

카메라 기반 문서 인식을 위한 적응적 이진화

(Adaptive Binarization for Camera-based
Document Recognition)

김 인 중*
(In-Jung Kim)

요 약 카메라 영상은 명도의 변화와 부정확한 초점으로 인해 스캐너 영상에 비하여 화질이 저하된다. 본 연구에서는 카메라 영상에서 자주 발생하는 화질 저하에 대한 적응력을 강화하여 카메라기반 문서 인식에 적합한 이진화 방법을 제안한다. 기존의 평가에서 우수하다고 보고된 이진화 방법을 기반으로 하되, 낮은 조도와 부정확한 초점으로 인해 명도 대비가 낮은 영상에 대한 적응력을 강화하였다. 또한 이진화 시 국소 윈도우를 이용하여 기존의 방법에서 공개지기 쉬운 문자의 세부 구조를 섬세하게 추출하도록 개선하였다. 실험에서는 기존에 우수하다고 평가된 이진화 방법들과 제안하는 방법을 문서 인식에 적용하여 다양한 카메라 문서 영상에 대한 성능을 비교하였는데, 그 결과 제안하는 방법이 카메라로 입력 받은 문서 영상의 인식에 효과적임을 확인하였다.

핵심주제어 : 적응적 이진화, 영상 처리, 카메라 기반 문서 인식

Abstract The quality of the camera image is worse than that of the scanner image because of lighting variation and inaccurate focus. This paper proposes a binarization method for camera-based document recognition, which is tolerant to low-quality camera images. Based on an existing method reported to be effective in previous evaluations, we enhanced the adaptability to the image with a low contrast due to low intensity and inaccurate focus. Furthermore, applying an additional small-size window in the binarization process, it is effective to extract the fine detail of character structure, which is often degraded by conventional methods. In experiments, we applied the proposed method as well as other methods to a document recognizer and compared the performance for many camera images. The result showed the proposed method is effective for recognition of document images captured by the camera.

Key Words : adaptive binarization, image processing, camera-based document recognition

1. 서 론

지난 수 십 년 동안 문서 인식 기술은 많은 발전을 이루었으며, 이로 인해 다양한 분야에

성공적으로 적용되어 왔다[1]. 현재 실용화된 많은 문서인식 방법론들은 이진 영상을 대상으로 하고 있다. 이진 영상이 아닌 명도 영상이 입력 될 경우 이를 이진화한 후 인식하는데, 이러한 시스템에서 이진화의 성능은 매우 중요하다[2].

수 년 전까지 실용화 된 문서 인식 시스템의

* 한동대학교 전산전자공학부

주된 인식 대상은 스캐너로 입력 받은 영상이었으나, 최근 디지털 카메라의 보급이 증가하면서 카메라로 촬영한 문서 영상을 인식하기 위한 시도가 증가하고 있다. 카메라 영상은 촬영 거리 및 조명, 초점 상태의 변이와 카메라 센서의 잡영 등으로 인해 스캐너 영상에 비하여 화질 저하가 많이 발생한다[1]. 카메라 영상을 효과적으로 인식하기 위해서는 이 같은 화질 저하에 강한 이진화 방법론이 요구된다.

본 연구에서는 카메라 영상에서 자주 발생하는 조명 변이 및 초점 저하, 센서 잡영으로 인한 화질 저하에 강한 이진화 방법을 제안한다. 제안하는 방법은 과거의 성능 평가에서 성능이 양호하다고 보고된 Sauvola의 방법[3]을 기반으로 하되, 낮은 조도와 초점 저하로 인해 명도의 대비가 뚜렷하지 않은 영상에 대한 적응력을 강화하였다. 또한 이진화 시 국소 윈도우를 이용하여 카메라 영상에서 조명 및 초점 문제로 인해 뭉개지기 쉬운 문자 획의 추출 성능을 크게 개선하였다.

실험에서는 기존에 우수하다고 평가된 이진화 방법들과 제안하는 방법을 문서 인식에 적용하여 다양한 카메라 문서 영상에 대한 성능을 비교하였다.

본 논문의 구성은 다음과 같다. 2장과 3장에서는 이진화와 관련된 기존의 연구와 과거의 성능 평가에서 우수하다고 보고된 방법들을 소개한다. 4장에서는 카메라 영상에서 자주 발생하는 왜곡에 대한 적응력을 강화시킨 이진화 방법을 제안한다. 5장에서는 실험 결과를 소개하고 마지막 6장에서는 결론을 맺는다.

