Journal of the Korea Academia-Industrial cooperation Society
Vol. 14, No. 2 pp. 885-892, 2013

http://dx.doi.org/10.5762/KAIS.2013.14.2.885

질감과 깊이 특징 기반의 문자영역 추출

장석우[1], 박영재[2], 허문행[1*]
[1]안양대학교 디지털미디어학과, [2]숭실대학교 컴퓨터학부

Character Region Extraction Based on Texture and Depth Features

Seok-Woo Jang[1], Young-Jae Park[2] and Moon-Haeng Huh[1*]
[1]Department of Digital Media, Anyang University
[2]School of Computing, Soongsil University

요 약 본 논문에서는 3차원의 입체영상으로부터 질감과 깊이 특징을 활용하여 영상 내에 존재하는 문자를 효과적으로 분할하는 방법을 제안한다. 제안된 문자 분할 방법은 크게 후보 문자영역 추출 단계, 문자영역 지역화 단계, 문자와 배경영역 분리 단계, 그리고 후보 문자영역 검증 단계의 네 가지 단계로 구성된다. 후보 문자영역 추출 단계에서는 입력된 영상에서 질감 특징을 이용해 문자영역이 존재하는 후보 영역을 분할하고, 문자영역 지역화 단계에서는 후보 문자영역 중에서 문자열만을 형성하는 영역을 추출한다. 그리고 문자와 배경 분리 단계에서는 지역화된 문자영역을 문자와 배경으로 분리하며, 후보 문자영역 검증 단계에서는 거리 특징을 활용하여 추출된 문자영역이 비 문자영역을 포함하지 않고 문자영역만을 포함하고 있는지를 최종적으로 검증한다. 실험에서는 제안된 방법을 여러 가지 영상에 적용하여 테스트 해 보았으며, 제안된 방법이 기존의 방법에 비해 보다 정확하게 문자영역을 추출함을 확인하였다.

Abstract In this paper, we propose a method of effectively segmenting character regions by using texture and depth features in 3D stereoscopic images. The suggested method is mainly composed of four steps. The candidate character region extraction step extracts candidate character regions by using texture features. The character region localization step obtains only the string regions in the candidate character regions. The character/background separation step separates characters from background in the localized character areas. The verification step verifies if the candidate regions are real characters or not. In experimental results, we show that the proposed method can extract character regions from input images more accurately compared to other existing methods.

Key Words : Character Region Extraction, Texture, Depth Feature, Localization, Verification

1. 서론

최근 들어, 빠른 속도의 유무선 인터넷, 다양한 저작도구의 출현, 대용량 저장장치의 발전 등에 힘입어 멀티미디어 자료가 기하급수적으로 보급되고 있다. 일반적으로, 멀티미디어 자료는 소리, 영상, 문자 등의 여러 미디어가 결합된 혼합 콘텐츠이다. 특히, 캡션과 같은 문자 정보는 영상의 내용 또는 의미를 파악하는 데 있어 매우 중요한 정보를 포함한다. 따라서 멀티미디어 자료의 색인 및 검색을 위한 문자 인식의 전 단계로 여러 동영상 내에 포함된 문자를 정확하게 분할하는 연구에 대한 필요성이 지속적으로 요구되고 있다[1,2].

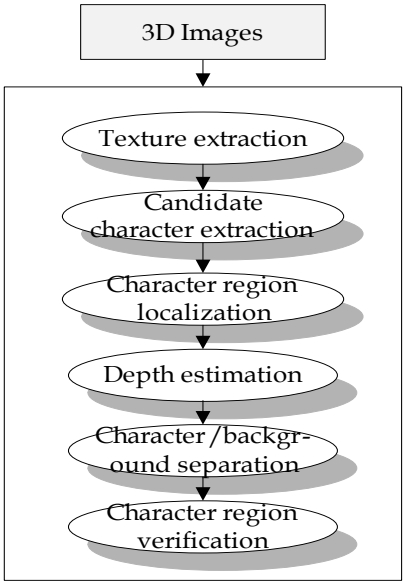
관련 문헌에서는 문자 영역을 분할하기 위한 다양한 연구 방법들이 소개되었다. [2]에서는 블록 단위로 영상을 분할한 후, 문자 영역에서 가로 및 세로 방향의 고주파 요소가 매우 크다는 특징을 활용하여 문자를 추출하

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2011-0021984)
*Corresponding Author : Moon-Haeng Huh (Anyang University)
Tel: +82-11-725-1511 email: moonh@anyang.ac.kr
Received November 14, 2012 Revised December 3, 2012 Accepted February 6, 2013

였다. [3]에서는 문자 영역의 에지를 강하게 남기고 나머지 에지들은 삭제함으로써 문자 영역을 획득하는 방법을 제안하였다. [4,5]에서는 채도 값을 이용하고 배경과 문자 사이의 색상 천이지도(transition map)을 생성하여 문자영역을 검출하였다. 이 외에도 문자 영역을 분할하는 다른 방법들이 계속 소개되고 있다[6].

