

## DCT계수와 천이지도 분석을 이용한 개선된 영상 내 자막영역 검출방법

안 권 재\*, 주 성 일\*\*, 김 계 영\*\*\*, 최 형 일\*\*\*\*

### An Improved Method for Detecting Caption in image using DCT-coefficient and Transition-map Analysis

Kwon-Jae An\*, Sung-Il Joo\*\*, Gye-Young Kim\*\*\*, Hyung-Il Choi\*\*\*\*

#### 요 약

본 논문은 DCT계수와 천이지도 분석을 이용하여 영상 내 자막영역을 검출하는 방법에 대해 제안한다. 기존 DCT계수 분석방법을 이용한 문자영역탐지 방법은 검출률은 높으나 오검출률이 매우 높은 단점이 있고, 천이지도를 이용한 문자영역 탐지 방법은 임계값이 정적이기 때문에 문자영역 검증단계에서 실제문자영역이 기각되는 일이 빈번히 발생한다. 이러한 문제점을 해결하기 위해 DCT계수 분석방법을 이용하여 유망문자영역맵을 작성하고 이를 천이지도를 이용한 문자영역탐지 방법에 적용하여 임계값을 단계별로 정한다. 그 결과로서 DCT계수 분석을 이용한 문자영역검출방법에 비해 오검출률이 크게 감소하였으며, 기존 천이지도를 이용한 문자영역검출 방법보다 검출률이 크게 향상되었다.

▶ 키워드 : 문자탐지, 문자인식, 이산코사인변환, 천이지도

#### Abstract

In this paper, we proposed the method for detecting text region on image using DCT-coefficient and transition-map analysis. The detecting rate of traditional method for detecting text region using DCT-coefficient analysis is high, but false positive detecting rate also is high and the method using transition-map often reject true text region in step of verification because of sticky threshold. To overcome these problems, we generated PTRmap(Promising Text Region map) through DCT-coefficient analysis and applied PTRmap to method for detecting text region using transition map. As the result, the false positive detecting rate decreased as compared with the method using DCT-coefficient analysis, and the detecting rate increased as compared with the method using transition map.

▶ Keyword : Text Detection, Character Recognition, DCT, Transition map

• 제1저자 : 안권재 • 교신저자 : 김계영

• 투고일 : 2011-01-01, 심사일 : 2011-01-15, 게재확정일 : 2011-01-20

\* 송실대학교 컴퓨터학과(Dept. of Computer, SoongSil University)

\*\* 송실대학교 미디어학과(Dept. of Media, SoongSil University)

\*\*\* 송실대학교 컴퓨터학부(School of Computer, SoongSil University)

\*\*\*\* 송실대학교 글로벌미디어학부(School of Global Media, SoongSil University)

※ 본 연구는 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 콘텐츠 산업기술지원사업의 연구결과로 수행되었음

## 1. 서론

멀티미디어 콘텐츠는 사람에게 다양한 정보를 복합적으로 전달한다. 동영상은 이러한 멀티미디어 콘텐츠의 대표적인 콘텐츠라고 할 수 있으며, 구성 미디어로 영상, 음성, 문자 등이 있다. 동영상은 위와 같은 다양한 미디어들을 각각으로 사용하거나 복합적으로 사용하여 제작자가 표현하고자 하는 의미를 시청자에게 전달할 수 있다. 또한 사용자는 이러한 동영상을 시청함으로써 영상, 음성, 문자 등과 같은 다양한 정보들을 한 번에 취득할 수 있다는 장점이 있다. 이중 동영상에 인위적으로 삽입된 자막은 동영상 제작자가 현재 장면에 대한 내용을 함축적으로 표현하고자 사용하는 경우가 대부분이기 때문에 이러한 자막에 대해 의미적정보를 추출하여 동영상 혹은 동영상의 일부 장면에 대한 색인을 할 수 있다. 동영상에서 자막을 이용하여 의미적정보를 추출하는 절차는 크게 3단계로 구분할 수 있다. 첫 번째로 문자영역 탐지 혹은 지역화이며, 두 번째는 문자추출이고, 마지막은 문자인식이다. 각 선행 단계의 검출률 혹은 정확도는 후행 단계 및 전체 성능에 영향을 미친다. 그렇기 때문에 최선행 절차인 문자영역 탐지는 매우 중요한 절차이며, 이에 대한 많은 연구들이 꾸준히 수행되어 왔다. 문자영역을 탐지는 주로 영상에서 질감, 에지, 색상 정보를 이용하여, 이들을 혼용하여 사용하기도 한다.

질감 정보를 이용하는 방법은 DCT 혹은 웨이블릿 계수를 분석하는 방법들이 대표적이다[1-3]. 이중 Zhong의 방법[3]은 DCT계수를 분석하여 문자영역을 찾는다. 이 방법은 I프레임 및 JPEG 영상에서 DCT계수를 바로 추출하여 사용하면 DCT 수행시간을 생략할 수 있어 매우 빠르며, 검출률이 99%로 매우 높고, 오검출이 전체 블록 개수에 비하면 미미한 수준이다. 하지만 검출된 블록의 개수와 오검출된 블록의 개수를 놓고 오검출률(FPR : False Positive Rate = 오검출 개수 / 총검출 개수)을 판단하였을 경우, 오검출률이 매우 높은 것을 알 수 있다. 또한 블록단위로 문자영역을 탐지하기 때문에 화소단위로 문자영역을 탐지하는 다른 방법들에 비해 정확한 문자영역을 검출 할 수 없으며 문자가 클 경우 블록단위로 볼 때 문자의 극히 일부만을 포함하여 고주파가 검출되지 않음으로 검출을 할 수 없는 단점이 있다. 에지 정보이용 방법은 문자영역 및 주변에서는 비문자영역에 비해 다수의 에지가 검출되는 특성을 이용하여, 에지의 군집지역을 찾는다[4-5]. Lyu 방법[5]은 이러한 에지 정보를 이용하는 방법의 하나다. 해당 방법은 영상을 레벨 단위로 축소하여 각 레벨에서 Sobel마스크를 이용한 에지를 검출을 하였다. 이후 문자 획에서 손실될