2. 관련 연구

이진화는 명도 영상 G 를 이진 영상 B 로 변환하는 과정이다. 명도 영상에서 화소 x 의 명도를 $G(x)$ 라고 할 때 일반적인 이진화 알고리즘은 먼저 각 화소에 대하여 흑화소와 백화소의 경계값인 임계치 $T(x)$ 를 설정한 후 $G(x)$ 가 $T(x)$ 보다 클 경우는 $B(x)$ 를 0으로, 그렇지 않을 경우 1로 설정함으로써 이진 영상 B 를 생성한다.

이진화 방법은 임계치 $T(x)$ 의 설정 방법에 따라 전역적 이진화와 지역적 이진화로 분류된다. 전역적 이진화 방법은 영상 전체에 대하여 단일 임계치를 적용하는 방법인데, Otsu의 방법이 가장 널리 사용된다[4]. 전역적 이진화는 속도가 빠르고, 명도가 균일한 스캐너 영상에 대하여는 성능이 우수하다. 그러나, 명도 변화에는 잘 적용하지 못하기 때문에 카메라 영상에는 적합하지 못하다.

지역적 이진화 방법은 화소 또는 영역별로 임계치를 다르게 결정하여 이진화하는 방법이다. Otsu의 방법을 지역적으로 확장한 Eikvil 등의 방법[5], 각 지역의 최대 및 최소 명도로부터 $T(x)$ 를 결정하는 Bernsen의 방법[6], 지역 명도 평균 및 표준편차로부터 $T(x)$ 를 결정하는 Niblack의 방법 등이 있다[7].

이와 같이 다양한 이진화 방법들이 발표되자 이들을 문서 인식의 관점에서 평가하기 위한 연구도 함께 수행되었다. Trier는 다양한 이진화 방법에 대하여 문서 인식의 관점에서 비교 평가를 수행하였다[8]. 그 결과 전역적 이진화로는 Otsu의 방법, 지역적 이진화로는 Niblack의 방법이 가장 우수한 것으로 보고되었다. 그 후 Sauvola등은 Niblack의 방법이 배경에 텍스처(texture)가 존재할 경우 많은 잡영을 생성한다는 문제점을 지적하고, 이를 개선한 방법을 제안하였다[3]. 박상은은 저화질의 카드 전표 영상에 대하여 Niblack의 방법과 Sauvola의 방법을 비교했는데, Sauvola의 방법이 잡영에 더 강하다고 보고하였다[9].

He등은 저화질의 고문서 영상에 대하여 다양한 이진화 방법들의 성능을 비교하였는데, 그 결과 Niblack의 방법이 가장 우수한 것으로 발표하였다[10]. 그러나, 역시 고문서를 대상으로 한 Gupta 등의 실험에서는 다중 해상도 Otsu 방법이 Niblack 등의 방법보다 우수한 것으로 보고되었다[11]. 이와 같이 He와 Gupta의 실험 결과가 상이하게 나타난 이유는 실험 대상이었던 문서 영상들의 특성이 서로 달랐기 때문인 것으로 추정된다.

그런데, 카메라 영상에서 발생하는 영상의 왜곡은 고문서에서 발생하는 왜곡과는 그 성질이

다르다. 고문서의 왜곡은 문서 자체의 촬영, 훼손 등에서 비롯된 반면 카메라 영상의 왜곡은 조명에 의한 명도 변이, 초점 저하, 카메라 센서의 촬영 등 영상 획득 과정에서 발생하고 그 특성 또한 고문서를 포함한 스캐너 영상과는 상이하다. 따라서 본 연구에서는 카메라 영상에 특화된 이진화 방법에 대한 연구를 수행하였다.

3. 주요 이진화 방법

기존 연구에서 성능이 좋은 것으로 보고된 방법은 Niblack의 방법[8][10]과 Sauvola의 방법[3][9], 다중 해상도 Otsu방법[11]이다. 각 방법은 다음과 같다.

3.1. Niblack의 방법[7]

Niblack의 방법에서는 각 화소 x 에 대한 $T(x)$ 를 계산하기 위하여 x 를 중심으로 하는 크기 $W \times W$ 인 윈도우를 설정한다. 그리고 윈도우에 포함된 화소들의 명도 평균 $m(x)$ 및 표준편차 $s(x)$ 를 계산한 다음 식(1)에 의하여 임계치 $T(x)$ 를 결정한다.

$$T(x) = m(x) + ks(x) \quad (1)$$

(1)에서 k 는 $s(x)$ 의 영향력을 조절하는 음의 상수이다. Trier의 실험에서는 $W = 15$, $k = -0.2$ 의 값을 사용했을 때 좋은 결과를 나타내었다.