이와 같이 문자를 분할하는 많은 기법들이 연구되어 오고 있으나 여전히 해결해야 할 문제들이 남아 있다. 특히, 기존의 많은 방법들은 주로 2차원 영상을 대상으로 문자 영역 추출을 시도하고 있다. 최근 들어, 3차원 스테레오 입체영상이 많이 보급되고 있으나 3차원적인 특징을 활용하여 3차원 영상에 존재하는 문자를 추출하는 연구는 현재 거의 찾아보기가 어려운 실정이다.

따라서 본 논문에서는 3차원의 스테레오 입체영상을 받아들여 질감과 깊이 특징을 기반으로 영상에 포함된 문자영역을 강건하게 추출하는 기법을 소개한다. Fig. 1은 본 논문에서 제안하는 문자영역 분할 알고리즘의 전체적인 흐름도를 보여준다.



[Fig. 1] Overall flow diagram

Fig. 1에서 확인할 수 있듯이 제안된 방법에서는 3차원으로 촬영된 입체영상을 받아들여 영상으로부터 문자가 존재할 것이라고 예상되는 후보 문자영역을 질감 특징을 이용하여 분할한다. 그런 다음, 후보 문자영역에서 문자열을 형성하는 부분만을 추출하는 문자영역 지역화를 수행한 후, 지역화된 문자열에서 배경을 분리하고 문자만을 추출한다. 마지막으로, 거리 특징을 활용하여 추출된 문자영역이 비 문자영역을 포함하지 않고 문자영역만을 포함하고 있는지를 최종적으로 검증한다.

1장에서는 본 연구를 수행하게 된 동기 및 배경, 그리고 전체적인 개요를 기술하였다. 2장에서는 후보 문자영역을 추출하는 기법에 대해 소개하고, 3장에서는 문자영역을 지역화하는 방법을 기술한다. 4장에서는 지역화된 문자 영역을 문자와 배경 영역으로 분리하는 기법을 설명하며, 5장에서는 후보 문자영역을 검증하는 방법을 설명한다. 그리고 6장에서는 제안한 문자영역 분할 방법의 성능을 비교 평가하기 위해서 수행한 실험결과를 보이며, 7장에서는 결론 및 향후 연구방향을 제시한다.

2. 후보 문자영역 추출

본 논문에서는 질감 특징을 이용하여 영상에서 후보 문자영역을 분할한다. 그리고 질감을 추출하기 위해서는 기본적으로 영상에서 문자가 존재하는 영역은 가로 및 세로 방향의 고주파 성분이 크다는 특성을 활용한다. 먼저, 영상을 8×8 화소의 블록으로 분할한 후, 블록 단위로 식 (1)과 같은 DCT(discrete cosine transform) 계수 c를 추출한다[7].

$$c_{uv} = \frac{1}{N} K_u K_v \sum_{x=0}^{N-1}\sum_{y=0}^{N-1} I_{xy} \cos\frac{\pi u(2x+1)}{2N} \cos\frac{\pi v(2y+1)}{2N} \quad (1)$$

식 (1)에서 u와 v는 수평과 수직 방향의 주파수를 나타내고, N은 블록 내의 화소의 개수를 나타내며, I는 영상의 명암값을 의미한다. 그리고 x와 y는 블록 내 화소의 좌표를 나타낸다. K는 계수로서 u 또는 v가 0일 경우에는 K=$1/\sqrt{2}$이며, 그렇지 않을 경우에는 K=1이 된다.

일반적으로, 문자영역은 비 문자영역 보다 수평방향 고주파 성분의 계수 값이 크므로, 각 블록에 대해 수평방향으로의 고주파 성분이 약한 블록들을 후보 블록에서 제거하기 위해 각 블록의 수평방향 고주파 성분을 식 (2)를 이용해 계산한다. 식 (2)에서 (i,j)는 각 블록을 나타내는 인덱스를 의미하고, v_1과 v_2는 수직 주파수의 범위를 나타낸다.