수 있는 에지정보를 보정하는 보정식을 사용하여 에지를 보정하였다. 또한 에지 정보를 이용하는 문자영역 탐지 기법의 단점인 복잡한 배경에 강건하지 못한 것을 극복하기 위하여 해당 논문에서 제안된 히스테리시스 마스크를 이용하여 문자영역에 지는 살리고 비문자영역 에지를 감쇠 시켰다. 이후 재귀적 X-Y컷방법인 coarse-to-fine projection을 이용하여 남은 에지가 군집된 지역을 찾음으로서 문자영역지역화를 하였다. Lienhart의 방법[4]은 RGB 각 채널 별 에지를 구하여 하나의 에지맵을 생성 한 후 일정 윈도우 크기로 탐색을 하며 인공 신경망을 이용하여 문자영역을 검증하는 방법이다. 색상 및 명암값을 정보를 이용하는 방법은 색상 및 명암값을 양자화나 전역적 이진화를 이용하여 주변에 비슷한 크기 및 성질의 객체가 존재하는지 검사는 방법이 대부분이다. 이러한 방법들은 비압축 영상에서는 좋은 성능을 보이지만, 압축영상에서는 화소정보가 훼손되어 정확도가 떨어진다[8-9]. 최근 Kim은 인위적으로 삽입된 영상이 원배경과 채도 및 명암값 차이가 많이 나는 점을 이용하여 문자영역을 찾고 압축영상에서 화소정보가 훼손되는 색상변경 효과가 비문자영역보다 문자영역에서 LBP 패턴을 다양하게 한다는 특성을 이용하여 문자영역을 검증하는 방법이 있다[6-7]. 이 방법은 단순한 처리로서 빠르고 비교적 복잡한 배경에서도 좋은 결과를 얻을 수 있으며, 결과로서 비교적 정확한 문자영역의 문자영역 경계를 기점으로 하는 MER(Minimum Enclosed Rectangle)을 얻을 수 있는 장점이 있다. 하지만 문자영역검출 단계 및 검증단계에서 고정적인 임계값을 사용하기 때문에 자막에 따라 채도대비와 명암대비가 낮은 경우는 문자영역을 찾지 못하는 경우가 있고 글자의 크기가 상이한 문자열 등에서 검증에 실패하는 경우가 종종 발생하게 된다. 마지막으로 프레임 간의 차연산을 이용한 문자영역 방법도 다수 존재하며, 질감, 에지, 색상 정보를 이용하는 방법과 혼용하여 사용하기도 한다. 본 논문에서는 정지영상 한 장을 기준으로 문자영역을 탐지하는 방법에 대해 설명한다. 단일 프레임에서 얼마나 문자영역을 정확히 탐지하는가가 선행적으로 해결되어야 연속되는 프레임에서 검출률이 상승하기 때문이다. 이를 위해 본 논문에서는 단일 영상에서 수행이 가능하고 검출률이 우수한 DCT 계수 분석을 이용한 문자영역검출 방법인 Zhong의 방법과 비교적 정밀한 문자열 단위 문자영역 탐지가 가능한 방법을 융합하여 장점을 살리고 Zhong의 방법이 FPR수치가 높은 단점과 Kim[7]의 검증단계에서 기각 될 수도 있는 단점을 보완한다.

본 논문의 구성은 2장에서는 블록 단위 DCT계수 분석과 정에서 유망 블록에 대해 설정하여 유망문자영역맵을 생성하는 방법을 설명한다. 3장에서는 유망문자영역맵을 기초로 하

여 Kim방법의 단점인 고정 임계값을 단계별로 설정으로 향상된 문자영역탐지 방법에 대해 설명한다. 4장은 실험결과에 대해 기술하고 5장에서 본 논문의 결론을 맺는다.

## II. DCT계수 분석을 이용한 유망문자영역맵 생성

유망문자영역맵은 입력영상에서 문자가 있을 가능성이 높은 영역을 탐지하여 화소 단위로 기록한 맵이며, 생성하는 방법은 Zhong의 DCT계수 분석을 이용한 문자영역탐지방법에서 각 단계별로 유망한 블록 내 존재하는 화소에 대해 해당 정보를 기록한다. Zhong의 방법을 사용하는 이유는 서론에서 언급하여 듯이 검출률이 99%로 매우 우수하여 후보영역을 찾기에 매우 적합한 방법이기 때문이다.

영상을 블록별로 DCT변환을 수행하였을 때 문자영역은 비문자영역과 비교하였을 때 수평, 수직방향의 주파수에 고주파가 존재하며, 가로, 세로 획이 주가 되는 알파벳과 다르게 한자의 경우 대각선 획도 다수 존재하기 때문에 대각 방향으로도 고주파가 존재한다[2-3]. 이러한 특성에 기반한 Zhong의 방법과 마찬가지로 입력영상에서 문자영역이 존재할 가능성이 큰 영역을 탐지하며 각 단계 별로 유망문자영역맵을 생성한다.

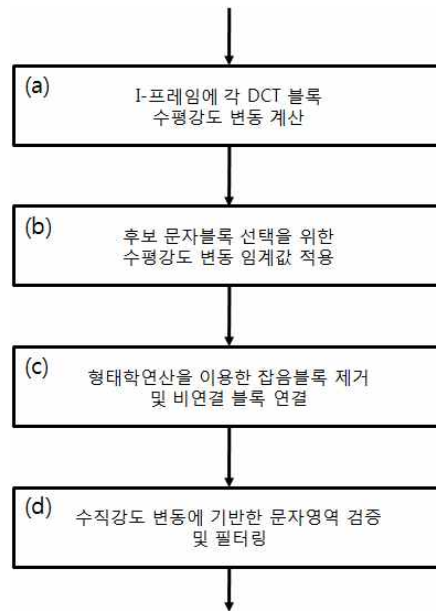


그림 1. Zhong방법의 절차도  
Fig. 1. Flow of Zhong's Method

[그림 1]은 Zhong 방법의 전체적인 흐름이며, 유망문자영역맵 생성에서 진행되는 전체적인 흐름은 같으나, Zhong 방법의 과정 중 유망문자영역맵이 생성되는 시점이 있다.

