

# 명도 정보를 이용한 자연 영상에서의 기울기 보정 및 텍스트 추출

최규담<sup>1</sup>, 김성동<sup>2</sup>, 최기호<sup>1</sup>

<sup>1</sup>광운대학교 컴퓨터공학과

<sup>2</sup>계원조형예술대학 정보통신과

## Text Extraction and Skew Compensation in Natural Scenes using Gray-level Information

Gyu-Dam Choi<sup>1</sup>, Sung-Dong Kim<sup>2</sup>, Ki-Ho Choi<sup>1</sup>

<sup>1</sup>Dept. of Computer Engineering, Kwangwoon University

<sup>2</sup>Dept. of Infomation Communication, Kaywon School of Art and Design

### 요 약

본 논문은 실내외에서 얻어진 자연 영상으로부터 기울어진 영상을 바로 보정하고 텍스트를 추출하는 방법을 제안한다. 본 연구는 명도 이미지를 대상으로 모든 과정이 4단계로 이루어진다. 첫째 자연 영상에서 에지 검출 처리를 위한 전처리 및 Canny 에지 추출을 수행하며, 둘째 영상의 기울기를 추출하기 위해 허프변환에 대한 전처리와 후처리를 한 후, 셋째로 잡음영상과 선을 제거하고 텍스트 특징을 이용한 후보영역 검출을 한다. 마지막으로 텍스트 후보영역 안에서 지역적 이진화를 수행하여 불필요한 비텍스트 연결 요소를 추려내기 위해 두 가지 텍스트 추출 방법을 수행한다. 본 연구는 게시판, 교통표지판, 책 표지 등 100장의 자연영상을 대상으로 실험한 결과 텍스트 추출에서 90.3% 추출 정확도를 가졌으며, 기울어진 각도 추출에서도 94.3%의 높은 추출률을 보였다.

### 1. 서 론

최근 컴퓨터의 발전과 인간의 시각기관과 유사하게 컴퓨터와 로봇에 부착된 카메라를 통해 인위적으로 삽입하거나 자연적으로 포함되어진 정보들을 적절히 분석하고 이해하고자 하는 많은 연구가 진행되었다. 이러한 정보들은 여러 비전 응용 시스템에 중요한 역할을 하고 있으며 이렇게 수많은 정보를 담고 있는 지능 시스템으로써 꼭 획득해야 하는 것이 바로 텍스트 정보이다. 텍스트 정보는 정보번역이나 편집, 해석이 용이하기 때문에 실시간에 추출 및 인식 할 수 있다면 시각 장애인 보행 보조 시스템, 자율 이동 로봇 운행, 무인 자동차 운행 등 다양한 분야에서 널리 활용될 수 있다. 하지만 이런 텍스트를 포함한 영상은 그림 형태로 구성되어 있기 때문에 당장 직접 활용에는 많은 제약점들이 따르게 된다[1][2].

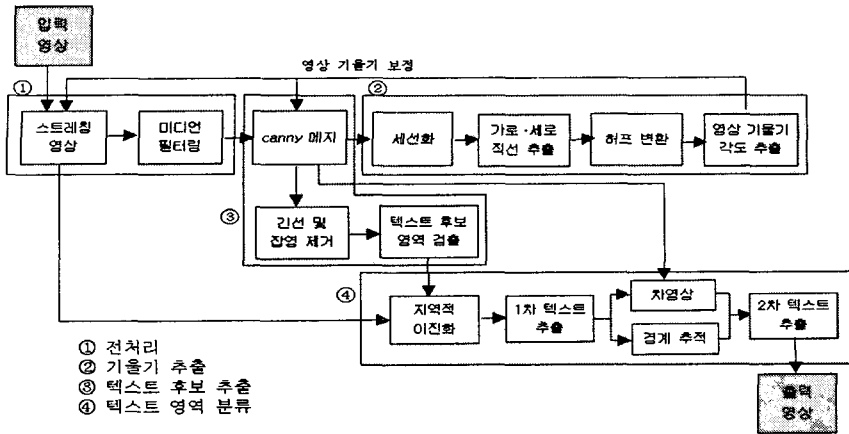
본 논문은 명도의 특징을 이용하여 자연영상의 텍스트를 추출하는 방법을 제안하고, 기울어진 영상에 대해서는 빠르게 되도록 보정한다.

