

# 에지와 색상 정보를 이용한 강의 영상의 학습 영역 추출

한은영\*, 서정희\*\*, 박홍복\*

\*부경대학교 전자컴퓨터정보통신공학부

\*\*동명대학교 컴퓨터공학과

e-mail:nowissue@korea.com

## Extraction of Study Regions from Lecture Video Using Edge and Color Information

Eun-Young Han\*, Jung-Hee Seo\*\*, Hung-Bog Park\*

\*Division of Electronic, Computer and Telecommunication  
Engineering, Pu-Kyong National University

\*\*Dept. of Computer Engineering, Tong-Myong University

### 요 약

본 논문에서는 강의 영상에 포함되어 있는 텍스트 정보를 바탕으로 학습에 의미 있는 영역을 추출한다. 실시간으로 입력되는 컬러 영상에서 Canny 연산자를 이용하여 에지 정보를 구하고, 모폴로지 연산(Morphological Operation)과 연결 성분(Connected Component)을 통해 후보 영역을 검색한다. 그리고 검색된 후보 영역내의 색 정보 분석을 통해서 의미 있는 학습 영역을 추출하였다. 본 논문에서 제안한 학습에 의미 있는 영역을 추출한 결과, 비교적 단순한 강의 영상과 복잡한 강의 영상 모두에서 정확한 학습 영역의 추출이 가능함을 알 수 있고, 학습에 의미 있는 정보만을 구성함으로써 적은 용량으로 최적화된 강의 영상을 제공할 수 있음을 확인할 수 있다.

### 1. 서론

최근 온라인 강의 및 원격 강의의 급격한 확산으로 인하여 언제 어디서나 학습하기 위한 많은 연구가 이루어지고 있다[1, 5]. 그러나 적은 용량과 배터리를 가진 모바일 디바이스(Mobile Device)에서는 고용량의 멀티미디어 정보를 저장할 수 없으며, 실행 시간에 있어서도 제약이 따르고 있어 원격 강의 서비스를 원활히 이용하기에는 아직 무리가 있다. 이러한 제한적인 디바이스 환경에서 짧은 시간에 효과적으로 학습을 하기 위해서는 적은 용량으로 학습에 필요한 최적화된 정보만을 구성할 수 있는 기술이 요구된다.

실시간으로 진행되는 강의 영상의 데이터 유형은 텍스트, 영상, 애니메이션 등 다양하다. 이러한 데이터 유형 중 텍스트는 이미지의 내용을 함축적이고 구체적으로 표현하는 중요한 정보이다. 따라서 본 논문에서는 강의 영상에서 학습에 의미 있는 영역을

정의하여 최적화된 강의 영상으로 구성하기 위해 에지 기반의 이미지 분할 기법을 이용하여 텍스트 영역을 추출하고, 추출된 텍스트 영역을 후보 영역으로 정의한 후 색 정보를 분석하여 추가적으로 학습에 필요한 객체를 포함함으로써 학습에 의미 있는 영역을 구성하고 추출하는 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 설명하고, 3장은 제안하는 알고리즘의 세부 사항을 설명한다. 4장은 구현 및 결과를 분석, 5장은 결론 순으로 기술한다.

### 2. 관련 연구

특정한 제약 없이 자연스럽게 촬영되어 영상에 내포된 텍스트들을 장면 텍스트[6]라고 한다. 장면 텍스트는 비구조적이고 다양한 형태의 텍스트 정보와 복잡한 배경을 포함할 경우가 많아 정확한 텍스트 영역을 추출하기가 쉽지 않다. 특히 교육 영상은 텍

스트 이외의 학습 중에 추가적으로 첨가되는 그림, 표, 그래프, 필기구, 포인터 등의 객체들로 인해 학습 영역만을 명확히 분류해 내기에는 한계가 있다.

2.1 장면 텍스트 추출

장면 텍스트 추출에 관한 연구는 크게 명도 정보를 이용한 텍스트 추출 연구[1, 2]와 색 정보[3, 4]를 이용한 텍스트 추출 연구로 분류할 수 있다.

명도 정보를 이용한 텍스트 추출에서 [2]는 텍스트는 주변 영상과 명도차가 크고 일정하다는 특징을 이용하여 입력 영상에서 지역적 임계값을 적용하여 텍스트 영역을 검출하고, 유사성을 비교한 후 요소들을 병합하는 방법을 제안하였다. 이 결과 배경이 복잡하지 않은 단순한 이미지에서는 좋은 결과를 얻었으나, 텍스트 영역의 기울어짐이 없고, 배경과 텍스트 영역이 정확히 구분되어야 하는 제약을 가지고 있다.