3.2. Sauvola의 방법[3]

Sauvola는 Niblack의 방법이 배경 영역에 텍스처나 명도가 불균일한 경우 많은 잡음을 발생시킨다는 문제점을 지적하고 이를 개선하기 위하여 다음 식을 제안하였다.

$$T(x) = m(x) \left[1 + k \left(\frac{s(x)}{R} - 1 \right) \right] \quad (2)$$

(2)에서 R 은 $s(x)$ 의 범위(dynamic range)로서 256 단계의 명도영상의 경우 128의 값이 많이 사용된다. (2)에서 k 는 역시 $s(x)$ 의 영향력을 조절하는 상수이나, (1)과는 달리 양의 상수가 사용된다. Sauvola의 실험에서는 0.5가 사용되었다.

Sauvola는 또한 지역적 이진화가 계산량이 많다는 문제점을 개선하기 위한 방법도 제안하였다. 모든 화소 x 에 대하여 $T(x)$ 를 직접 계산하는 것이 아니라 가로 세로 각각 n 번째의 기준 화소에 대하여만 (2)에 의해 $T(x)$ 를 계산하고 나머지 화소의 임계치는 주변의 기준 화소 임계치로부터 내삽(interpolation)에 의해 결정하였다.

3.3. 다중 해상도 Otsu 방법[11]

Otsu의 이진화 방법은 화소들의 명도 분포로부터 흑화소와 백화소 각각의 클래스간 가중 분산(weighted between-class variation)을 최대화하도록 임계치를 결정하는 방법이다.

Gupta는 영상을 동일한 크기의 정사각형 블록으로 나눈 후 각 영역에 Otsu의 방법을 적용함으로써 중심점에 대한 $T(x)$ 의 값을 결정하였다. 이 때 결정된 $T(x)$ 를 이용하여 해당 블록을 이진화하였을 때 백화소/흑화소의 비율이 2:1를 넘는다면 그 블록의 크기가 너무 작아서 뭉개진 것으로 판단하여 그 비율이 2:1보다 작아질 때까지 블록의 크기를 동적으로 확장하였다. Gupta의 실험에서는 블록의 초기 크기를 문자 높이의 2배로 지정했으며 뭉개진 블록에 대해서는 2배씩 확장하였다. 각 블록의 중심점 이외의 화소에 대한 $T(x)$ 는 주변의 중심 화소 임계치로부터 내삽에 의해 결정하였다.

4. 카메라 기반 문서인식을 위한 적응적 이진화

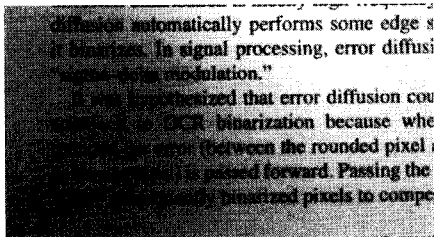
본 연구에서는 카메라로 촬영한 문서 영상의 인식을 위한 이진화 방법을 목표로 하였기 때문에 카메라 영상에서 주로 발생하는 조명의 불균일, 초점 저하로 인한 문자 세부구조의 뭉개짐,

그리고 낮은 조도에서 촬영 시 발생하는 잡영에 중점을 두었다.

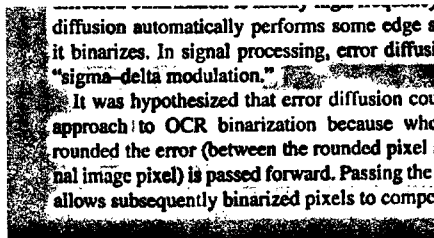
4.1. 기존 이진화 방법에 대한 사전 실험

먼저 이러한 왜곡에 대한 기존 방법들의 적응력을 알아보기 위한 간단한 사전 실험을 수행하였다. 소수의 카메라 영상 샘플에 대하여 적용시키며 그 결과를 관찰하였다.

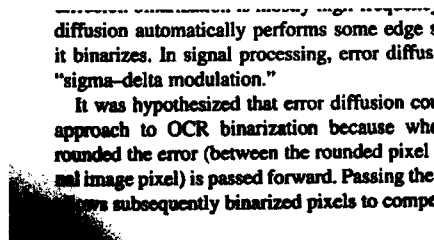
그림 1은 카메라 센서로 인한 잡영이 많고 조명이 불균일한 영상에 대한 이진화 결과들이다. Niblack의 방법은 문자 영역의 이진화에는 효과적이었으나 배경 영역의 잡영에 민감하였다. 특히 어두운 조명에서 촬영되어 센서 노이즈가 포함된 영상에서는 그림 1(b)와 같이 매우 많은 잡영이 발생하였다. 반면 다중 해상도 Otsu 방법은 잡영에 강인한 성능을 보였다. 그러나, 명도 변화가 많은 영상에 대한 적응력이 부족하여 밝은 영역은 문자가 가늘게, 어두운 영역은 두껍게 나오는 문제점이 관찰되었다.



