가능한 카메라 문자인식 기술의 개발이 쉽지 않았다.

본 논문에서는 문서영상을 카메라로 획득할 때 발생하는 3차원의 비선형적 왜곡을 2차원의 선형변환을 이용하여 근사화함으로써 왜곡된 글자모양을 교정한다. 또한 변환과정에서 글자와 글자줄을 추출하여 상하로는 글자줄의 배열방향을, 좌우로는 문자획의 기울어진 각도를 측정함으로써 왜곡으로 인해 사변형 형태로 변형된 글자영역 즉 변환영역을 자동으로 설정함으로써 부가적인 정보의 획득 없이 문서영상 자체만으로 기하학적 왜곡보정이 가능하다.

ABSTRACT

Document images captured by scanners have only skewing distortion. But camera captured document images have not only skew but also vignetting effect and geometric distortion. Vignetting effect, which makes the border areas to be darker than the center of the image, make it difficult to separate characters from the document images. But this effect has been decreased, as the lens manufacturing skill is developed. 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.

키워드

character recognition, distortion correction of document image, image processing

I. 서 론

기존의 문자인식 방법들은 대상 문서를 스캐너를 이용

하여 영상형태로 획득하고 있으나 카메라에 비해 영상 획득 시간이 길고 휴대하거나 이동하기가 불편하며 책의 경우 스캐너를 이용하기 어려운 단점이 있어 사용이 비교적

* 특허청 전지전자심사본부 전기심사팀
** 공주대학교 컴퓨터교육과

간편한 카메라를 이용한 문자인식 방법을 찾으려는 노력들이 있어 왔다. 그러나 카메라를 이용하여 문서영상을 획득할 경우 영상의 가운데 부분에서 가장자리로 갈수록 어두워지는 비네팅현상때문에 배경으로부터 문자의 분리가 어려워지며 카메라와 피사체간 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 위치 정보를 알아야 보정이 가능하지만 이러한 정보를 카메라로부터 얻는 것은 현실적으로 어려움이 있다.

본 논문에서는 임의의 사변형 모양을 직사각형으로 변환하는 양선형변환 방법을 이용하여 디지털 카메라, 휴대폰 카메라, PDA 카메라로부터 획득한 문자영상의 기하학적 왜곡을 보정함으로써 카메라를 이용한 문자인식 성능을 향상시키는 방법을 제안한다.

제안한 방법은 문서영상으로부터 개별 문자와 문자줄을 추출하고 첫번째와 마지막 문자줄을 지나는 가로방향 2개의 직선과 각 문자줄의 가장자리에 위치한 세로방향 문자획들의 기울기를 이용하여 결정된 세로방향 2개의 직선을 이용하여 양선형변환을 위한 사변형의 변환영역을 설정하고 이 영역을 직사각형의 영역으로 변환함으로써 카메라와 피사체간 위치와 각도의 차이에 의한 기하학적 왜곡을 보정한다. 그리고 실험에서는 왜곡보정 전, 후의 영상을 대상으로 PSNR과 문자인식률 비교하여 그 성능을 검정한다.

II. 전처리

전처리과정에서는 칼라문서영상을 이진화하고 연결요소를 기반으로한 상향식 방법을 이용하여 문자를 추출하며 문자들 사이의 연결관계를 이용하여 문자줄을 추출한다.

1. 칼라영상 이진화

카메라로 획득한 칼라문서영상을 256레벨의 그레이 영상으로 변환한 다음 각종 잡음, 왜곡 및 외부 조명에 덜 민감한 영상을 얻기 위하여 영상개선 알고리즘을 적용하고 문서의 배경으로부터 인식대상 문자 영역을 추출하기

위한 이진화 과정을 수행한다.