색상은 영상에서 객체를 식별하게 하는 중요한 요소이다. 그러나 색상 정보만을 이용하게 될 경우 텍스트 영역과 배경색이 유사하거나, 조명 상태가 변하는 경우 효과적인 영역 추출이 어려워진다. 색 정보를 이용한 방법으로는 색 양자화(Quantization)를 수행하고 각 색 면에 대한 연결 요소를 분석하여 텍스트를 추출한 방법[3], 색 병합 후 모폴로지를 적용하여 텍스트를 추출하는 방법[4] 등이 있다. 최근의 경향은 어느 하나의 정보만을 이용하기 보다는 복합적으로 사용하여 영역을 검출한다. 본 논문에서는 이러한 명도 정보와 색 정보의 장점을 통합적으로 이용하여 학습 영역을 추출하는 방법을 제안한다.

2.2 모폴로지(Morphology) 연산

에지 추출 후 이진화된 결과 영상은 그림자나 작은 구멍과 같은 잡음(Noise)을 포함하게 된다. 따라서 이러한 작은 구멍들을 채우고 돌출 부분을 제거하여, 불연속적인 에지 점들을 연결하기 위해 이진화 과정을 거친 후 영상에 모폴로지 기법인 팽창(Dilation)과 침식(Erosion)을 반복적으로 수행한다 [7]. 식(1)과 식(2)는 각각 팽창 연산과 침식 연산으로 물체의 최외각 픽셀을 확장 또는 수축시키는 역할을 한다. 식(3)은 Opening 연산으로 침식 연산 후 팽창 연산을 하는 것으로 이와 같은 제거 연산을 통해서 일반적으로 물체의 윤곽을 부드럽게 만들어 가고, 가는 돌출부나 협곡 같은 부분은 없애준다.

$$A \oplus B = \{z | (B)_z \cap A \neq \Phi\} \quad A : \text{Binary Image} \quad \text{식(1)}$$

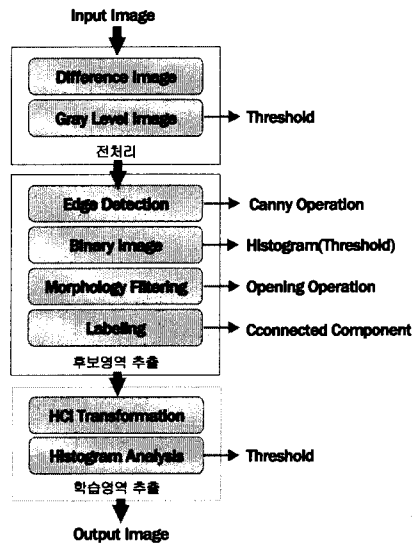
$$A \otimes B = \{z | (B)_z \subseteq A\} \quad B : \text{Structuring element(Mask)} \quad \text{식(2)}$$

$$A \circ B = (A \otimes B) \oplus B \quad (B)_z : B \text{ translated by } z \quad \text{식(3)}$$

3. 설계

3.1 학습 영역 추출 단계

본 논문에서 학습 영역(Region of Study)을 추출하기 위한 과정은 그림 1에서 볼 수 있듯이 전처리 단계, 후보 영역 추출 단계, 학습 영역 명확화 단계로 나뉜다.



(그림 1) 학습 영역 추출 단계

학습 영역을 추출하기 위한 제안한 단계는 다음과 같은 과정을 거친다.

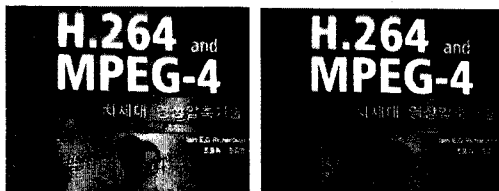
- Step1. 학습 영역내의 변화가 발생할 경우 차영상을 추출한다(단, 학습 영역의 초기값은 320x240이며, 이전 프레임에서 추출된 학습 영역이 있을 경우 갱신된다).
- Step2. 실시간 처리의 계산량을 줄이기 위해 컬러 영상을 그레이 영상으로 변환한다.
- Step3. 그레이 영상에서 Canny 에지 연산을 사용하여 에지 정보를 구한 후 히스토그램을 통해 임계값(Threshold)을 적용시켜 이진화 정보를 구한다.
- Step4. 이진 에지 영역에 형태학적(Morphological) 필터링 연산인 Opening 연산을 수행하여 명확한 에지만을 남기고 불필요한 잡음을 제거한다.
- Step5. Connected Component 라벨링(Labeling) 알고리즘을 통해 학습 후보 영역을 결정한다.