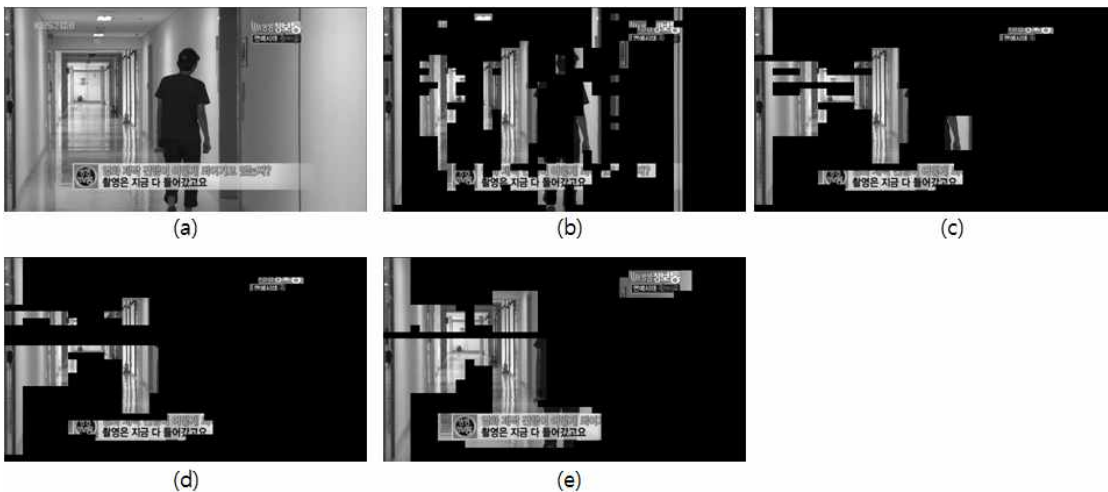


그림 2. Zhong방법의 문자영역 탐지 절차. (a) 입력영상. (b) 수평문자에너지가 높은 블록 선택. (c) 형태학연산을 이용한 잡음 제거 및 후보영역 검출. (d) 조건에 의한 후보문자영역 제거. (e) 팽창연산 후 최종 결과 영상.

Fig. 2. Flow of Zhong's Method for text region detection. (a) input image. (b) selecting high horizontal text energy (c) Reducing noise block and detecting candidate region using morphological operation (d) filtering candidate text region through condition (e) result image after dilation

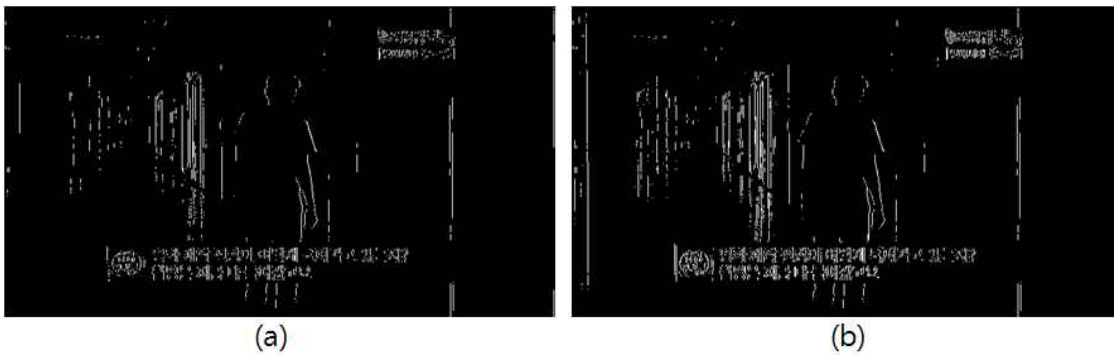


그림 3. 기존 천이지도와 수정된 천이지도 비교 (a) 기존 천이지도 (b) 본 논문에서 수정된 천이지도  
 Fig. 3. Comparisons between the traditional transition map and the modified transition map in this paper.  
 (a) the traditional transition map. (b) the modified transition map.

1차적으로 유망문자영역맵이 생성되는 시점은 [그림 1]의 (d)가 종료되는 직후이며, 최종 유망문자영역맵이 생성되는 시점은 형태학 연산 적용 이후이다. [그림 1] (a)에서 I프레임 내부의 각 블록에서 수평강도의 변동에 대해 계산하여 후보 블록을 찾는다. 방법은 각 블록에서 수평문자에너지  $E_{hor}$ 를 식(1)을 이용하여 계산하고, 블록들의 수평문자에너지 평균값을 구하여 이러한 평균의 1.45배 이상의 에너지값을 갖는 블록들만 선택적으로 남긴다. 해당 결과는 [그림 2] (b)에서 보는 것과 같다.

$$E_{hor}(i,j) = \sum_{v_1 \leq v \leq v_2} |c_{0v}(i,j)| \dots\dots\dots (1)$$

식(1)에서  $i, j$ 는 블록 단위의 인덱스를 나타내며,  $c$ 는 해당 블록의 DCT 계수 값들을 뜻하고,  $v$ 는 수직 주파수이며,  $v_1, v_2$ 는  $E_{hor}$ 를 구하기 위한 수직주파수의 범위로서 각각  $v_1 = 2, v_2 = 6$ 의 값을 갖는다.

후보블록들은 [그림 2]의 (b)에서 보는 것과 같이 잡음 블록들이 섞여 있고 문자영역을 충분히 포함하지 못한다. 이러한 문제를 해결하기 위해 [그림 1]의 (c)에서 보는 것과 같이 형태학연산을 수행하여 준다. 형태학연산은 기본구조 1 X 3의 마스크를 이용하여 열림 후 닫힘을 수행하며, 해당 연산과정을 통해 1차후보문자영역을 얻게 된다. 결과는 [그림 2]의 (c)와 같다.