(a) 명도영상



(b) Niblack의 이진화



(c) 다중 해상도 Otsu 이진화

diffusion automatically performs some edge s
it binarizes. In signal processing, error diffusi
"sigma-delta modulation."

It was hypothesized that error diffusion cou
approach to OCR binarization because whe
rounded the error (between the rounded pixel
nal image pixel) is passed forward. Passing the
allows subsequently binarized pixels to compe

(d) Sauvola의 이진화

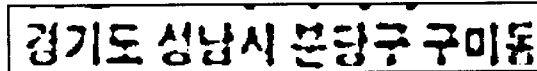
<그림 1> 잡영이 많고 명도가 불균일한 영상

따라서 본 연구에서는 명도 변화와 센서 잡영에 모두 안정적인 성능을 보인 Sauvola의 방법을 기반으로 하였다. 그러나, 화질이 좀 더 낮은 영상에 대하여 Sauvola의 방법을 적용한 결과 다음과 같은 문제점들을 발견하였다.

먼저 Sauvola의 방법은 명도 대비(contrast)가 낮고 초점이 부정확한 영상에 대한 적응력이 충분하지 못했다. 그림 2(a)는 이러한 영상의 한 예이며, 그림 2(b)는 이 영상에 대하여 파라미터를 변화시켜가며 얻은 최적의 결과이다. 문자 획 중 배경과의 명도차이가 작은 부분이나 획의 끝 부분은 소실되는 문제점을 관찰할 수 있었다. 이는 Sauvola 방법의 명도 대비에 대한 적응력이 조명과 초점 상태의 변화가 심한 카메라 영상에는 충분하지 않음을 보여준다.



(a) 명도 영상



(b) Sauvola의 이진화 (W=15, k = 0.1)

<그림 2> 명도 대비가 낮고 초점이 부정확한 영상

두 번째 문제점은 문자 획이 밀집한 부분이 묻개지는 현상이다. 그림 3은 이러한 영상의 예이다. 이는 카메라 영상에서는 획과 획 사이의 배경 화소들이 주변의 흑색화소의 영향으로 상당히 낮은 명도를 갖기 때문에 발생하는 현상으로써 카메라 영상에서 자주 발생하는 문제점이다.



통신

<그림 3> 명도가 낮고 문자 획이 밀집한 문자 영상

이 같은 두 가지 문제점은 저화질의 카메라 영상에서 자주 발생하는 문제점으로써 인식 성능 저하의 주요 원인이 되므로 카메라 기반 문서 인식을 위해서는 이에 대한 개선이 필요하다. 따라서, Sauvola의 방법에서 관찰된 두 가지 문제점을 개선하기 위한 연구를 수행하였다.

4.2. 지역 명도 최대/최소값을 이용한 명도 적응적 이진화

그림 2와 같은 문제점의 원인을 분석하기 위해서 식 (2)를 다음과 같이 전개하였다.

$$T(x) = m(x) \left[1 + k \left(\frac{s(x)}{R} - 1 \right) \right] = m(x) \left[1 - k \left(1 - \frac{s(x)}{R} \right) \right]$$

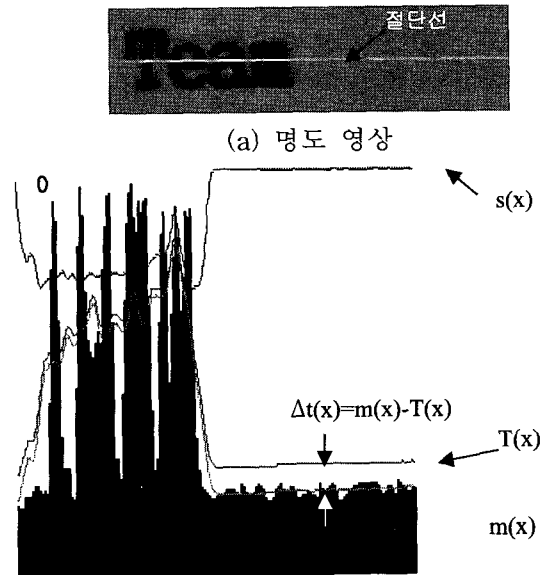
$$= m(x) - m(x)k \left(1 - \frac{s(x)}{R} \right) = m(x) - \Delta t(x) \quad (2')$$

여기에서 $\Delta t(x)$ 는 $m(x)k \left(1 - \frac{s(x)}{R} \right)$ 이다. (2')에서 $T(x)$ 는 주변 화소들의 평균 명도인 $m(x)$ 와 $\Delta t(x)$ 로 나눌 수 있다. 그림 4는 명도 영상을 수평으로 절단한 단면이다. 그림 4(b)에서 볼 수 있듯이 $T(x)$ 는 $m(x)$ 와 근접한 값을 갖는데, 그 결과 주변의 화소들보다 명도가 낮은 화소는 문자 영역으로, 그렇지 않은 화소는 배경으로 판별된다. 그런데, $m(x)$ 를 직접 임계치로 사용할 경우 명도 변화에 지나치게 민감해 지기 때문에 잡영이 많이 발생하는데, 이를 막기 위해 $m(x)$ 에서 $\Delta t(x)$ 을 뺀으로써 $T(x)$ 를 어두운 쪽으로 이동시켰다.