- Step6. 학습 후보 영역에 한해서 밝기 값과 HCI 컬러성분을 이용하여 학습 개체를 찾는다.
- Step7. Histogram 분석을 통해 학습 후보 개체에 대해서 위치 및 거리 정보가 학습 영역의 특징과 부합하는지를 비교하여 적합한 개체를 검출한다.
- Step8. 학습 개체가 검출되면 그 영역을 학습 영역으로 최종 확정하고, step1에서는 다시 갱신된 학습 영역을 비교하게 된다.

### 3.2 차영상 추출 및 전처리

전처리 단계는 실시간 영상에서 비디오 정보를 요약하기 위한 선행 작업으로 그림 2와 같이 단계적으로 변화되는 학습 영역내의 변화가 어느 임계값 이상이 될 경우 새로운 학습 내용임을 감지하고 추출 (a)한 후 그레이 영상으로 변환(b)하는 단계이다. 이 단계는 학습 영상의 특징이 고려된 부분으로 영상의 변화가 학습 영역 위주가 아니라면 중요도가 낮다는 전제를 두었다. 그레이 영상으로 변환하는 식은 (식 3.1)과 같으며, 그림 2는 전처리 단계를 통해 추출된 차영상(a)과 그레이 영상(b)을 나타낸다.

$$\begin{aligned}
 \text{BYTE* pData} &= (\text{BYTE*})\_lp\text{LastVHdr} \rightarrow \\
 \text{lpData} &+ (\text{height} - 1 - (y - 1)) * \text{width} * 3 + (x - 1) * 3; \\
 \text{GrayImg} &= (\text{BYTE}) (\text{pData}[2] * 0.299) \\
 &+ (\text{pData}[1] * 0.587) \\
 &+ (\text{pData}[0] * 0.114)
 \end{aligned}
 \tag{식 3.1}$$



(a) 차영상 추출 (b) 그레이 변환  
(그림 2) 전처리 단계

### 3.3 학습 후보 영역 추출

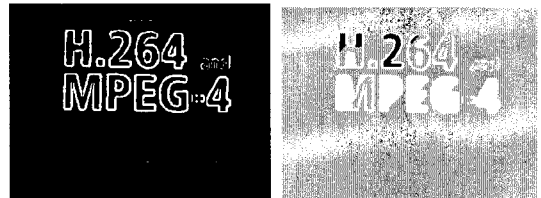
강의 영상의 관심 영역은 텍스트가 1차적이다. 따라서 그레이 영상 변환 후 그림 3과 같이 Canny 연산(a)을 통해 에지를 추출하고 이진 영상으로 변환한 후 명확한 에지만을 남기고 불필요한 잡음을 제거하기 위해 Projection 연산(b)과 오프닝(Opening) 연산(c)을 각각 수행한다. 그리고 라벨링(Labeling) 과정(d)을 통해 학습 후보 영역을 정의한다. 에지 추출 후 모폴로지의 Opening 과정의 침식(식 3.2)과 팽창(식 3.3)에 대한 연산 식은 아래와 같다.

$$\begin{aligned}
 \text{if } (\text{img\_data}[i][j-1] == 1 \ \&\& \ \text{img\_data}[i-1][j] \\
 == 1 \ \&\& \ \text{img\_data}[i][j] == 1) \\
 \text{s\_tmp}[i][j] &= (\text{unsigned char})1;
 \end{aligned}
 \tag{식 3.2}$$

$$\begin{aligned}
 \text{if } (\text{img\_data}[i+1][j]==1 \ \parallel \ \text{img\_data}[i][j+1]==1 \ \parallel \\
 \text{img\_data}[i+1][j+1]==1 \ \parallel \ \text{img\_data}[i][j]==1) \\
 \text{s\_tmp}[i][j] &= (\text{unsigned char})1;
 \end{aligned}
 \tag{식 3.3}$$



(a) Canny Operation (b) Opening

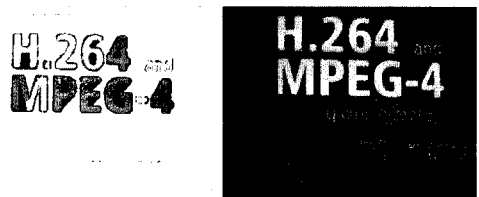


(c) projection (d) Labeling

(그림 3) 학습 후보 영역 추출 과정

### 3.4 색상 정보를 이용한 학습 영역 명확화

라벨링된 영역 중 실제 학습 영역임을 판단하기 위해 최소 학습 영역 비율과 에지 색상과 배경색의 분석을 통해 정확한 영역을 추출해 내는 단계이다. 따라서 그림 4는 최종 학습 영역을 추출한 결과를 나타내며, (a)는 색상 분석 후 추출된 요소이며, (b)는 텍스트의 위치 정보와 추출된 요소를 결합하여 최종 학습 영역을 Bounding Box 처리한 결과이다.