RGB 칼라영상은 수식 (1)을 이용하여 256레벨의 그레이 영상으로 변환한다.

$$\text{Gray level} = 0.11*B + 0.59*G + 0.30*R \quad (1)$$

카메라로 획득한 칼라문서영상을 256레벨의 그레이 영상으로 변환할 경우 명도레벨이 밝은 쪽이나 어두운 쪽으로 편중되는 현상이 자주 발생한다. 즉 이진화 과정에서 문자를 구성하는 화소가 손실되거나 배경이 문자를 구성하는 화소로 잘못 선택될 수 있다. 따라서 부분 화소값들을 전체 명도 레벨에 고루 분포시키는 명도레벨 정규화가 필요하며, 정규화는 화소(x,y)의 레벨을 f(x,y)라 할 때 정규화 후의 화소 레벨 g(x,y)는 식 (2)를 이용하여 구한다.

$$g(x,y) = (L-1) \frac{f(x,y) - \min[f(x,y)]}{\max[f(x,y)] - \min[f(x,y)]} \quad (2)$$

$$\text{for } 0 \leq x < M, 0 \leq y < N$$

여기서 L은 변환될 영상의 명도레벨 범위이고 M, N은 각각 영상의 가로 및 세로 크기이다.

이진화 방법은 white-class와 black-class간의 between-class variance를 최대화하는 임계값을 찾는 Otsu 이진화 방법[1]을 로컬영역 단위로 적용하여 각 영역마다 임계치를 구하고 이를 기준으로 이진화한다[2].

2. 문자 및 문자줄 추출

문서영상에서 영역분할 방법은 기본이 되는 화소단위에서 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. 제안한 방법은 상향식 방법인 X. Li의 방법[3]을 사용하여 자소들을 분할하며 이 방법은 연속된 화소의 집합인 연결요소를 추출하고 이들을 연결하여 영역을 생성하는 방법으로 4방향 또는 8방향으로 탐색하여 연결요소를 생성하지 않고 줄단위로 연결요소들을 추출한 후 인접줄간 연결요소들을 병합하여 영역을 생성함으로써 계산량을 감소시켜 처리시

간을 단축시킨 상향식 영역분할 방법이다.

카메라 문서영상에서의 글자줄은 기하학적 왜곡에 의하여 기울어지고 휘어지므로 이웃하는 글자들 간의 연결 관계를 검사함으로써 글자줄을 확인할 수 있다. 그 과정은 분할한 가장 왼쪽 상단에 위치한 자소를 시작으로 가장 인접한 오른쪽 자소를 찾고 이 과정을 더 이상의 인접 자소가 없을 때까지 기준위치를 오른쪽 영역으로 옮겨가며 반복함으로써 하나의 문자줄을 추출한다. 나머지 자소들을 대상으로 같은 방법을 적용함으로써 문서영상내의 모든 글자줄들을 찾아낸다. 그림 1은 두 자소간 교차여부를 판단하는 예로 수식 3을 만족하는 경우 두 자소는 교차 관계에 있게 된다.

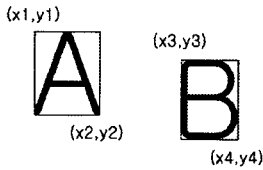


그림 1. 글자간 교차여부 검사 예
Fig. 1 The example of crossing examination among the characters

$$(MinY \geq MaxY) \cap (MinX < MaxX) \quad (3)$$

$$MinY = \min[\max(y1,y2),\max(y3,y4)]$$

$$MaxY = \max[\min(y1,y2),\min(y3,y4)]$$

$$MinX = \min[\max(x1,x2),\max(x3,x4)]$$

$$MaxX = \max[\min(x1,x2),\min(x3,x4)]$$

그림 2는 글자와 글자줄의 추출 예로 각 문자는 외각을 둘러싸는 직사각형 영역으로 표현하였다.

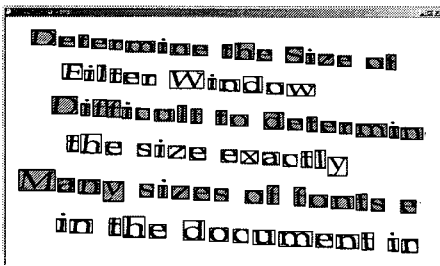


그림 2. 글자와 글자줄 추출 예
Fig. 2 The example of extractions of characters and their lines

III. 기하학적 왜곡보정

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 양선형변환을 이용하여 3차원적이고 비선형적인 왜곡현상을 2차원의 선형으로 근사화하여 보정한다.