식(2')에서 $\Delta t(x)$ 는 $m(x)$ 에 비례한다. 따라서 그림 4의 좌측과 같은 문자 영역에서는 평균 명도가 낮기 때문에 $m(x)$ 가 작아서 $\Delta t(x)$ 도 역시 작아지는데, 그 결과 $T(x)$ 가 $m(x)$ 와 근접한다. 따라서 명도 변화에 민감해져서 문자 획들이 잘 추출된다. 반면 우측과 같은 배경 영역에서는

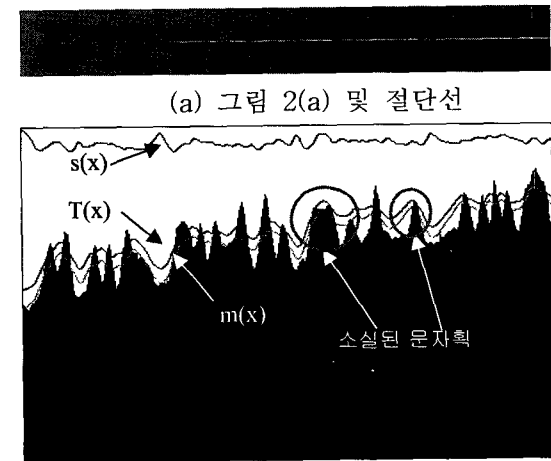
명도가 높아 $m(x)$ 가 크기 때문에 $\Delta t(x)$ 가 커져서 작은 명도 변화는 무시되므로 잡영이 많이 발생하지 않는다.

그런데, 그림 5(a)와 같이 명도의 대비가 약하고 초점이 부정확한 영상은 문자 영역에서도 $m(x)$ 가 제법 큰 값을 갖는다. 그 결과 $\Delta t(x)$ 가 충분히 작아지지 않기 때문에 그림 5에서 동그라미로 표시한 것과 같이 흐린 획은 주변에서의 명도 변화가 무시되어 결국 소실된다.



(b) 명도영상의 단면 (W = 15, k = 0.1)

<그림 4> 단면 영상에서의 Sauvola 임계치



(b) 그림 2의 절단면

<그림 5> 그림 2(a)의 Sauvola 임계치

본 연구에서는 이 같은 문제점을 보완하기 위하여 $\Delta t(x)$ 에 주변 화소 명도의 최대/최소값을 반영하도록 수정하였다. $\Delta t(x)$ 가 단순히 $m(x)$ 에 비례하는 대신 주변 화소의 최소 및 최대 명도에 대하여 정규화된 $m'(x)$ 에 비례하게 함으로써 영상의 명도 대비 및 초점 저하에 적용하도록 하였다. $m'(x)$ 은 다음과 같이 정의된다.

$$m'(x) = 255 \frac{m(x) - L_{\min}}{L_{\max} - L_{\min}} \quad (3)$$

여기에서 L_{\min} 과 L_{\max} 는 x 를 중심으로 하는 크기 $W_{\text{large}} \times W_{\text{large}}$ 의 윈도우 내 화소들의 최대/최소 명도이다. (3)을 (2)에 반영할 경우 $T(x)$ 는 (4)와 같이 정의된다.

$$\begin{aligned} T(x) &= m(x) - m'(x)k \left(1 - \frac{s(x)}{R}\right) \\ &= m(x) - 255 \frac{m(x) - L_{\min}}{L_{\max} - L_{\min}} k \left(1 - \frac{s(x)}{R}\right) \end{aligned} \quad (4)$$

(4)의 k 는 (2)에서와 유사한 의미를 가지나 (2)보다 작은 값을 갖는다. 실험적으로 (2)의 k 에 비하여 약 3/4정도의 값을 가질 때 좋은 결과를 나타내었다. 그림 2(a)를 제안하는 방법 (4)를 이용하여 이진화한 결과는 그림 6과 같다 ($W=15$, $W_{\text{large}}=30$, $k=0.075$).