(a) 색상 영역 추출 (b) 학습 영역 지정

(그림 4) 결과 이미지

## 4. 구현 및 결과분석

본 논문에서 제안된 알고리즘의 성능을 평가하기 위해서 저해상도 캡을 이용하여 3분 가량의 강의 영

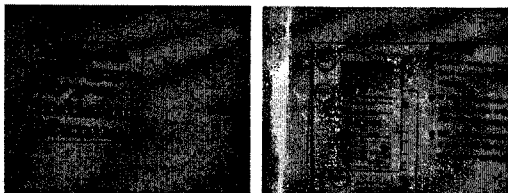
상을 촬영하여 AVI 형태로 저장하였다. 개발 환경은 WindowsXP를 기반으로, Microsoft Visual C++6.0의 MFC를 사용하여 개발하였다.

실험 영상은 강의 영상의 특징을 고려하여 텍스트 색과 배경 색은 다양하지만 복잡하지 않고, 학습 개체의 배치가 불규칙적으로 구성되어 프리젠테이션되는 영상을 이용하였다. 제안된 알고리즘의 성능 평가 결과는 표 1과 같다.

표 1. 성능 분석

	실 영상	제안된 영상	최적화
초당 프레임 수	90 frame	22 frame	25%
추출된 학습 영역 정확도	100%	98%	-
요약 후 용량	34.4MB (225kb x 90)	5.72MB (223kb x 22)	최적화 84.5%
실행 시간	90 sec	22 sec	

320x240 크기의 실시간 영상은 한 프레임 당 225 kb 크기를 가지며 3분가량의 동영상으로 촬영될 경우 34.4MB의 크기로 만들어진다. 즉 한 프레임 당 동영상으로 구성되기 위해 추가되는 크기는 37kb이다. 제안된 영상에서는 총 22 프레임이 추출되었으며 초당 25%의 프레임 감소율을 가져왔고 추출된 프레임으로 동영상을 구성하였을 경우 5.72MB의 크기로 84.5%의 크기 감소를 가져왔다. 그러나 실제로 음성이 포함된 강의 영상이 구성될 경우 용량은 더 클 것으로 예상되나, 압축이 되지 않은 AVI 유형의 파일이라고 본다면 코덱을 적용할 경우 압축율이 높을 것으로 예상된다.



(a)

(b)



(c)

(d)

(그림 5) 검증 영상

## 5. 결론

본 논문은 강의 영상에 포함되어 있는 텍스트 정보를 바탕으로 학습에 의미있는 영역을 추출하기 위해 에지와 색상 텍스트 검출 방법을 상호 보완하여 추출하는 방법을 제안하였다. 제안된 알고리즘의 시험결과 텍스트 영역만을 추출하는 부분은 그림 5의 (a)(b)와 같이 명확히 추출 되었는데 색 정보를 이용하여 배경이 복잡한 영역에서 학습에 의미 있는 개체를 포함시키는 부분은 그림 5의 (c)(d)와 같이 잘못 인식하는 경우가 발생했다.

제안된 알고리즘은 강의 영상만을 이용하여 프레임을 추출하였는데, 보다 효율적인 알고리즘으로 구성하기 위해서는 강사의 음성과 강의 영상을 동기화하여 제공되어야 하지 않을까하며, 모바일 디바이스에서 효율적으로 학습하기 위한 연구가 계속되어야 할 것이다.

## 참고문헌

- [1] C.Ngo, F. Wang & T. Pong "Structuring Lecture Videos for Distance Learning Applications," Proceedings of the IEEE Fifth International Symposium on Multimedia Software Engineering, 2003
- [2] X. Hua, X. Chert, L. Wenyin, H. Zhang "Automatic Location of Text in Video Frames," in Multimedia Information Retrieval Ottawa Canada of ACM, 2001
- [3] L.Gu, T.Kaneko, n.Tanaka, and R.M.Haralick, "Robust extraction of characters from color scene image using mathematical morphology", in Proceedings of Internantaional Conference on Pattern Recognition, vol.2, pp. 1002-1004, 1998
- [4] Y.Zhong, K.Karu, and A.K.Jain, "Locating text in complex color images," in Pattern Recognition vol.28, pp. 1523-1535, 1995
- [5] 임영진, 서정희, 박홍복, "모바일 단말에서의 SMIL을 이용한 멀티미디어 교육 알고리즘 설계 및 구현", 한국해양정보통신공학회, 2006, 5
- [6] 강오형, 황대훈, 이양원, "다중 프레임 병합을 이용한 스포츠 비디오 자막 영역 추출", 한국멀티미디어학회지, 2004, 4
- [7] Rafael C. Gonzales, Richard E. Woods, Digital Image Processing, Addison-Wesley, 1998