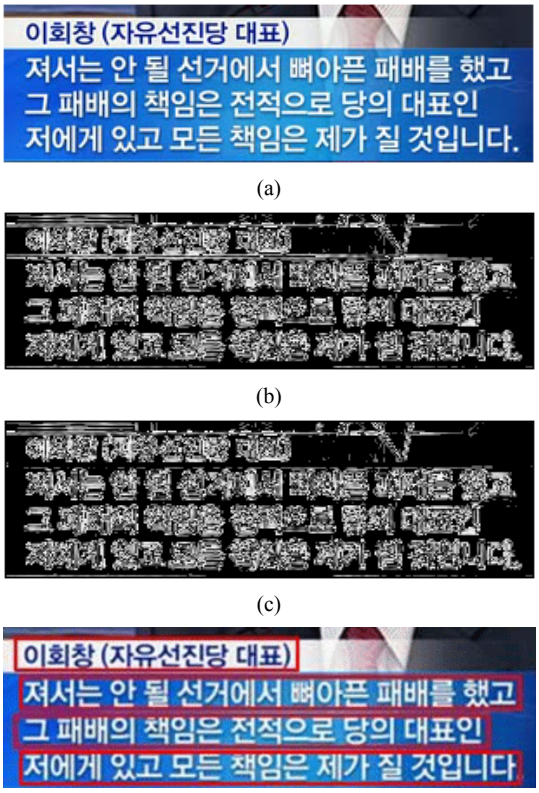
$$E_{hor}(i,j) = \sum_{v_1 \le v \le v_2} |c_{0v}(i,j)| \quad (2)$$

[Fig. 2] Candidate character region extraction
(a) Input image (b) Horizontal high frequency selection (c) Morphological operation (d) Vertical low frequency elimination (e) Dilation operation (f) Acquisition of character region MERs

본 논문에서는 전체 블록들의 수평 방향 고주파 성분의 평균보다 큰 블록들만을 후보 블록으로 선택하는데, Fig. 2 (b)와 같은 영상이 획득된다. 그리고 잡음 블록을 제거하고, 비 연결 블록들을 연결하기 위해 1×3 마스크를 사용하여 열림 후 닫힘 형태학 연산[8]을 수행하면 Fig. 2 (c)와 같은 영상을 획득한다. 그런 다음, Fig. 2 (d)와 같이 후보 문자영역 중에서 수직 방향 고주파 성분이 작은 블록들을 제거한다. 그리고 3×3 마스크를 이용하여 블록들에 팽창연산을 수행하면 문자영역이 Fig. 2 (e)와 같이 확대된다. 따라서 문자영역이나 문자 획의 끝부분이나 극히 일부분만을 포함하여 고주파 성분이 충분히 나오지 않아 검출되지 않았던 문자영역을 복원할 수 있다. 최종적으로, Fig. 2 (f)와 같이 검출된 문자영역에 대응하는 최소인접사각형(MER: minimum enclosing rectangle)을 획득한다. 그런 다음, 최소인접사각형에 해당하는 원본영상을 획득하여 다음 장에서 언급하게 될 문자영역 지역화의 입력으로 사용한다.

3. 문자영역 지역화

앞 단계에서 MER 단위로 추출한 후보 문자영역은 문자열의 정확한 위치도 파악하기 어렵고 비 문자영역도 다수 포함하여 있으므로, 문자열만을 형성하는 영역으로 최소화하여 추출하는 문자영역 지역화를 수행해야 한다. 이를 위해, 문자가 존재하는 영역은 문자의 획 주변에서 많은 양의 에지가 군집(cluster)을 형성한다는 특성을 활용한다. 본 논문에서는 소벨(Sobel) 에지 마스크를 이용해서 후보 문자영역에 해당하는 에지 맵을 추출한다.



[Fig. 3] Localization of character regions
(a) Input image (b) Edge map extraction (c) Horizontal line edge elimination (d) Character localization

그런 다음, 가로 방향 및 세로 방향으로 길게 연결되어 있는 라인 형태의 잡음 에지를 제거시킨다. 예를 들어, Fig. 3 (b)의 에지맵에서 가로 방향으로 라인 형태의 잡음 에지가 존재하는데, 수평 방향으로의 연속된 에지의 개수가 영상 너비를 기준으로 일정 비율 이상이 된다면 Fig. 3 (c)와 같이 해당되는 라인 에지는 제거한다. 그런 다음,

에지 맵을 가로 및 세로 방향으로 투영(projection)하여 줄 단위의 문자열을 산출한다. 이를 위해, 식 (3)과 같이 에지 히스토그램을 생성하고 임계값을 적용하여 행 또는 열에 에지가 밀집되어 있는지를 판별한다.

$$BiHisto(k) = \begin{cases} 1, \text{if } EHisto(k) \geqq Edge_{th} \\ 0, \text{ if } EHisto(k) < Edge_{th} \end{cases} \quad (3)$$

식 (3)에서 $EHisto(k)$는 k번째 에지 히스토그램을 의미하고, $Edge_{th}$는 임계값을 의미한다. 식 (3)을 이용해 구한 $BiHisto(k)$을 활용하여 행 또는 열 단위의 문자열을 추출할 수 있다.