이후 [그림 1]에는 나와 있지 않지만 1차후보문자영역을 조건에 의해 나누어 2차후보문자영역을 만든다. 조건은 각 1차후보문자영역들에서 가장 너비가 긴 행의 길이의 50%보다 작은 행의 길이를 갖는 행을 기준으로 1차후보문자영역을 나

눠 2차후보문자영역을 생성한다. 이후 2차후보문자영역에 대해 [그림 1]의 (d)에 해당하는 검증을 수행하게 된다.

검증은 문자영역의 수직강도 변동은 수평강도 변동과 마찬가지로 크다는 특성을 이용한다. 방법은 식(2)와 같이 각 블록단위로 수직강도 변동을 수직문자에너지  $E_{ver}$ 를 산출하여 구한다. 이후 식(3)과 같이 2차후보문자영역을 이루는 블록들을 행 단위로 평균 수직문자에너지  $\epsilon_{ver}^R$ 를 구한다. 만약 어떠한 2차후보문자영역이 한 행이라도 높은  $\epsilon_{ver}^R$ 를 갖지 못한다면 해당 2차후보문자영역은 기각이 되며, 검증은 종료된다.

$$E_{ver}(i,j) = \sum_{u_1 \leq u \leq u_2} |c_{u0}(i,j)| \dots\dots\dots (2)$$

$$\epsilon_{ver}^R(i) = \sum_{j:(i,j) \in R} E_{ver}(i,j) / \sum_{j:(i,j) \in R} 1 \dots\dots\dots (3)$$

위의 식(2), 식(3)에서  $u$ 는 수평 주파수이며, 수평주파수 범위  $u_1, u_2$ 는 각각  $u_1 = 1, u_2 = 6$ 이고,  $R$ 은 2차후보문자영역을 뜻한다.

검증이 종료되면, 남은 문자영역은 문자가 존재할 확률이 매우 높다. 이러한 특성을 기반으로 해당 위치를 기억하는 유망문자영역맵을 생성한다. Zhong의 방법을 [그림 1]의 (d)까지 수행하고 남아있는 문자영역의 블록 내 존재하는 화소들의 좌표를 원본크기로 확대 혹은 축소하였을 때 대응되는 좌표에 대한 유망문자영역맵의 값을 2로 할당하고, 그 외 부분은 0으로 할당한다. 이후 팽창연산을 수행하여 주는데 그 결과는 [그림 2]의 (e)와 같다. 팽창연산을 수행하는 이유는 블록 단위로 영상을 나누기 때문에 실제 문자영역의 경계에서

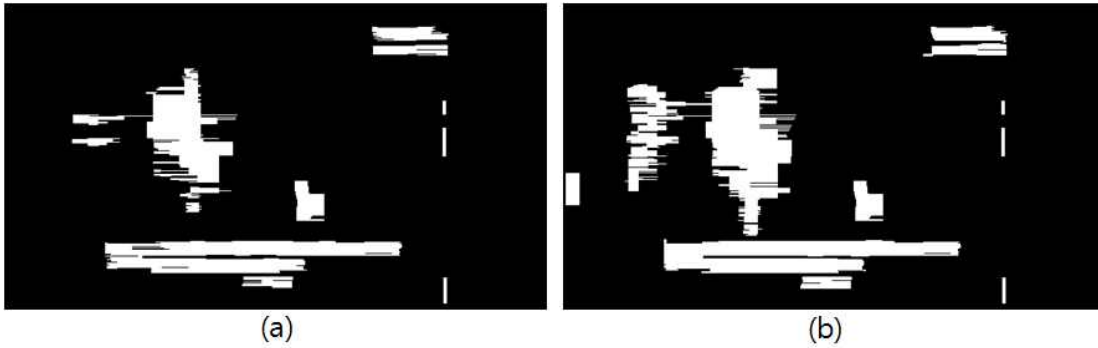


그림 4. 천이지도를 이용한 연결지도 생성결과. (a) 기존 천이지도를 이용한 연결지도 (b) 수정된 천이지도를 이용한 연결지도  
 Fig 4. The result of Generating the liked-map using transition maps. (a) the liked-map using traditional method. (b) the liked-map using modified transition map

문자영역이 극히 일부만 포함이 되어 앞선 단계들에서 비문자 영역으로 판정이 될 수 있으며, 실제 문자영역을 포함 하더라도 그 문자의 형태에 따라 고주파요소가 적어 검출이 안됐을 수도 있기 때문이다.

앞서 수행된 팽창연산은 앞선 단계들에서 검출이 안됐을 수도 있는 영역들을 얻은 것이므로 실제적으로 문자영역을 포함할 수도 있고 아닐 수도 있으며, 문자영역을 포함하더라도 극히 일부만을 포함하기 때문에 문자가 존재할 확률이 팽창연산을 수행하기 전보다 떨어진다. 그러므로 유망문자영역맵에 팽창으로 얻어진 블록들에 대응되는 화소들에 대해서는 1을 할당하여 준다.

### III. 문자영역 검출

Kim의 방법은 서론에서 언급한 것과 같이 정적인 임계값을 이용하여 천이 지도를 생성하며, 검증단계에서도 정적인 임계값을 적용한다. 하지만 자막은 종류, 배경과의 대비, 꾸밈 등이 매우 다양한데 정적인 임계값을 적용하는 것은 적절치 않다. 그렇기 때문에 경우에 따라 임계값을 조절할 필요가 있다.