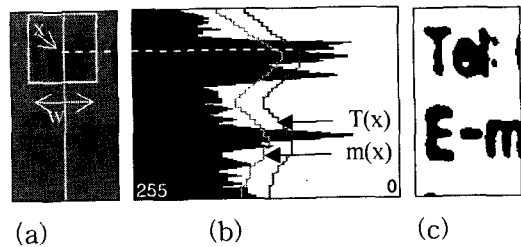
경기도 성남시 분당구 구미동

<그림 6> 그림 2(a)에 대한 제안하는 방법 적용 결과

$\Delta t(x)$ 을 $m(x)$ 대신 L_{\min} 과 L_{\max} 를 이용해 정규화된 $m'(x)$ 에 비례하도록 수정 함으로써 명도 대비가 낮은 영상에 대한 적응력이 강화되었음을 알 수 있다.

4.3. 국소 윈도우를 이용한 문자 획 추출 최적화

그림 3과 같이 문자 획이 밀집한 영역이 뭉개지는 것은 문자 획과 근접한 백화소들이 주변 화소의 영향으로 낮은 명도 값을 가져서 이진화 시 흑화소로 판별되기 때문이다. 예를 들어 그림 7(a)의 x 와 같은 배경 화소는 주변의 흑화소보다는 명도가 다소 높으나 일반적인 배경 영역의 백화소보다는 명도가 매우 낮다. 따라서 해당 화소를 중심으로 하는 $W \times W$ 윈도우 내에 명도가 높은 배경 화소가 많이 포함될 경우 흑화소로 판별되는데, 그 결과 문자 획 사이의 배경 영역이 뭉개지게 된다.



<그림 7> 이진화 시 뭉개지는 영상의 단면

이와 같은 문제점을 해결하기 위해서는 문자 획이 밀집한 영역을 이진화할 때 명도가 높은 배경 화소가 많이 포함되지 않도록 해야 한다. 따라서, $T(x)$ 를 계산할 때 크기가 작은 국소 윈도우를 사용함으로써 주변 영역의 배경 화소를 제외되도록 하였다. 그 결과 그림 8과 같이 획들 사이의 배경 영역을 이진화 할 때 문자 외부의 배경 화소의 영향을 적게 받아서 백화소로 판별되었다. 실험에서는 국소 윈도우의 크기를 문자 획의 굵기로 지정했을 때 좋은 이진화 결과를 얻을 수 있었다.



<그림 8> W 를 획의 굵기로 지정했을 때 이진화 결과

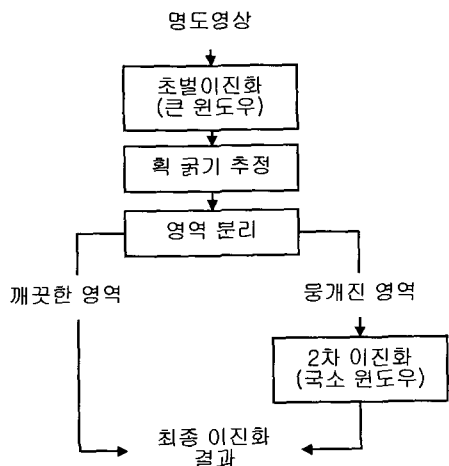
그러나, 이와 같이 작은 윈도우를 이용해 이진화 할 경우 $T(x)$ 가 명도 변화에 지나치게 민

감하게 되기 때문에 그림 9과 같이 많은 잡영이 발생하였다.



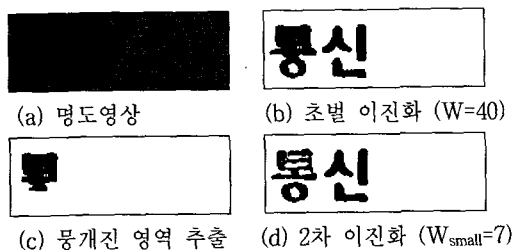
<그림 9> 그림 3에 대하여 W=7로 이진화한 결과

이 같은 문제점을 보완하기 위하여 본 연구에서는 크기가 다른 두 개의 윈도우를 이용하여 이진화하는 방법을 개발하였다. 그림 10은 제안하는 방법의 전체 흐름을 보여준다.



<그림 10> 제안하는 방법의 흐름도

먼저 충분한 크기의 윈도우로 이진화를 수행한 후 문자 획이 뭉개진 부분과 그렇지 않은 부분을 구분한다. 그리고, 뭉개진 영역에 대해서만 선택적으로 국소 윈도우를 이용한 2차 이진화를 수행한다.