그리고 행 단위의 문자열 검출 시에는 문자의 띄어쓰기를 고려해야 한다. 일반적으로, 문자열에서 띄어쓰기의 폭은 문자열을 이루는 높이보다 작으므로, 식 (4)를 만족할 경우에는 문자열이 연결된 구간으로 판단한다. 식 (4)에서 $Height_{MER}(t)$는 t번째 문자열 MER의 높이를 나타내고, $Valley(t,k)$는 t번째 문자열의 k번째 띄어쓰기 구간에 해당한다. 그리고 g는 상수값으로 0.6에서 0.8 사이의 값으로 설정한다.

$$Valley(t,k) \leq g \cdot Height_{MER}(t) \quad (4)$$

마지막으로, 지역화된 후보 문자열 중에서 MER의 높이가 문자인식이 불가능한 높이 이하이거나 이상일 경우, 그리고 후보 문자열 MER의 너비가 높이보다 작을 경우에는 문자 인식이 무의미하므로 후보 문자열에서 제외시킨다.

4. 문자와 배경 분리

문자영역을 지역화한 후에는 지역화된 문자영역을 문자와 배경 부분으로 이진화한다. 보통, 문서에서는 글씨는 검정색이고 배경은 흰색이라는 선행지식이 존재하나 본 논문의 대상 영상에는 이런 선행지식이 없기 때문에 문자영역에서 문자와 배경을 구분하고, 각각에 대한 명암값을 결정하는 과정이 필요한데, 다음과 같은 단계로 진행한다.

첫째, 컬러 영상을 그레이 영상으로 변환하고, Otsu의 방법[10]을 이용해 최적의 임계값을 히스토그램 이진화를 통해 통계적으로 구한 후, 이를 이용해 변환된 그레이 영상을 식 (5)와 같이 이진화한다.

$$B(x,y) = \begin{cases} 255, & \text{if } I(x,y) \geq T_1 \\ 0, & \text{if } I(x,y) < T_1 \end{cases} \quad (5)$$

식 (5)에서 $I(x,y)$는 그레이 영상이며, $B(x,y)$는 이진화된 영상이다. 둘째, 이진화된 영상으로부터 식 (6)을 이용해 두 개의 이진화 맵을 생성한다.

$$MAP_{white}(x,y) = \begin{cases} True, & \text{if } B(x,y) = 255 \\ False, & \text{if } B(x,y) = 0 \end{cases} \quad (6)$$

$$MAP_{black}(x,y) = \begin{cases} True, & \text{if } B(x,y) = 0 \\ False, & \text{if } B(x,y) = 255 \end{cases}$$

셋째, 영상을 스캔하면서 식 (7)을 k가 0인 시점부터 만족할 때까지 반복적으로 수행하며, 검사가 된 화소는 후에 검사가 재 수행되지 않도록 값을 False로 변경한다. 여기서 Map_q는 MAP_{black}과 MAP_{white}를 포함한다.

$$\begin{aligned} &\text{if } ((Map_q(x+k,y+k) == True) \text{ and} \\ &\quad (Map_q(x+k-m,y+k) == True) \text{ and} \\ &\quad (Map_q(x+k,y+k-n) == True)) \\ &\qquad then\ k = k+1 \\ &where\ 0 \leq m,n \leq k,\ k\ begins\ with\ 0 \end{aligned} \quad (7)$$

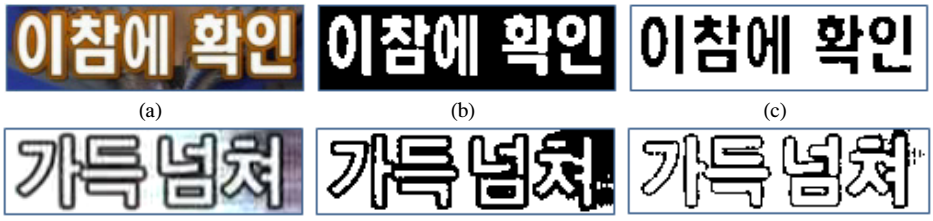
넷째, 세 번째 과정 수행 중 식 (8)을 만족하면 i번째 최대내접사각형의 한 변의 길이 IS가 획득된다.