#### 1. 천이지도 생성

천이지도는 영상에서 현재 화소에서 좌우화소들과 채도값[10] 변화와 명암값 변화가 큰 화소를 표시한 지도이다. 천이지도  $T$ 는 식(4)에 의해 정의 됐으며, 이하 값들은 식(5), 식(6), 식(7)에 의해 얻어진다[6-7].

$$T(x,y) = \begin{cases} 1, & \text{if } D_H > D_L + TH \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots (4)$$

$$D_L(x,y) = (1 + dS_L(x,y)) \times |I(x-1,y) - I(x,y)|, \dots\dots (5)$$

$$D_H(x,y) = (1 + dS_H(x,y)) \times |I(x,y) - I(x+1,y)|$$

where

$$dS_L(x,y) = |nS(x-1,y) - nS(x,y)|,$$

$$dS_H(x,y) = |nS(x,y) - nS(x+1,y)|$$

$$nS(x,y) = \frac{S(x,y)}{N} \dots\dots\dots (6)$$

where

$$N = \begin{cases} 2 \times (0.5 - nI(x,y)), & \text{if } nI(x,y) > 0.5 \\ 2 \times nI(x,y) & , \text{otherwise} \end{cases}$$

$$S(x,y) = 1 - \frac{3}{(R+G+B)} [\min(R, G, B)] \dots\dots\dots (7)$$

위의 식(4-7)에서  $x,y$ 는 좌표이며,  $S$ 는 [10]에서 정의된 채도값을 나타낸다.  $nI$ 는 명암값  $I$ 를 명암값의 최대값으로 나누어 0부터 1까지의 값으로 정규화한 값이며,  $nS$ 는  $nI$ 의 각 단계에 조건을 추가하여  $S$ 를 정규화한 값이다.  $dS_L, dS_H$ 는 현재화소와 좌측, 우측화소 간의  $nS$ 의 변동을 나타내는 값이며,  $D_L, D_H$  현재화소와 좌측, 우측화소 간의  $I$ 의 변동과  $nS$ 의 변동이 얼마나 있는지 나타내는 값으로서  $I, nS$ 의 변동과 정비례관계이다.

Kim방법의 단점은 임계값이 정적이란 것에 있다. 식(4)에서 천이지도를 생성할 때 사용되는 임계값  $TH$ 는 80으로 정의되어 있으며, 여러 가지 형태, 색상, 명암, 꾸밈이 있는 자

막들을 고려하였을 때 적절치 않다. 그렇기 때문에 본 논문에서는 앞서 구한 유망문자영역맵을 이용하여 TH를 식(8)과 같이 단계별로 조절하여 천이지도를 재정의한다. 여기서 PTR<sub>map</sub>은 앞서 얻어진 유망문자영역맵이며, D<sub>L</sub>과 D<sub>H</sub>를 구하는 방법은 Kim방법과 동일하다.

$$T(x,y) = \begin{cases} 1, & \text{if } D_H > D_L + TH(x,y) \\ 0, & \text{otherwise} \end{cases} \dots\dots\dots (8)$$

where

$$TH(x,y) = \begin{cases} 80, & \text{if } PTR_{map}(x,y) = 0 \\ 60, & \text{if } PTR_{map}(x,y) = 1 \\ 50, & \text{if } PTR_{map}(x,y) = 2 \end{cases}$$

[그림 3]은 Kim방법의 천이지도와 본 논문에서 수정된 천이지도를 나타낸 것이다. 본 논문에서 수정된 천이지도는 적용되는 임계값들이 유망문자영역맵에 따라 기존 천이지도에 적용되는 임계값보다 같거나 작기 때문에 DCT계수 분석을 이용한 문자영역으로 판별이 난 영역에 대해 좀 더 많은 후보 화소들이 검출된다.

검출된 화소들은 문자영역에 집중적으로 분포하기 때문에 가로방향 스미어링을 통해 주변화소들과 연결하여 문자후보영역을 생성한다[11]. 스미어링 시 연결되는 화소간의 최대 거리는 Kim의 방법과 마찬가지로 영상 너비의 5%이다.

[그림 4]는 각 천이지도를 이용한 연결지도의 그림이다. 본 논문에서 수정된 천이지도를 이용한 연결지도가 기존 천이지도를 이용하는 것보다 더 많은 연결된 화소들이 존재하는 것을 볼 수 있다.

**2. LBP를 이용한 검증**

이후 Kim 방법은 상태 밀도인 DOS 개념을 이용하여 문자후보영역을 검증한다. 구체적인 방법은 각 후보영역 내 각 후보화소에서 식(9)을 이용한 LBP를 산출한다. Kim은 문자영역은 영상압축 시 발생하는 정보손실로 색상 번짐현상이 일어나는 것과, 이 때문에 복잡한 구조의 삼입 문자의 경계 주변화소 패턴이 다양해진다는 것을 실험을 통해 보였으며, 이러한 특성을 LBP의 결과값이 다양하게 나온다는 특성을 검증식에 적용하였다.

$$LBP_{P,R} = \sum_{i=1}^{P-1} s(g_i - g_c)2^i, \text{ where } (x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases} \dots\dots\dots (9)$$

위의 식(9)에서 P는 현재 화소 주변의 원형으로 분포된 이웃 화소의 개수를 의미하고 R은 원의 반지름을 의미한다. g<sub>c</sub>는 현재 화소의 값이고, g<sub>i</sub>는 P에 포함된 화소의 값이다. Kim의 방법과 본 논문에서 적용되는 구체적인 매개변수값은, P는 8이며 주변 8이웃[10]을 뜻한다. 또한 현재화소와 주변화소가 서로 비교되는 값은 명암값이다.

$$DOS_i = w_i \frac{B_i(LBP_{8,1})}{256}, \text{ where } i = 1, \dots, N \dots\dots\dots (10)$$

식(10)은 최종 문자영역을 결정하는 검증식이다. N은 후보영역개수이며, w는 각 후보영역의 삼입 자막 후보 화소의 밀도다. 간단히 하면 후보화소를 후보문자영역의 최소인접사각형(MER)로 나눠준 값이다. 식(10)에서 상태밀도의 의미를 살펴보면, 후보문자영역 내에 후보화소가 다수 존재할수록, 그리고 LBP의 값 종류가 다양할수록 밀도값이 상승한다.

Kim의 방법은 검증단계에서 DOS의 임계값을 0.05로 정하고 임계값 이상의 값을 갖는 후보문자영역만을 최종문자영역으로 검출한다. 하지만 정적인 임계값에 때문에 경우에 따라 검증단계에서 실제 문자영역이 기각되는 경우가 많다.

이러한 문제점을 해결하기 위해 본 논문에서는 앞서 구한 유망문자영역맵을 이용하여 상태밀도 임계값을 단계별로 구한다. 본 논문에서는 상태밀도 DOS를 구하는 과정에서 w를 산출 할 때 이용되는 화소들은 연결지도 생성 전 본 논문에서 재정의된 천이지도에 존재하는 화소만을 이용한다.