텍스트 추출에 대한 기존 연구로써는 크게 상향식

방법(bottom-up)과 하향식(top-down)방법으로 나눌 수 있는데 상향식 방법은 대표적으로 연결요소(connected components)방법[3]과 인접 선분 밀도(neighborhood line density)분석방법[4]이 있다. 이 방법은 여러 가지 다양한 형태의 텍스트를 처리 할 수 있으나 많은 계산량을 필요로 한다. 또한, 하향식 방법으로는 런 길이 평활화 방법[5], Fourier 변환방법[6] 등이 있으며 장점으로 매우 빠르게 처리될 수 있는 반면에 임계값을 미리 결정해야 하며 텍스트가 기울어져 있는 경우 효과적으로 적용이 어렵다는 단점이 있다.

기울기 추출 관련 기존 연구로는 허프변환(Hough Transform)[7], 투영 윤곽(projection profile)분석에 의한 방법[8] 등을 사용하고 있으나 복잡한 배경을 가진 자연 영상에서 기울기를 찾는 것은 다소 어려운 점이 있다.

본 논문에서는 명도의 특성을 이용하여 기존연구와 차별화 된 방법을 제안함으로써 자연영상에서 텍스트 추출 정확성을 보이고, 또한 기울기 추출 방법에서도



[그림 1] 자연영상에서의 텍스트 추출 과정

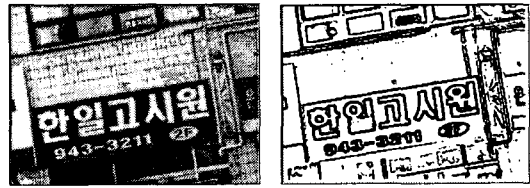
허프변환에 대한 전처리와 후처리 알고리즘을 제안하여 기울어진 각도를 추출함에 있어 정확성을 높이는 데 중점을 두었다.

## 2. 자연 영상에서 텍스트 추출의 설계

본 논문에서 제안하는 텍스트 추출과정은 네 단계를 거친다. 첫 번째는 전처리와 에지 검출을 수행하고 두 번째는 영상의 기울기 추출 및 보정단계를 세 번째는 텍스트 후보영역화(segmentation) 단계를 거치고 마지막으로 텍스트 영역 분류(Block classification)단계로 비텍스트와 텍스트를 분류해낸다. [그림 1]은 이 과정을 요약하여 나타낸 그림이다.

### 2.1 전처리 및 에지 추출

자연영상에서의 텍스트는 중요한 정보를 알려줘야 하기 때문에 사람의 눈으로 잘 구별할 수 있도록 그 주변 배경과 보색관계를 가지고 있다. 따라서 본 연구는 텍스트 영역과 주변 배경과의 명도차를 이용한 에지를 사용한다. 우선 에지 추출을 위한 전처리로 명도의 밝기 차를 더 부각시키기 위해 어떤 특징부분에 집중되어 있는 화소값을 명암대비가 곱고루 분포 되도록 화소 밝기를 스트레칭(Stretching)시키고 잡영의 일부를 제거하기 위해 3x3 마스크 미디언(Median)필터를 적용한다. 다음 단계로 에지 수행을 하는데 대부분의 에지 추출 검지기는 잡음에 대해 민감하므로, 작은 잡음이라도 그것을 에지로 간주하여 추출하는 경우가 많다. 텍스트를 추출하기 위해서는 정확한 경계선의 검출을 위해 끊어짐 현상이 방지되어야 하는데 많은 텍스트 결과 prewitt 마스크, 가우시안의 표준편차  $\sigma=0.6$ , 이중 경계값, high값=70, low값=40을 이용한 Canny 에지 검지기를 사용하였을 때, [그림 2]의 (b)와 같이 텍스트를 검출을 하기 위한 가장 좋은 영상을 얻는다.