$$\begin{aligned} &\text{if}((Map_q(x+k,y+k) == False) \text{ or} \\ &\quad (Map_q(x+k-m,y+k) == False) \text{ or} \\ &\quad (Map_q(x+k,y+k-n) == False)) \\ &\qquad then\ IS_i^q = k+1 \\ &where\ 0 \leq m,n \leq k \end{aligned} \quad (8)$$

이 과정을 수행하면 Map_q가 여러 개의 최대내접사각형으로 분할된다. 그리고 MAP_{black}과 MAP_{white}에서 각각 진행되기 때문에 나누어진 최대내접사각형의 개수는 상이할 수도 있다. 다섯째, 현재까지의 과정으로부터 얻어진 최대내접사각형들의 한 변의 길이를 이용하여 최대내접사각형 넓이의 평균 E와 분산 σ^2을 식 (9)를 이용하여 구한다. 여기서N_q는 Map_q에서 나누어진 최대내접사각형의 개수이다.

$$E(IS^q) = \frac{1}{N_q}\sum_{i=1}^{N_q}(IS_i^q)^2 \quad (9)$$

$$\sigma_q^2 = \frac{1}{N_q}\sum_{i=1}^{N_q}\left\{E(IS^q) - (IS_i^q)^2\right\}^2$$



(a) (b) (c)

(d) (e) (f)

[Fig. 4] Character extraction
(a) Input image (b) Binarization (c) Separation characters from background (d) Input image (e) Binarization (f) Separation characters from background

여섯째, 각 맵에서 추출된 최대내접사각형 분산의 비교를 통해 각 화소의 명암값을 식 (10)과 같이 결정한다. 만일, black에 해당하는 맵의 분산이 작다면 첫 번째 과정에서 구한 이진화 영상을 그대로 유지하고, 그렇지 않을 경우에는 이진화 영상을 역변환한다. 이런 과정을 수행하면 문자와 배경 화소의 명암값이 결정된 영상 B'(x,y)가 생성된다. 본 논문에서는 기본적으로 문자의 명암값은 0, 그리고 배경의 명암값은 255로 설정하였다.

$$B^{'}(x,y)=\begin{cases} B^{-1}(x,y), & \text{if } \sigma^2_{white} < \sigma^2_{black} \\ B(x,y), & \text{if } \sigma^2_{white} \geq \sigma^2_{black} \end{cases} \quad (10)$$

Fig. 4는 위에서 설명한 내용을 문자영상에 적용한 예를 보여준다. Fig. 4 (a)는 원본영상을 나타내고, Fig. 4 (b)는 Otsu의 방법을 기반으로 이진화한 결과를 나타내며, Fig. 4 (c)는 문자와 배경을 분리한 결과를 보여주며, Fig. 4 (d), (e), (f)도 동일한 처리 결과를 보여준다. Fig. 4에서 확인할 수 있듯이 Fig. 4 (a)는 글자의 테두리 명암값이 배경의 명암값과 유사하여 글자의 테두리가 배경에 흡수됨으로써 Fig. 4 (c)와 같이 문자와 배경의 분리가 잘 수행되었다. 그러나 Fig. 4 (d)는 글자의 테두리 명암값이 문자 또는 배경과 상이하여 Fig. 4 (f)와 같이 문자와 배경의 분리가 정상적으로 진행되지 않았다. 이런 문제를 해결하기 위해 기존의 여러 가지 방법들이 제안되었으나 결과가 좋지 못하였다. 그러나 본 논문에서는 3차원 특징인 거리 정보를 이용해 이 문제를 해결한다. 이에 대한 자세한 설명은 다음 장에서 기술한다.

5. 문자영역 검증

앞에서 언급한 문자영역 지역화 과정을 통해서 추출된 후보 문자영역들도 여전히 비 문자영역들을 포함할 수 있다. 따라서 본 연구에서는 3차원의 깊이 정보와 인공신경망(artificial neural network)을 활용하여 후보 문자영역을 검증함으로써 비 문자영역을 제외하고 실제 문자영역만을 선택한다.

기존의 인공 신경망만을 이용한 후보 문자영역 검증 알고리즘도 어느 정도의 정확성을 보이지만 문자영역과 유사한 특징이 추출되는 나뭇잎, 창문, 그리고 철조망 등과 같은 후보 영역들에 대해서는 오 분류가 많이 발생한다.