상태밀도 임계값인 TH<sub>DOS</sub>의 단계별 산출과 최종문자영역 검증은 식(12)를 따르며, 상태밀도의 단계 TH<sub>DOS</sub><sup>FLAG</sup>는 식(11)에 의해 구해진다.

$$TH_{DOS}^{FLAG} = \left[ \frac{1}{n_i} \sum_{m(x,y) \in R_i^{MER}} PTR_{map}(x,y) \right] \dots\dots\dots (11)$$

여기서 괄호[ ]의 의미는 반올림이다. n은 i번째 후보영역의 MER크기이며, m은 i번째 후보영역 R의 MER 내 존재하는 화소를 뜻한다. 간단히 식(11)이 의미하는 뜻은 후보영역의 MER에 대응되는 위치의 유망문자영역맵의 값의 평균을 반올림한 것이다. 위의 과정을 수행하면 유망문자영역맵의 존재하는 0, 1, 2로 표시된 위치의 포함정도에 따라 TH<sub>DOS</sub><sup>FLAG</sup>가 0, 1, 2 단계로서 나온다.

표 1. 기존의 방법과 제안된 방법에 대한 실험 결과  
Table 1. The result of experiment for traditional and proposed method.

방법	전체영상 개수	전체자막 개수	정검출 개수	일부검출 개수	과검출 개수	오검출 개수	미검출 개수
Kim의 방법	206	351	127	56	1	168	167
본 논문에서 제안된 방법	206	351	273	22	10	14	47

$$TextRegion_i = \begin{cases} TRUE & , \text{ if } DOS_i \geq TH_{DOS} \\ FALSE & , \text{ otherwise} \end{cases}$$

..... (12)

$$where TH_{DOS} = \begin{cases} 0.025, & \text{ if } TH_{DOS}^{FLAG} = 2 \\ 0.030, & \text{ if } TH_{DOS}^{FLAG} = 1 \\ 0.050, & \text{ if } TH_{DOS}^{FLAG} = 0 \end{cases}$$

최종문자영역은 식(12)에서 상태밀도  $DOS$ 가 단계별 상태 밀도 임계값  $TH_{DOS}$ 보다 크거나 같은 영역이다.

#### IV. 실험 결과

본 논문에 실험은 크게 두 가지에 대해 실험한다. 여기서 첫 번째 실험은 Kim의 방법과 본 논문에서 제안하는 방법과

의 비교이고, 두 번째 실험은 본 논문에서 검증에 단계별로 적용되는 값을 다르게 하였을 경우이다. DCT 계수 분석을 이용한 문자영역 탐지와 비교를 하지 않은 이유는 Zhong는 문에 해당 방법에 대한 실험결과가 검출, 오검출, 미검출에 대해 언급 되어있으며, 결과 영역을 표현하는 방법이 Zhong은 블록단위이고 제안된 방법은 화소단위이기 때문에 적절치 않아 정량화하여 비교하지 않았다.

실험은 Visual Studio 2005를 사용 하였으며, 실험 영상은 동영상에서 206장의 정지영상을 수작업으로 얻었다. 실험에 사용된 영상의 해상도는 640 X 360 11장, 1280 X 720 20장, 640 X 480 48장, 624 X 352 98장, 800 X 600 14장, 720 X 406 7장, 704 X 396 8장이다. 정지영상을 얻은 동영상은 뉴스(SBS, MBC, KBS) 48장, 리얼 버라이어티(KBS, MBC) 54장, 일본쇼프로그램 21장, 음악방송(KBS) 32장, 기타 51장이다.

본 실험에서 사용된 집계방법은 다음과 같다. 만약 검출된 영역이 자막을 포함하고 있다면 검출이다. 만약 검출된 영역이 자막을 포함하지 않는다면 오검출이며, 실제 자막이 존재

표 2. Kim의 방법과 제안된 방법에 대한 검출률, 오검출률, 정검출률 비교  
Table 2. a comparison of detecting rate, false positive detecting reate, true positive detecting for Kim's method and proposed method.

방법 및 검출분류		검출률	오검출률	정검출률
Kim의 방법	일부검출을 정검출에 포함 과검출을 정검출에 포함	50.4%	49.8%	50.2%
	일부검출은 미검출 포함 과검출은 오검출에 포함	34.2%	59.5%	40.5%
본 논문에서 제안된 방법	일부검출을 정검출에 포함 과검출을 정검출에 포함	86.9%	4.4%	95.6%
	일부검출은 미검출 포함 과검출은 오검출에 포함	77.8%	8.1%	91.9%

하지만 검출되지 않았다면 미검출이다. 검출된 결과 분류에서, 실제 자막영역의 MER의 넓이를 80%이상 130%이하를 포함하는 MER을 문자영역으로 검출하였을 경우 정검출이라 하였다. 검출된 MER영역이 실제 자막영역 MER의 넓이에 80%미만을 검출하였다면 일부검출이라 하였으며, 실제 존재하는 자막의 MER넓이에 130%를 초과하여 MER을 찾은 경우 과검출이라 하였다. 실험영상 중 좌측상단에 방송국의 이름을 표시하는 자막이 나오는데, 이러한 자막들은 검출여부에서 제외하였다.

다음 [표 1]은 Kim의 방법과 제안된 방법에 대한 실험 결과이다. Kim의 방법이 검출률 수치가 낮은 이유는 후보영역 검출까지는 정상적으로 이뤄지지만, 검증단계에서 후보문자 영역들의 기각이 빈번히 일어났기 때문이다. 반면에 제안된 방법은 검증단계에서 임계값을 단계적으로 적용함에 따라 이러한 미검출을 줄일 수 있었다.

다음 [표 2]는 이러한 결과에 따른 검출률과 오검출률이다. 검출률은 (정검출 개수 / 실제 자막 개수)로 구했으며, 정검출률은 (정검출 개수 / 전체 검출 개수), 오검출률은 (오검출 개수 / 전체 검출 개수)로 구했다.