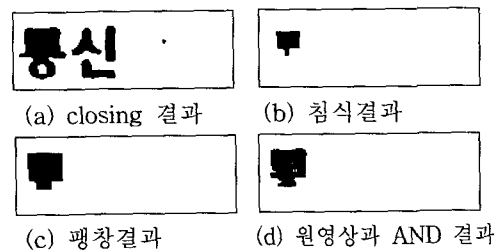


<그림 11> 국소 윈도우를 이용한 이진화

국소 윈도우의 크기로 사용하기 위한 문자 획의 굵기는 초벌 이진화로 얻은 이진 영상으로부터 추정할 수 있다. 먼저 이진 영상으로부터 수평, 또는 수직 런(run)들을 추출하여 런의 길이에 대한 히스토그램을 구성한 후 빈도수가 가장 높은 런의 길이를 획의 굵기로 선택한다.

이와 같이 획의 굵기를 추정하였다면 그 값을 이용하여 뭉개진 영역을 추출할 수 있다. 일반적으로 문서 영상에서는 획의 굵기가 균일하나, 뭉개진 영역은 그 보다 두꺼운 형태를 갖는다. 본 연구에서는 이점을 이용하여 획의 굵기보다 두꺼운 영역을 뭉개진 영역으로 판정하였다. 이진 영상에서 두꺼운 영역을 추출하는 방법은 다음과 같다.

먼저 이진 영상에 대하여 closing을 수행한다. 그 후 문자 획 굵기의 1/2만큼 침식(erosion)을 수행한다. 이렇게 하면 그림 12(b)와 같이 뭉개지지 않은 부분은 사라지고 굵은 영역의 중심부만 남게 된다. 이 영상을 동일한 횡수만큼 팽창(dilation)하면 그림 12(c)와 같은 영상을 얻을 수 있다. 이 영상을 원래의 영상과 AND하면 그림 12(d)와 같이 뭉개진 영역을 찾을 수 있다.



<그림 12> 뭉개진 영역의 추출

5. 실험

본 논문에서 설명한 이진화 방법들을 실제 문서 인식에 적용하였을 때의 효과를 알아보기 위해 세 종류의 카메라로 촬영한 다수의 문서 영상에 대하여 실험하였다. 데이터 수집을 위해서는 Kenox α5, Nikon Coolfix S5 등 두 대의 디지털 카메라와 큐리텔 PT-L1400 카메라폰을 사용하였다. 교과서, 논문, 신문 등 다양한 문서를 각 카메라로 30장씩 촬영하여 모두 90장의

문서 영상을 사용하였다. 각 영상 당 문자의 수는 평균 711개였다. 인식기로는 상용 인식기인 Omnipage SE 2.0을 이용하였다[12]. 비교 대상은 다중 해상도 Otsu방법(MROtsu), Sauvola의 방법, 제안하는 방법, 마지막으로 Omnipage의 내장 이진화 등 4종이다. Niblack의 방법은 센서 잡영으로 인해 지나치게 낮은 인식 성능을 나타내었으므로 비교에서 제외하였다. 각 이진화 방법의 파라미터는 다양하게 변화시켜가며 실험적으로 가장 높은 성능을 내는 값을 선택하였다.

<표 1> 카메라 및 이진화 별 인식 성능

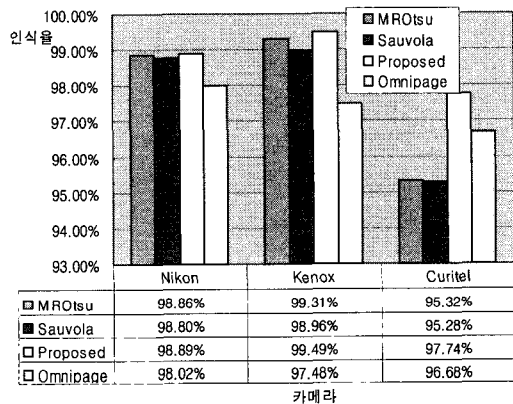
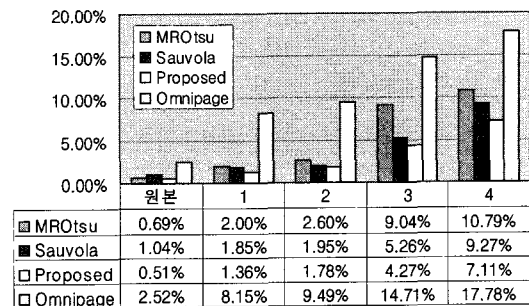


표 1은 각 카메라 및 이진화 방법 별 인식 성능이다. 모든 경우에 있어서 제안하는 방법이 우수한 성능을 나타내었다. 화질이 좋은 디지털 카메라 영상에서는 그 차이가 비교적 작았으나 상대적으로 화질이 낮은 카메라폰 영상에서는 그 차이가 더욱 컸다.