본 논문에서는 3차원의 깊이 정보와 인공 신경망을 결합하여 문자영역을 검증하는데, 다음과 같은 단계를 통해 수행한다.

(1) 좌 · 우 입력영상으로부터 스테레오 정합을 통해 깊이 영상을 추출한다.
(2) 지역화된 문자영역의 에지 부분에 해당하는 화소의 깊이 정보를 획득한 후, 추출된 깊이 값들의 평균(μ) 및 표준편차(σ)를 구한다.
(3) 보통, 문자영역에 해당하는 깊이 값은 유사하므로 σ가 Th_σ 이상인 문자영역들은 비문자영역이라고 판단하고 1차적으로 제거한다.
(4) 인공신경망을 이용하여 2차적으로 문자영역을 검증한다. 즉, 1차적으로 검증된 문자영역에 대한 에지의 방향을 16개 방향으로 양자화하고, 각 방향에서의 에지의 크기를 인공신경망의 입력으로 받아들인 후, 배경 영역에 대해 학습한 값과 비교하여 2차적으로 비 문자영역을 제거한다.

본 논문에서는 3차원의 깊이 정보를 추출하기 위해서 여러 가지 기존의 방법 중에서 최근 들어 좋은 성능을 발휘한다고 알려진 그래프 컷(graph cut)[9] 기반의 스테레오 정합(stereo matching) 알고리즘을 적용한다. 이를 위해, 입력영상을 그레이 영상으로 변환한 후 좌우 영상 사이의 정합을 통해서 변이 정보를 추출한다.

그리고 본 논문에서 사용한 인공 신경망은 바이어스 노드를 포함하지 않는 32개의 입력노드, 바이어스 노드를 포함하지 않는 28개의 노드로 구성된 1개 층의 은닉층, 그리고 2개의 노드로 구성된 출력층으로 이루어져 있다. 인공신경망에 사용되는 특징은 후보 문자영역의 전경과 배경 각각에 대해 에지의 방향에 따른 크기를 사용하였다. 식 (11)과 식 (12)은 특징에 사용되는 에지의 방향과 크기를 구하는 식이다.

$$E_{theta}(x,y) = \tan^{-1}\frac{I(x,y+1)-I(x,y-1)}{I(x+1,y)-I(x-1,y)} \quad (11)$$

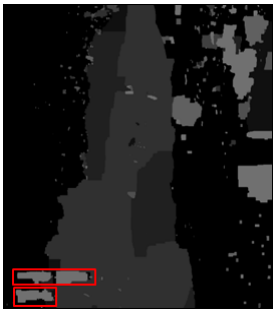
$$E_{magnitude}(x,y) = \sqrt{(I(x,y+1)-I(x,y-1))^2+(I(x+1,y)-I(x-1,y))^2} \quad (12)$$

후보 문자영역의 전경과 배경 각각에 대해 산출된 에지의 방향을 16개 방향으로 양자화해서 인공신경망의 입력으로 사용되는 32개의 히스토그램을 생성한다. 여기서 1번에서 16번까지는 전경에서 추출한 에지의 크기를 각 방향에 따른 빈(bin)에 누적한 것이고, 17번부터 32번까지는 배경에서 추출한 에지의 크기를 각 방향에 따른 빈에 누적한 것이다. 그런 다음, 해당하는 히스토그램을 전경 에지의 전체 크기와 배경 에지의 전체 크기로 나누어 정규화한다.

6. 실험결과

본 논문의 실험을 위한 컴퓨터는 인텔 Pentium Core 2 Duo의 2.66GHz CPU와 8GB의 메모리를 사용하였고, 운영체제로는 마이크로소프트사의 Windows 7을 사용하였다. 그리고 본 논문에서 제안한 3차원의 깊이 정보를 활용한 문자영역 분할 알고리즘의 구현을 위해 마이크로소프트사의 Visual C++ 2008 통합 개발 환경을 이용하였다. 또한, 실험에서 사용할 문자 영상 데이터베이스를 구축하기 위해서 다양한 종류의 입체영상을 수집하여 활용하였다.

Fig. 5는 본 논문에서 제안한 방법을 사용하여 최종적으로 추출한 문자영역을 보여준다. Fig. 5 (a)는 좌 · 우 영상으로부터 추출한 깊이 영상을 보여주는데, 좌측 하단에 표시된 사각형 영역이 문자라고 판단된 영역이다. 그리고 Fig. 5 (b)는 좌 · 우 원본 영상에서 추출된 문자영역에 해당하는 부분을 사각형으로 표시하여 보여주고 있다.