제안된 방법에 있어서 최초 천이지도를 생성할 때 임계값을 단계별로 할당한 80, 60, 50을 하향 조정하여 적용하였을 경우 검출률이 떨어지고 과검출률이 늘었다. 검출률이 떨어진 원인은 배경에서 후보화소들이 검출되어 후보문자영역이 원 배경에서 검출된 후보화소를 다수 포함하는 MER이 검출되어 MER크기는 커지고 MER 내부에 존재하는 후보화소의 개수가 상대적으로 감소하여 DOS를 구할 때  $w$  값이 급격히 작아짐에 따라 검증 단계에서 기각이 되었기 때문이며, 기각이

되지 않더라도 원배경을 다수 포함하는 과검출이 많았다. 반대로 위의 값을 상향 조절하였을 경우 미검출 및 일부검출이 늘었다.

또한 검증단계의 임계값을 0.5, 0.3, 0.25는 그 값의 조절에 따라 실험 결과가 달라졌다. [표 3]은  $TH_{DOS}^{FLAG}$ 에 따른  $TH_{DOS}$ 값을 조절하여 실험한 결과이다. [표 3]에서 보는 것과 같이 임계값을 낮추수록 검출률은 높아졌지만, 오검출률 또한 늘었다.

Zhong 방법의 결과는 블록단위로 찾은 것이기 때문에 본문문과 정확히 비교하기는 어려우나, Zhong 방법의 정검출 블록 140,983개와 오검출 블록 50,759개에 대한 오검출율(FPR)이 약 26.5%인 것을 고려하였을 때, 제안된 방법에서 검출률은 비록 낮지만 오검출률이 획기적으로 줄어든 것을 알 수 있다. 제안된 방법이 검출률은 Zhong의 방법보다 낮지만, Zhong의 방법이 일부검출하는 것에 대해 정검출이 가능한 경우도 [그림 5]에서 보는 것과 같이 존재한다.

Kim의 방법과 제안된 방법 모두 [그림 4]에서 보는 것과 같이 우측상단 자막을 포함하는 후보영역을 얻었으나 [그림 5]를 보면 Kim의 방법은 검증단계에서 해당 자막이 기각된 것을 알 수 있다.

[표 4]는 Kim방법과 제안된 방법의 결과영상 예이다. 최좌측영상은 원본영상이며, 각각의 방법에 의해 최종적으로 검출된 문자영역은 흰색 사각형으로 표시하였다. 결과영상에서 보는 것과 같이 제안된 방법은 Kim의 방법이 검출하지 못하는 문자영역까지 검출을 하며, 또한 오검출 영역이 적은 것을 알 수 있다.

표 3. 제안된 방법의 임계값 설정에 따른 검출률, 오검출률, 정검출률 비교  
Table 3. a comparison of detecting rate, false positive detecting rate, true positive detecting rate by adjusting the threshold in proposed method.

$TH_{DOS}^{FLAG}$ 에 따른 $TH_{DOS}$ 와 검출분류	검출률	오검출률	정검출률	
$TH_{DOS} = \begin{cases} 0.020, & \text{if } TH_{DOS}^{FLAG} = 2 \\ 0.025, & \text{if } TH_{DOS}^{FLAG} = 1 \\ 0.050, & \text{if } TH_{DOS}^{FLAG} = 0 \end{cases}$	일부검출을 정검출에 포함 과검출을 정검출에 포함	89.5%	9.2%	91.8%
	일부검출은 미검출 포함 과검출은 오검출에 포함	79.2%	14.7%	86.3%
$TH_{DOS} = \begin{cases} 0.015, & \text{if } TH_{DOS}^{FLAG} = 2 \\ 0.020, & \text{if } TH_{DOS}^{FLAG} = 1 \\ 0.050, & \text{if } TH_{DOS}^{FLAG} = 0 \end{cases}$	일부검출을 정검출에 포함 과검출을 정검출에 포함	92.2%	9.2%	91.8%
	일부검출은 미검출 포함 과검출은 오검출에 포함	81.7%	15.6%	84.4%



표 4. 제안된 방법과 Kim방법의 결과영상 비교  
Table 4. a comparison of result image for proposed method and Kim's method.

원본영상	제안된 방법의 결과영상	Kim방법의 결과영상

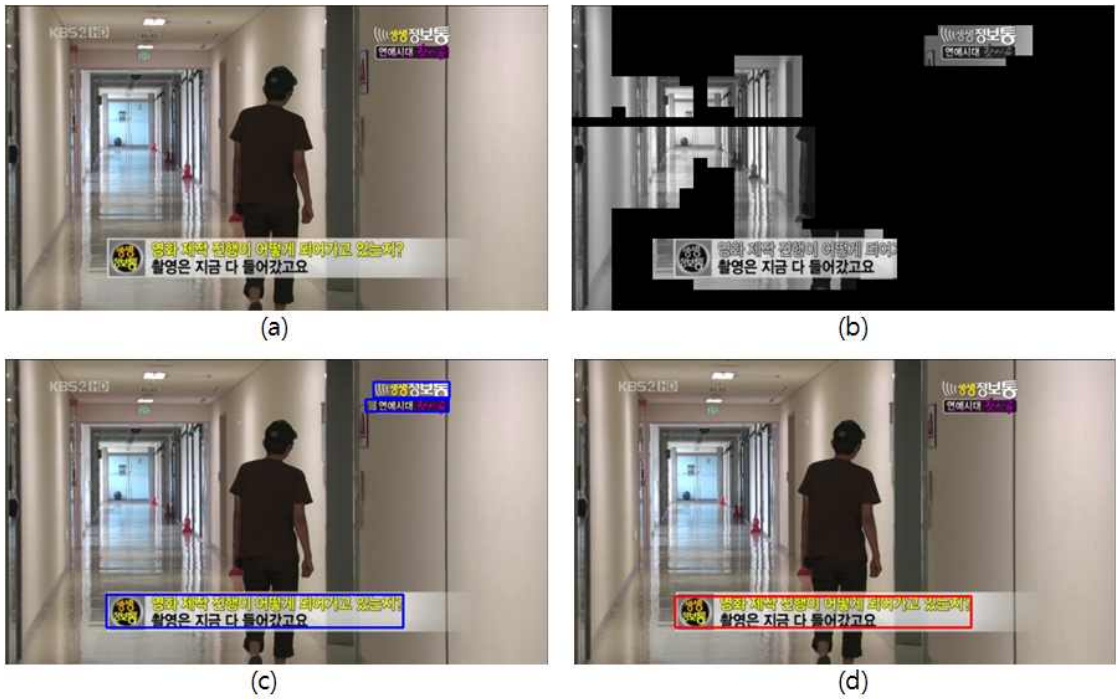


그림 5. 실험결과영상 비교 (a) 입력영상. (b) Zhong 방법의 결과영상. (c) 제안된 방법의 결과영상. (d) Kim 방법의 결과영상  
 Fig. 5. a comparison about result images. (a) input image. (b) Zhong's method result image. (c) proposed method result image. (d) Kim's method result image.