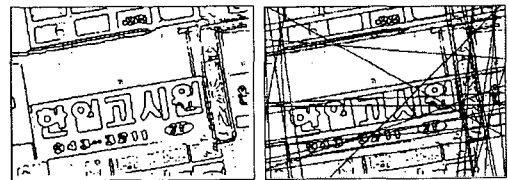


(a) 스트레칭 영상 (b) Canny(Prewitt) 에지 영상

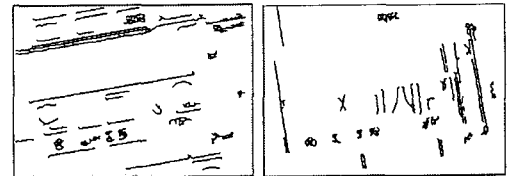
[그림 2] 전처리 및 에지 영상

### 2.2 텍스트 기울기 추출 및 보정

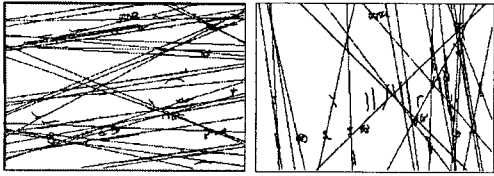
에지영상에서 기울기를 추출하기 위한 전처리로 세선화(Thinning) 처리를 수행하여 선의 두께를 한개의 픽셀로 만든다. 세선화에 의해 생긴 여러 선분에서 각도를 알아내기 위해 허프변환(Hough Transform)을 수행한다. 그러나 허프변환만을 사용하면 [그림3]의 (b)와 같이 너무나 많은 기울기를 가지고 있는 직선이 발생하여 정확한 기울기를 추출하는데 어렵다.



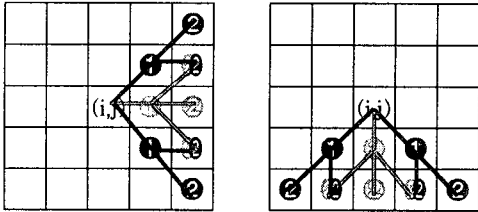
(a) 세선화 (b) 허프변환



(c) 가로 직선 추출 (d) 세로 직선 추출



(e) 가로 각도 추출(허프변환) (f) 세로 각도 추출(허프변환)  
[그림 3] 기울기 추출 과정



(a) 가로 추출 윈도우 (b) 세로 추출 윈도우  
[그림 4] 가로·세로 직선 추출 윈도우

따라서 허프변환을 하기 전 [그림 4]의 각 가로·세로  $-45^\circ \sim +45^\circ$ 의 직선을 추출하는  $5 \times 5$  윈도우로  $(i,j)$ 에 있는 흑화소에 대해 색상이 같은 ①과 ②가 연속되어 있는 화소만 남기고 나머지 화소는 제거한다. [그림 3]의 (e),(f)와 같이 각 추출된 가로·세로 직선들은 허프 변환 알고리즘과 다음과 같은 단계로 각도를 추출한다

- (1) 세로 각도의 빈도수와 가로 각도의 빈도수를 가장 많은 것부터 적은 것으로 차례로 서로 비교해 가면서 차가 90도를 가지면 그때 가로 각도는 영상의 기울기가 된다. (단 빈도수가 1인 것은 연산에서 제외시킴)  
위의 예)  $99^\circ - (9^\circ) = 90^\circ$  영상 기울기 :  $9^\circ$
- (2) (1)에서 기울기 각도를 찾지 못하면 가로·세로 기울기의 전체 빈도수 히스토그램을 구해 첨두값을 영상의 기울기 각도로 지정한다.

추출된 각도는 스트래칭 영상, Canny(Prewitt) 영상에 식(1)을 이용하여 기울어진 영상을 바로 보정한다.

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x + C_x \\ y + C_y \end{pmatrix} + \begin{pmatrix} C_x \\ C_y \end{pmatrix} \dots \dots \dots \text{식(1)}$$

$(x, y)$  : 회전 전의 원 영상 좌표값  
 $(x', y')$  : 회전 후의 새로운 영상 좌표값  
 $(C_x, C_y)$  : 원 영상 회전 중심의 좌표값  
 $\theta$  : 회전각

### 2.3 후보 영역화

기울기가 보정된 에지영상의 텍스트 주변에는 텍스트 후보영역을 추출하는데 불필요한 잡영과 필요없는 긴 선들이 많이 발생하게 된다. 작은 잡영의 경우 비록 이것이 텍스트라고 하더라도 인식이 어렵기 때문에 제거해 줄 필요가 있다. 이런 잡영 및 복잡한 선들을 제