(a)

(b)

[Fig. 5] Segmentation of character regions
(a) Depth image
(b) Segmented left and right character regions

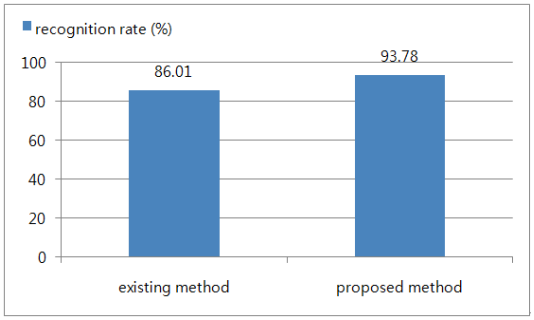
본 논문에서는 제안된 문자영역 분할 방법의 성능을 정량적으로 평가하기 위해서 비 문자영역을 포함하지 않은 정확하게 지역화된 문자열과 전체 지역화된 문자열의 비율을 백분율(%)로 표시한 식 (13)과 같은 정확도 척도를 사용하였다. 그리고 Table 1은 기존의 인공신경망을 이용한 방법과 제안된 3차원의 깊이 정보를 활용한 문자영역 검증 방법의 정확도를 수치적으로 보여준다. 전반적으로, Fig. 3의 문자영역과 같이 인위적으로 삽입된 배경이 있는 경우에는 배경 영역이 단순하여 이진화가 정확하게 수행되므로 기존의 방법과 제안된 방법 모두 문자 분할 정확도가 높았다. 그러나 Fig. 4와 같이 인위적인 배경이 없는 문자의 경우 기존의 방법에서는 문자와 배경의 이진화가 어려워 결과가 좋은 못한 반면 제안된 방법은 깊이 정보를 활용하여 좋은 결과를 얻을 수 있었다.

$$accuracy = \frac{No\ of\ accurately\ localized\ characters}{No\ of\ localized\ characters} \times 100 \quad (13)$$

[Table 1] Verification of character regions

Method	Number of localized characters	Number of accurately localized characters	Verification accuracy (%)
Existing NN-based method	278	253	91.0
Proposed 3D depth-based method	263	251	95.4

Table. 1에서 확인할 수 있듯이 제안된 방법은 3차원 깊이 정보의 도움으로 비 문자영역 추출을 상당수 감소시키므로 문장 영역 검증의 정확도를 향상시켰다. 그리고 Fig. 6은 앞 단계에서 추출한 문자열 영역을 문자 인식기에 입력하여 문자인식을 수행한 결과를 그래프로 보여준다. 본 논문에서는 문자 인식률이 좋다고 알려진 READ IRIS Pro 10[12]을 이용하여 문자 인식을 수행하였다. 예상했던 대로 기존의 방법보다 제안된 문자분할 방법에 기반한 문자 인식이 보다 높은 인식률을 보였다.



[Fig. 6] Result of character recognition

7. 결론

본 논문에서는 3차원으로 제작된 스테레오 입체영상에 존재하는 문자영역을 질감과 깊이 정보를 활용하여 효과적으로 분할하는 알고리즘을 제안하였다.

먼저, 입력 영상으로부터 문자가 위치한다고 예상되는 후보 문자영역을 질감 정보를 이용하여 추출하였으며, 후보 문자영역 중에서 배경 영역을 제외하고 문자열이 존재하는 영역만을 추출하는 지역화를 수행하였다. 그런 다음, 깊이 정보를 활용하여 후보 문자영역을 검증함으로써 비 문자영역을 제외하고 문자영역만을 분할하였다. 실험에서는 제안된 방법이 기존의 2차원 위주의 문자영역 분할 방법에 비해 강건하게 문자영역을 추출함을 확인하였다.

향후에는 좀 더 다양한 입체영상을 입력받아 본 논문에서 제안한 문자영역 분할 알고리즘의 성능을 최적화할 예정이다. 그리고 왜곡이 심한 문자 영역에 대한 분할도 시도할 계획이다.

References

[1] K.-J. Ahn and G.-Y. Kim, "Determing Intensity Value of Characters and Backgrounds on Caption," In Proc. of the Conference of the Korea Society of Computer and Information, Vol. 18, No. 2, pp. 125-128, 2010.