1. 양선형변환(bilinear transform)

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

그림 3에서 $q_0(u_0, v_0) \sim q_3(u_3, v_3)$ 는 변환전 사변형의 좌표이고 $p_0(x_0, y_0) \sim p_3(x_3, y_3)$ 는 변환후 직사각형의 좌표이다. 변환은 수식 (4) ~ (10)을 이용하며 변환후의 좌표 (x, y) 에 해당하는 변환전의 좌표 (u, v) 를 구한다.

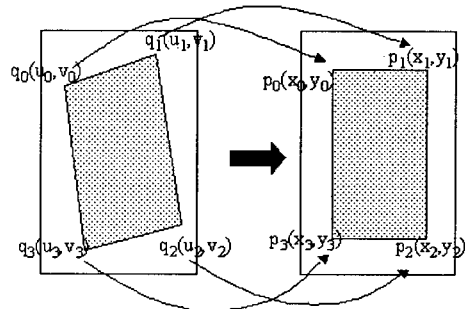


그림 3. 양선형변환
Fig. 3 Bilinear transform

$$u = u01 + (u23 - u01) \times dy \quad (4)$$

$$v = v03 + (u12 - v03) \times dx \quad (5)$$

$$u01 = u0 + (u1 - u0) \times dx \quad (6)$$

$$u23 = u2 + (u3 - u2) \times dx \quad (7)$$

$$v03 = v0 + (v3 - v0) \times dy \quad (8)$$

$$v12 = v1 + (v2 - v1) \times dy \quad (9)$$

$$dx = \frac{x}{x1 - x0}, dy = \frac{y}{y3 - y0} \quad (10)$$

2. 변환영역 자동설정

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.

2.1 가로방향 직선 결정

변환영역을 둘러싸는 4개의 직선 중 2개의 가로방향 직선은 변환영역의 상,하 2개의 최외각 문자줄과 그 구성 문자들을 이용하여 구한다. 그림 4에서 두 문자가 하나의 문자줄을 구성하는 최외각 문자들이고 문자줄이 최상위 문자줄일 경우 직선의 기울기는 $(x1,y1)$ 점과 $(x4,y3)$ 점간의 기울기 θ_u 가 되고 최하위 문자줄일 경우 기울기는 $(x1,y2)$ 점과 $(x4,y4)$ 점간의 기울기 θ_l 이 된다.

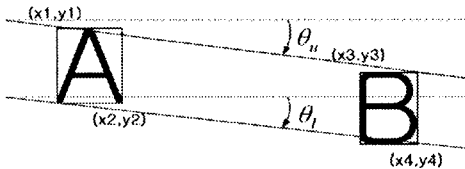


그림 4. 수평방향 직선의 기울기 결정
Fig. 4 The slope determination of horizontal line

직선은 기울기 θ_u 와 θ_l 를 갖는 각각의 직선 l_u 와 l_l 이 그림 5와 같이 변환영역 최외각을 지나도록 결정한다.

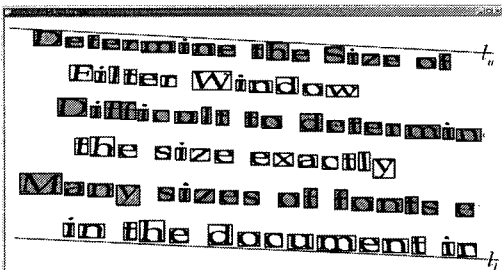


그림 5. 2개의 가로방향 직선 결정
Fig. 5 The determination of two horizontal lines

2.2 세로방향 직선 결정

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 6과 같이 문자의 바닥선 중간점 지나는 y축을 기준으로 $-20^\circ \leq \theta \leq 20^\circ$ 범위내에서 1° 씩 회전한다. $|\theta| > 20^\circ$ 인 경우는 기하학적 왜곡보정 전, 후의 문자인식률이 거의 차이가 나지 않을 정도로 왜곡이 심한 상태이므로 재촬영이 효과적이다.