거하기 위해 라벨링(Labeling)의 알고리즘을 사용한다. 라벨링을 수행할 때 연결요소의 수가 매우 작거나 연결요소를 둘러싸고 있는 블록의 넓이보다 5분에 1보다 작으면 제거한다. 텍스트 후보 영역(블록)은 에지 영상을 조사하는 과정에서 연결요소가 처음 발견된 위치  $X_{start} = X_c, Y_{start} = Y_c, X_{end} = X_c + \Delta X, Y_{end} = Y_c + \Delta Y$  ( $\Delta X$ : 블록의 길이,  $\Delta Y$ : 블록의 높이)로 표현된다. 계속해서 연결요소를 좌  $\rightarrow$  우, 위  $\rightarrow$  아래로 조사하면서 같은 연결성의 화소의 블록을 찾아 낼 수 있다. 각 연결요소의 블록을 찾는 식은 다음과 같다.

$$\begin{aligned} X_{start} &= \min(X_{min}, X_c) \\ Y_{start} &= \min(Y_{min}, Y_c) \\ \Delta X &= \max(Y_{max}, X_c) \dots \dots \dots \text{식(2)} \\ \Delta Y &= \max(Y_{max}, Y_c) \end{aligned}$$

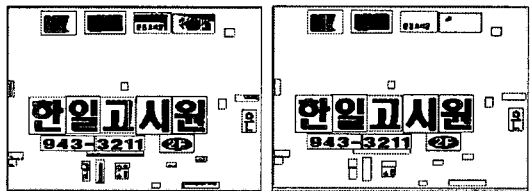
비교 대상 블록 :  $(X_{min}, Y_{min}, Y_{max}, Y_{max})$

### 2.4 텍스트 분류

텍스트 영역 분류는 우선 스트래칭 된 영상에 각 후보영역의 위치 정보를 가지고 있는 블록을 매칭시켜 각 영역별로 지역적 이진화 작업을 수행한다. 이렇게 이진화 된 영역은 1차 텍스트 분류와 2차 텍스트 분류 과정을 통해 텍스트만을 추출 하게 된다.

#### (1) 1차 텍스트 분류

이진화 된 후보 텍스트 영역은 <그림5>의 (a)와 같이 텍스트는 블록 선 안에 존재하지만 비텍스트인 경우는 선에 존재하는 경우가 있다. 따라서 각 블록 안에 라벨링을 위해 연결요소들을 조사하여 선에 붙어있는 연결요소는 <그림5>의 (b)와 같이 제거한다.



(a) 지역 이진화 후 영상 (b) 1차 텍스트 추출 후 영상  
[그림 5] 1차 텍스트 추출

#### (2) 2차 텍스트 분류

2차 텍스트 추출은 Canny(Prewitt) 대신 Canny(Sobel) 연산 알고리즘을 사용한 에지와 1차 텍스트 추출 영상을 식(3)과 같이 차를 구하고, 1차 텍스트 추출의 연결요소를 경계 추적을 한 영상을 식(4)를 이용하여 나머지 비텍스트를 삭제한다.

$$DT(i, j) = CS(i, j) - T1(i, j) \dots\dots\dots(3)$$

$$\text{if } (T1(i, j) < 0) \quad DT(i, j) = 0 \dots\dots\dots(3)$$

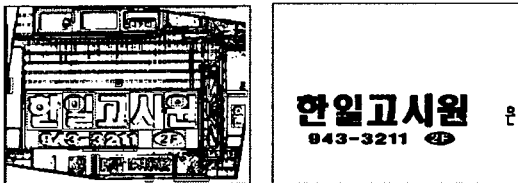
CS : Canny-소벨 에지 영상

T1 : 1차 텍스트 추출 후 영상 (가정:1-흑화소, 0 - 백화소)

실험 결과 비텍스트는 1차 텍스트 추출의 영상을 경계 추적한 흑화소의 수보다 2배 이상을 가지고 있다. 2차 텍스트 추출 연산 수행 후 (b)와 같이 최종 텍스트만 추출한 영상을 얻을 수 있다.