[2] X. Qian, G. Liu, H. Wang, and R. Su, "Text Detection, Localization, and Tracking in Compressed Video," Signal Processing: Image Communication, Vol. 22, No. 9. pp. 752-768. Oct. 2007.
DOI: http://dx.doi.org/10.1016/j.image.2007.06.005

[3] M. R. Lyu, J. Song, and M. Cai. "A Comprehensive Method for Multilingual Video Text Detection, Localization, and Extraction," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp. 243-255, Feb. 2005.
DOI: http://dx.doi.org/10.1109/TCSVT.2004.841653

[4] W. Kim and C. Kim, "A New Approach for Overlay Text Detection from Complex Video Scene," Journal of the Korean Society of Broadcast Engineers, Vol. 13, No. 4, pp. 544-553, Jul. 2008.
DOI: http://dx.doi.org/10.5909/JBE.2008.13.4.544

[5] W. Kim and C. Kim, "A New Approach for Overlay Text Detection and Extraction from Complex Video Scene," IEEE Transactions on Image Processing, Vol. 18, No. 2, pp. 401-411, Feb. 2009.
DOI: http://dx.doi.org/10.1109/TIP.2008.2008225

[6] P. P. Roy, U. Pal, J. Lladós, and M. Delalandre, "Multi-oriented Touching Text Character Segmentation in Graphical Documents Using Dynamic Programming," Pattern Recognition, Vol. 45, No. 5, pp. 1972-1983, May 2012.
DOI: http://dx.doi.org/10.1016/j.patcog.2011.09.026

[7] K.-J. An, S.-I. Joo, G.-Y. Kim, and H.-I. Choi, "An Improved Method for Detecting Caption in Image Using DCT-Coefficient and Transition- map Analysis," Journal of the Korea Society of Computer and Information, Vol. 16, No. 4, pp. 61-71, Apr. 2011.
DOI: http://dx.doi.org/10.9708/jksci.2011.16.4.061

[8] O. Strauss and F. Comby, "Variable Structuring Element-based Fuzzy Morphological Operations for Single Viewpoint Omnidirectional Images," Pattern Recognition, Vol. 40, No. 12, pp. 3578- 3596, Dec. 2007.
DOI: http://dx.doi.org/10.1016/j.patcog.2007.05.003

[9] A. Jama and S. Rakshit, "Augmenting Graph Cut with TV-L Approach for Robust Stereo Matching," In Proc. of the International Conference on Image Information Processing, pp. 1-6, 2011.
DOI: http://dx.doi.org/10.1109/ICIIP.2011.6108868

[10] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," IEEE Transactions on Systems, Man and Cybernetics, Vol. 9, No. 1, pp. 62-66, 1979.
DOI: http://dx.doi.org/10.1109/TSMC.1979.4310076

[11] J. Song, M. Cai, and M. R. Lyu., "A Robust Statistic Method for Classifying Color Polarity of Video Text," In Proc. of the International Conference on Multimedia and Expo, Vol. 1, pp. 385-388, 2003.

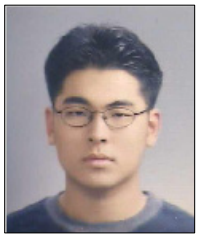
[12] http://www.irislink.com

장 석 우(Seok-Woo Jang) **[정회원]**

- 1995년 2월 : 숭실대학교 전자계학과 (공학사)
- 1997년 2월 : 숭실대학교 컴퓨터학과 (공학석사)
- 2000년 8월 : 숭실대학교 컴퓨터학과 (공학박사)
- 2009년 3월 ~ 현재 : 안양대학교 디지털미디어학과 조교수

<관심분야>
로봇비전, 증강현실, HCI, 비디오 색인 및 검색, 등

박 영 재(Young-Jae Park) **[정회원]**



- 2005년 2월 : 청운대학교 컴퓨터과학과 (공학사)
- 2008년 2월 : 숭실대학교 컴퓨터학과 (공학석사)
- 2008년 3월 ~ 현재 : 숭실대학교 컴퓨터학과 박사과정

<관심분야>
컴퓨터 비전, 영상처리, 패턴인식 등

허 문 행(Moon-Haeng Huh) **[정회원]**

- 1979년 2월 : 숭실대학교 전자계학과 (공학사)
- 1989년 2월 : 연세대학교 전자계산학과 (공학석사)
- 2003년 8월 : 충북대학교 전자계산학과 (공학박사)
- 2004년 3월 ~ 현재 : 안양대학교 디지털미디어학과 교수

<관심분야>
디지털콘텐츠, 유비쿼터스, 미디어 컨버전스 등