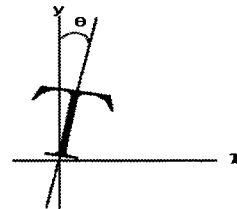


그림 6. 문자의 회전각 측정방법
Fig. 6 The rotation angle measurement of the character

그리고 1° 씩 회전할 때마다 수식 11을 이용하여 세로방향 흑화소의 최대 길이를 측정한다. 수식 11에서 y_1, y_2, \dots, y_n 은 그림 7과 같이 수직선을 $0 \leq k \leq x_e$ 까지 변화시키는 동안 각 수직선에서의 연속된 흑화소의 길이를 나타낸다.

$$y_{max} = \max(y_1, y_2, \dots, y_i, y_{i+1}, \dots, y_n) \quad (11)$$

그림 7과 같이 세로방향 흑화소 길이 중 최대 값을 대상문자 중 그 문자의 높이(h_{char})에 해당하는 세로방향 확이 있는 경우 θ 가 변화함에 따라 특정 θ 에서 수식 12를 만족하는 문자가 존재하면 θ 가 찾고자하는 문자의 기울어진 각도이며 수식 12를 만족하는 모든 θ 들의 평균 값(θ_{av})을 문자의 기울기로 설정한다.

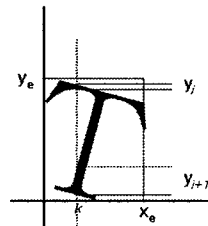


그림 7. 세로방향 흑화소 길이 측정
Fig. 7 The measurement of the vertical length of black pixel

$$\frac{3}{4}h_{char} \leq y_{max} \leq h_{char} \quad (12)$$

2개의 세로방향 직선은 변환영역의 좌, 우 양쪽에서 구한 각각의 θ_{av} 기울기를 이용하여 그림 8과 같이 직선이 영역의 최외각을 지나도록 긋는다. $\theta_{av} \leq 2^\circ$ 인 경우는 측정의 오차를 감안하여 $\theta_{av} = 0^\circ$ 로 수정한다.

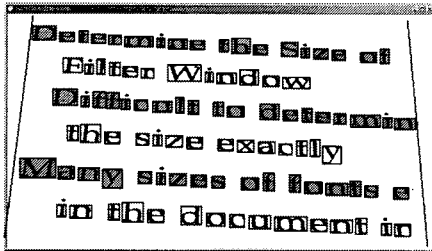
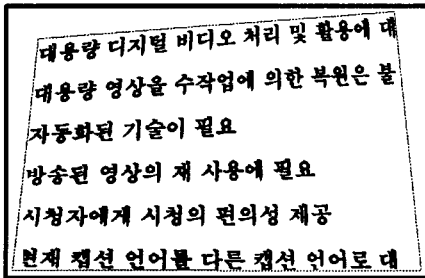


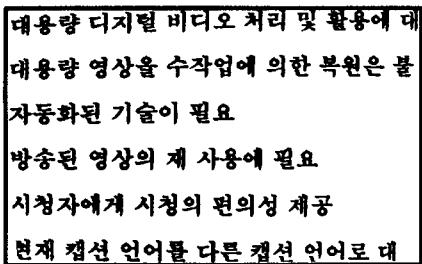
그림 8. 2개의 세로방향 직선 결정
Fig. 8 The determination of two vertical lines

3. 기하학적 왜곡보정

2에서 결정한 4개의 직선이 만든 사변형의 영역을 각 직선의 중점을 지나는 직사각형의 영역으로 3.1에서 설명한 양선형 방법을 이용하여 변환함으로써 기하학적 왜곡을 보정한다. 그림 9는 변환전,후의 예이다.



(a) 변환전 영상



(b)변환후 영상

그림 9. 기하학적 왜곡보정 예

Fig. 9 The example of geometric distortion correction

IV. 실험 및 고찰

카메라 문자영상 기하학적 왜곡보정은 PSNR(power signal to noise ratio)과 문자인식률 2가지를 측정하여 그 성능을 시험한다. PSNR측정에서는 문서영상을 그래픽 도구를 이용하여 인위적으로 기하학적 왜곡을 가한 후 왜곡보정을 수행하여 복원된 영상과 원영상과의 PSNR을 비교하여 그 결과를 표 1에 나타내었다. 또한 왜곡보정 전, 후의 문자인식률을 실험하여 그 결과를 표 2에 나타내었다. 실험은 200만 화소 디지털 카메라로 촬영한 20장의 문서영상과 130만 화소와 200만 화소 휴대폰 카메라로 촬영한 20장의 문서영상과 100만 화소급 PDA 카메라로 촬영한 20장의 문서영상을 대상으로 제안한 방법을 이용한 기하학적 왜곡보정 전후의 평균 PSNR과 문자인식률을 기록하였다.