V. 결론

본 논문은 DCT계수 분석을 이용한 문자영역탐지방법에서 유망문자영역맵을 작성하고 친이지도 생성 및 후보문자영역 검증에 이를 활용함으로써 두 개의 문자영역탐지방법의 장점을 살리고 단점을 보완하는 방법을 제시하였다. 비록 DCT계수 분석을 이용한 문자영역탐지방법보다 검출률이 낮지만, 오검출률이 크게 낮았으며, 기존 친이지도를 이용한 문자영역탐지방법보다 검출률을 높이고, 오검출률을 낮췄다.

현재는 유망문자영역맵에서 각 위치의 값을 0, 1, 2 세 단계로 구분하고 있는데, 향후에는 이를 단계를 나누지 않고 DCT계수 분석 시 나오는 값을 이용한 실수의 값이 적용될 수 있도록 유망문자영역맵 작성에 대한 연구할 것이며, 또한 이렇게 개선된 유망문자영역맵을 친이지도에 적용함에 있어 임계값을 단계별로 정하는 것이 아니라, 어떠한 수식에 의해 정량적으로 계산되어 얻어지는 동적임계값을 만들고자 한다. 또한 추출된 자막영역을 개선하여 문자추출이 용이하도록 문자열 단위로 추출할 수 있는 방법과, 문자추출의 방법에 대해 연구할 것이다.

참고문헌

- [1] Palaiahnakote Shivakumara. Trung Quy Phan. Chew Lim Tan, "New Wavelet and Color Features for Text Detection in Video.", 20th International Conference on Pattern Recognition, 2010., vol.20 no.6 pp.3006-3009.
- [2] Xuening Qian. Guizhong Liu. Huan Wang. Rui Su, "Text detection, localization, and tracking in compressed video.", Signal processing, Image communication., 2007., vol.22 no.9 pp.752-768
- [3] Yu Zhong. HongJiang Zhang. AnilJain, "Automatic Caption Localization in Compressed Video.", IEEE transactions on pattern analysis and machine intelligence, 2000., vol.2 no.11 pp.385-392
- [4] Rainer Lienhart. Axel Wernicke, "Localizing and segmenting text in images, videos and web pages.", IEEE transactions on circuits and systems for video technology., 2002., vol.12, no.4, pp.256-268

- [5] Michael R.Lyu, Jiqiang Song, Min Cai., "A comprehensive method for multilingual video text detection, localization, and extraction", IEEE transactions on circuits and systems for video technology, a publication of the Circuits and Systems Society 2005, vol.15 no.2 pp.243-255
- [6] Wonjun Kim Changick Kim, "A new approach for overlay text detection from complex video scene.", Journal of Broadcast Engineering., 2008., vol.13 no.43 pp.544-553
- [7] Wonjun Kim Changick Kim, "A New Approach for Overlay Text Detection and Extraction From Complex Video Scene.", IEEE transactions on image processing, a publication of the IEEE Signal Processing 2009., vol.18 no.2 pp.401-411
- [8] Anil K. Jain. and Bin Yu., "Automatic text location in images and video frames.", Pattern Recognition, vol.31 no.12 pp.2055-2076, 1998
- [9] Rainer Lienhart, Frank Stuber., "Automatic text recognition in digital videos.", SPIE Internat. Soc. Opt. Eng., 1996, pp.180-188
- [10] Rafael C. Gonzalez. Richard E. Woods., "Digital Image Processing, 3rd Edition", PEARSON, 2008
- [11] Ah-hyun Cho. Hye-hyun Lee. Jae-uk Ryu. Kwang-Baek Kim., "The Extraction of Character from an English Name Card by Using Smearing Method and Contour Tracking Algorithm.", Proceedings of the Korea Intelligent Information System Society Conference 2002 May., 2002, pp.410-413

## 저 자 소개



### 안 권 재

2009: 숭실대학교  
컴퓨터학부 공학사.  
현 재: 숭실대학교  
컴퓨터학부 석사과정.  
관심분야 : 문자탐지, 문자인식,  
이진화  
Email : ankwonjae@naver.com



### 주 성 일

2008: 한국산업기술대학교  
컴퓨터공학과 공학사.  
현재: 숭실대학교  
미디어학과 박사과정.  
관심분야 : 영상처리, 패턴인식, 추적  
Email : sijoo82@ssu.ac.kr



### 김 계 영

1990: 숭실대학교 전산학과 공학사.  
1992: 숭실대학교  
컴퓨터학과 공학석사.  
1996: 숭실대학교  
컴퓨터학과 공학박사.  
2001년 ~ 현재 : 숭실대학교 컴퓨  
터학과 교수.  
관심분야 : 컴퓨터비전, 형태인식,  
생체인식 증강현실 등  
Email : gylim1@ssu.ac.kr



### 최 형 일

1979: 연세대학교 전자공학과 공학사.  
1984: 미시간대학 전산학과 공학석사.  
1987: 마시간대학  
전산학과 공학박사.  
1987년 ~ 1999년: 숭실대학교  
컴퓨터학부 교수.  
2000년 ~ 현재: 숭실대학교 미디어  
학부 교수.  
관심분야 : 컴퓨터비전, 패턴인식,  
증강현실 등  
Email : hic@ssu.ac.kr