$$T2B(i, j) = \begin{cases} T1B(i, j) & \text{if } DTB_{cnt} > DBB_{cnt} \times 2 \dots\dots\dots(4) \\ 255 & \text{else} \end{cases}$$

·T2B : 2차 텍스트 추출 후 텍스트 후보 블록  
 T1B : 1차 텍스트 추출 후 텍스트 후보 블록  
 DTB<sub>cnt</sub> : 차영상 텍스트 후보 블록의 흑화소 수  
 DBB<sub>cnt</sub> : 경계추적영상 텍스트 후보 블록의 흑화소 수



(a) 차영상 (b) 2차 텍스트 추출 후 영상  
 [그림 6] 2차 텍스트 추출

### 3. 실험결과 및 고찰

실험은 게시판 영상, 교통 표지판, 책 표지, 상가간판 등 100장의 영상을 가지고 실험을 하였으며 실험자의 눈으로 인식 가능한 크기의 대상으로 단순한 영상과 복잡한 영상들을 크기에 상관없이 사용하였다. 실험결과 텍스트 추출률은 94.2%를 보였으며, 텍스트 추출 정확도는 90.3%를 보였다. 이는 텍스트와 비텍스트가 붙어 있거나 또는 텍스트와 배경사이의 명암 대비의 차가 현저히 적을 때 텍스트를 추출하지 못하는 경향을 보였다.

기울기 추출에 대해서는 전체 이미지에 대해서 94.3%의 추출률을 보였으며, 기존 연구에서 단지 허프 변환을 사용한 방법보다 기울기 추출률의 정확성을 현저하게 높일 수 있었다.

### 4. 결 론

본 논문은 텍스트가 포함되어 있는 자연 영상에서 텍스트는 주변 배경과 명도의 대비차가 크다는 것을 착안하여 명도 이미지에서 에지를 사용하여 텍스트들이 가지고 있는 특징들로 텍스트만을 추출하였다. 또한 기울기를 추출하기 위해 5x5 윈도우를 사용하여 가로, 세로 -45° ~+45°의 직선을 추출 후, Hough 변환과 본 논문에서 제시한 각도 추출 단계를 거쳐 각도를

추출하였다. 실험 결과 다양한 자연영상에서 텍스트 추출률과 기울기 추출에 대해 좋은 결과를 얻을 수 있었다. 앞으로 원근감이 있는 영상에 대해 원근 정보를 이용한 보정과 텍스트 후보영역을 보다 더 정확하게 추출하기 위한 연구가 필요하다.

### [참고문헌]

- [1] Jiangying Zhou, Daniel Lopresti, "Extracting Text from WWW Image", Proceedings of the 4th International Conference Document Analysis and Recognition (ICDAR '97), pp. 248-252, 1997
- [2] Anil K. Jain, Fellow, Bin Yu, "Document Representation and Its Application to Page Decomposition", IEEE Transactions On Pattern Analysis and Machine Intelligence, Vol. 20, No.3, pp. 294-305, 1997
- [3] Chuang Li, Xiaoqing Ding, Youshou Wu, "Automatic text location in natural scene images", Document Analysis and Recognition, 6th International Conf.. pp. 1069-1073, 2001
- [4] Yong-Bin K., Se-Young O., Hwan-Gue C., "Character grouping technique using 3D neighborhood graphs in raster map", Pattern Recognition Proc., Fourteenth International Conf., Vol. 2, pp. 1092-1094, 1998
- [5] E.G Johnson, "Printed text discrimination," Computer Graphics Image Processing, Vol. 3, No. 1, pp. 83-89, 1974
- [6] Bong-Kee Sin, "Locating characters in scene images using frequency features" Pattern Recognition, 2002. Proc. 16th International Conf. , Vol. 3, pp. 489-492, 2002
- [7] Y. Nakano, Y. Shima, H. Fujisawam J. Higashino and M. Fujinawa, "An Algorithm for the Skew Normalization of documnet Image" Proc. 10th Int. Conf. on Pattern Recognition, Atlantic City, USA, June pp. 8-13, 1990.
- [8] J. L. Fisher, "Logical Structure Descriptions of Segmented Document Images," Proc. 1st Int. Conf. on Document Analysis and Recognition, Saint-Malo, France, pp. 302-310, Sep. 1991