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

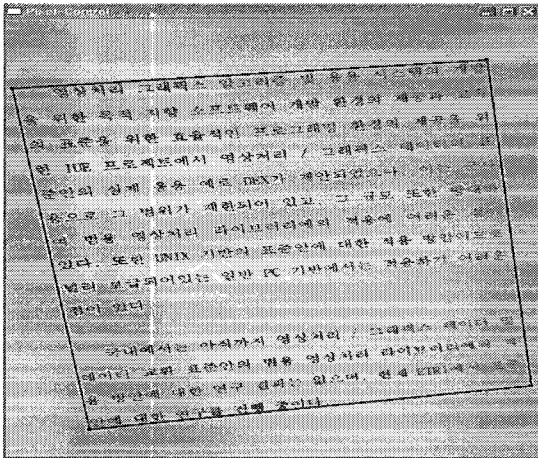
표 1. 기하학적 왜곡보정 전, 후의 PSNR 비교
Table. 1 The comparison of PSNR before and after geometric distortion correction

항목	기하학적 왜곡보정 전			기하학적 왜곡보정 후		
	디지털 카메라 영상 [20장]	휴대폰 영상 [20장]	PDA 영상 [20장]	디지털 카메라 영상 [20장]	휴대폰 영상 [20장]	PDA 영상 [20장]
PSNR						
평균 [dB]	7.2	6.5	6.1	12.4	11.7	11.2
합계평균 [dB]	6.6			11.8		

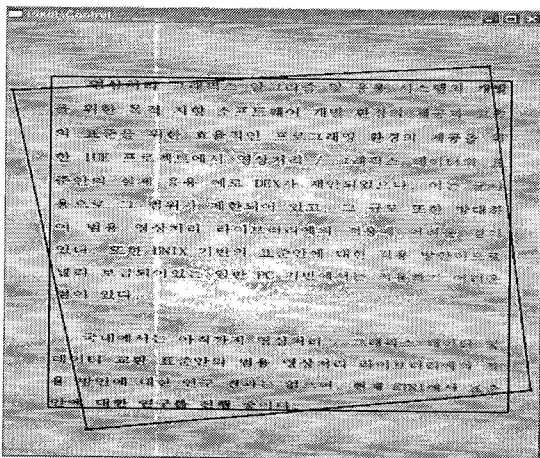
표 2. 왜곡보정 전, 후의 문자인식률 비교
Table. 2 The comparison of character recognition rate before and after geometric distortion correction

항목	기하학적 왜곡보정 전			기하학적 왜곡보정 후		
	디지털 카메라 영상 [20장]	휴대폰 영상 [20장]	PDA 영상 [20장]	디지털 카메라 영상 [20장]	휴대폰 영상 [20장]	PDA 영상 [20장]
문자인식률						
평균 [%]	90.3	89.1	89.3	93.5	95.0	93.8
합계평균 [%]	89.6			94.1		

그림 10, 11의 문서영상은 각각 130만 및 200만 화소 휴대용 카메라로 획득하였으며 기하학적으로 왜곡된 예로 제안한 방법을 이용하여 왜곡을 보정한 예이다. 그림 (a)에서는 설정된 사변형의 변환영역을, (b)에서는 (a)에서 설정한 영역과 변환후의 직사각형 영역을 함께 표시하였다.