또한 초점 저하에 대한 적응력을 평가하기 위해 촬영 영상을 평활화(smoothing)한 후 이진화하여 인식하였을 때 오류를 측정하였다. 본 실험에서는 7x7 Gaussian 윈도우를 이용하였고, 표준편차 σ 를 다양하게 변화시키며 인식 오류율의 변화를 측정하였다. 두 번째 실험에서는 세 가지 카메라의 영상 중 평균 인식률이 가장 높은 Kenox 카메라의 영상을 이용하였다.

σ 를 증가시킴에 따라 모든 이진화 방법론에서 인식 오류율이 증가하였다. 그러나, 제안하는 방법이 다른 방법에 비하여 오류의 증가량이

<표 2> σ 에 따른 인식 오류율의 변화



상대적으로 적었는데, 이로 인하여 제안하는 방법이 카메라 영상에서 많이 나타나는 초점 저하에 적응력을 개선하는데 효과가 있음을 보여준다.

6. 결론

카메라로 촬영한 문서 영상을 효과적으로 인식하기 위해서는 카메라 영상에서 자주 발생하는 명도의 변이나 부정확한 초점, 카메라 센서의 잡영 등에 강인한 이진화 알고리즘이 요구된다. 본 논문에서는 이와 같은 카메라 문서 영상의 특징에 효과적으로 적용할 수 있는 이진화 방법을 제안하였다.

제안하는 방법은 기존 평가에서 우수하다고 보고된 Sauvola의 방법을 기반으로 하되, 지역 화소들의 최대 및 최소 명도를 반영함으로써 명도의 대비가 낮고 초점이 부정확한 카메라 영상에 대한 적응력을 강화하였다. 또한 초벌 이진화 결과로부터 뭉개진 영역을 자동으로 추출한 후 국소 윈도우를 이용하여 2차 이진화를 수행함으로써 문자 획이 밀집한 영역에서도 문자의 구조를 효과적으로 추출할 수 있게 하였다.

다수의 카메라로 촬영한 문서 영상에 대하여 제안하는 방법과 기존 연구에서 우수하다고 알려진 이진화 방법들을 적용한 후 인식기에 입력하여 인식률을 비교하였다. 그 결과 제안하는 방법이 카메라 기반 문서 인식 시스템에 적합한 성능을 제공함을 확인할 수 있었다.

참 고 문 헌

- [1] J. Kim, "New chances and new challenges in camera-based document analysis and recognition," 1st Int. Workshop on CBDAR, Keynote speech, 2005.
- [2] ㈜인지소프트, Inzi i-Form™ 제품 소개서.
- [3] J. Sauvola and M. Pietikainen, "Adaptive document image binarization," Pattern Recognition, vol. 33, pp. 225-236, 2000.
- [4] N. Otsu, "A threshold selection method from grey level histogram," IEEE Trans. SMC, vol. 9, no. 1, pp. 62-66, 1979.
- [5] L. Eikvil, T. Taxt and K. Moen, "A fast adaptive method for binarization of document images," Proc. 1st ICDAR, pp. 435-443, Saint-Malo, France, 1991.
- [6] J. Bernsen, "Dynamic thresholding of grey-level images," Proc. 8th ICPR, pp. 251-255, Paris, France, 1986.
- [7] W. Niblack, "An Introduction to Image Processing," Prentice-Hall, Englewood Cliffs, NJ, pp.115-116, 1986.
- [8] O. D. Trier, "Goal-directed evaluation of binarization methods," IEEE PAMI vol. 17, no. 12, pp 1191-1201, 1995.
- [9] 박상은, "인식 영역의 추출과 이미지 개선을 통한 신용카드 전표 이미지 인식", KAIST 전산학과 석사논문, 2001.
- [10] J. He, Q.D.M. Do, A.C. Downton and J.H. Kim, "A Comparison of Binarization Methods for Historical Archive Documents," Proc. of 8th ICDAR, Seoul, Korea, pp. 538-542, 2005.
- [11] M.R. Gupta, N.P. Jacobson and E.K. Garcia, "OCR binarization and image pre-processing for searching historical documents," Pattern Recognition, vol. 40, no. 2, pp. 389-397, 2007.
- [12] Omnipage SE
(<http://www.nuance.com/omnipage>)



김 인 중 (In-Jung Kim)

- 1994년 KAIST 전산학과 (학사)
- 1995년 KAIST 전산학과 (석사)
- 2001년 KAIST 전산학과 (박사)
- 2001.2~2006.2 (주)인지소프트 책임연구원
- 2006.3~현재 한동대학교 교수
- 관심분야 : 패턴인식, 영상처리, 인공지능
- 주요 수상 경력: 2001년 KAIST 전산학과 우수논문상(박사논문), 2005년 49주차 IR52 장영실상