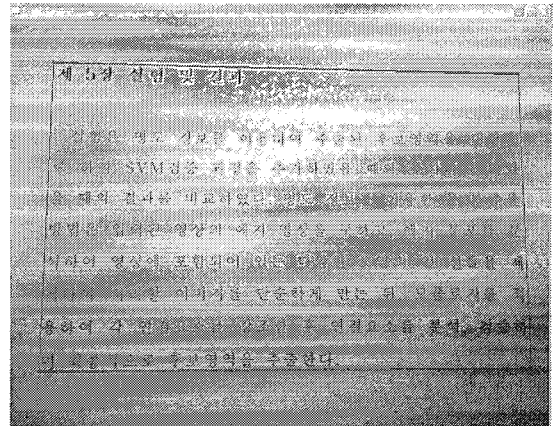


(a) 변환전 문서영상

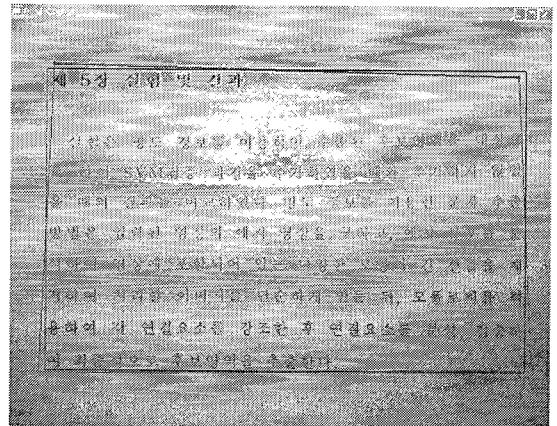


(b) 변환 후 결과영상

그림 10. 130만 화소 폰카로 획득한 영상 실험 예
Fig. 10 The experiment example using 1.3 mega pixel mobile phone camera



(a) 변환전 문서영상



(b) 변환 후 결과영상

그림 11. 200만 화소 폰카로 획득한 영상 실험 예
Fig. 11 The experiment examples using 2 mega pixel mobile phone camera

V. 결론

본 논문에서는 카메라를 이용하여 문서영상을 획득할 때 발생하는 기하학적 왜곡현상을 문자와 문자줄의 기술어진 정도를 측정하여 변환영역을 설정하고 양선형 변환을 이용하여 보정하는 방법을 제안하였다. 성능실험에서는 디지털 카메라, 휴대폰 카메라, PDA 카메라로 획득한 문서영상을 대상으로 보정 전후의 PSNR과 문자인식률을 비교하여 그 성능을 측정하였다. 결과는 왜곡보정 전 평균PSNR 6.6dB가 왜곡보정 후 11.8dB로 향상됨을 확인

할 수 있었고 문자인식률도 89.6%가 94.1%로 4.5% 향상됨을 확인할 수 있었다.

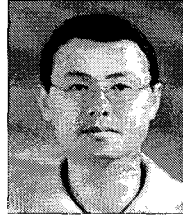
제안한 방법은 카메라로 획득한 컬러 문서영상을 이진화한 영상을 이용하여 문자줄을 추출한다. 따라서 이진화가 잘되지 않을 경우 문자줄 추출이 어려워 제안한 방법을 적용할 수 없는 문제점이 있다. 또한 제안한 방법은 글자의 기울어진 정도를 측정할 수 있는 대상글자가 존재하지 않는 경우도 있을 수 있으므로 개선이 필요하다.

참고문헌

- [1] 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.
- [2] K. K. Kim, S. Y. Chi, and Y. K. Chung, "Region-oriented local thresholding of camera document image," *IATED SPPRA*, pp. 260-264, June, 2003.
- [3] 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.
- [4] George Wolberg, *Digital Image Warping*, *IEEE Computer Society Monograph*, 1990.
- [5] K. K. Kim, J. Y. Lee, and J. H. Kim, "Character segmentation of camera document image," *IATED SPPRA*, Rhodes, Greece, pp. 256-259, June, 2003.
- [6] K. K. Kim, S. Y. Chi, Y. K. Chung and S. K. Park, "Recognition of camera document image," *Proc. of the 9th Korea-Japan Joint Workshop, Frontiers of Computer Vision(FCV)*, pp. 211-215, Feb. 2003.

저자소개

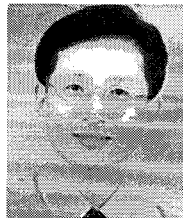
장 대 근(Dae-geun Jang)



2003년 경북대학교 전자공학과 박사
1996년~2005년 ETRI 선임연구원
2005년~현재 특허청 전기전자
심사본부 통신사무관

※ 관심분야 : 문서영상처리, 영상인식, 로봇비전

김 의 정(Eui-jeong Kim)



1997년 충남대학교 컴퓨터공학과
박사
1997년~1998년 SERI 연구원
1998년~현재 공주대학교 컴퓨터
교육과 교수

※ 관심분야 : 컴퓨터비전, 패턴인식, 가상현실, Web-3D