

의사결정 트리를 이용한 한글 자막 추출

정제희^o 이승훈 김재광 이지형
성균관대학교 정보통신공학부

E-mail: gulingi@skku.edu, reinblame@skku.edu, linux@ece.skku.ac.kr, jhlee@ece.skku.ac.kr

Korean Caption Extraction with Decision Tree

Je-Hee Jung^o · Seunhoon Lee · Jaekwang Kim · Jee-Hyong Lee
School of Information and communication engineering

요 약

자막은 영상과 관련이 있는 정보를 포함한다. 이러한 영상의 정보를 이용하기 위해서 자막을 추출하는 연구가 진행되고 있다. 기존의 자막 추출 연구는 언어 독립적인 특징으로 자막을 이루는 획의 예지는 일정한 간격을 유지하거나 수평라인으로 존재하는 글자의 분포를 이용한 방법을 제안하였다. 이러한 방법들은 획의 간격이 일정한 자막이나 하나의 글자가 하나의 획으로 이루어진 글자에서만 정상적인 동작을 보장하였다. 본 논문에서는 한글 자막 특징을 고려한 자막 추출 방법을 제안한다. 먼저, 한글 자막의 특징인 가로 획의 다수 분포를 고려한 적응형 예지 이진화를 수행하여 예지 영상을 생성하고 예지 연결 객체를 생성한다. 그 후에 생성한 연결 객체를 특징을 추출하여 사전에 생성한 의사결정 트리로 연결 객체를 자막과 비자막 연결객체로 분류한다. 의사결정 트리를 생성하기 위해서 사용한 연결 객체는 뉴스, 다큐멘터리 프로그램에서 획득하였으며, 성능 평가를 위해서 뉴스, 다큐멘터리, 스포츠 프로그램과 같은 대중 방송에서 획득한 영상에서 자막을 추출하였다. 평가 방법은 찾아진 연결 객체 중에 자막 연결 객체의 비율과 전체 자막 중에서 찾아진 자막 연결 객체의 비율로 분석하였다. 실험 결과에서는 제안한 방법이 한글 자막의 추출에 적용 가능함을 보여준다.

1. 서 론

지금 이 순간에도 정보를 포함하는 콘텐츠는 생성 저장, 검색 및 교환되고 있다. 이러한 콘텐츠는 문서, 음악, 영상, 동영상 등과 같이 다양하게 존재한다. 영상은 직관적으로 정보 전달이 가능하지만 다른 콘텐츠에 비해서 내용기반의 검색이 어렵다. 영상에 존재하는 자막은 영상의 내용과 밀접하게 존재한다.

영상에 존재하는 자막은 표준이 존재하지 않아서 다양하게 존재한다. 이러한 특징에 의해서 영상에 포함된 다양한 형태의 자막을 추출하기 위해 자막의 특징을 이용한 자막 추출 연구가 진행되어 왔다. 특히, 다양한 크기의 자막을 추출하기 위해서 기존의 연구에서는 영상에서 자막에 포함된 예지를 추출하기 위해 해상도를 변경하여 동일한 과정을 수행하거나[1], 각 예지마다 대응 예지와 획 간격을 조사하여 자막 예지를 추출하는 방법을 제안하였다[2]. 하지만, 이러한 방법들은 특정 언어의 특징만을 고려하였고, 한글이나 일정한 획의 간격을 만족하지 못하는 자막에서는 정상적인 성능을 보장하지 못한다.

본 논문에서는 한글 자막의 특징을 고려한 자막 추출 방법을 제안한다. 먼저, 자막 예지의 이진화를 상하방향의 예지의 분포를 이용하여 수평 라인을 경계로 수행한

다. 추출한 예지들로 연결 객체를 구성하고 구성된 연결 객체는 사전에 입력된 연결 객체의 패턴으로 생성된 의사결정 트리에 의해서 자막과 비자막 연결 객체로 분류한다. 영상에 포함된 자막 연결 객체를 포함하는 최소의 자막 영역을 다중 프로젝션 방법[1]을 사용하여 결정하고, 결정한 영역 내부에서 획 기반 이진화 알고리즘[2]을 사용하여 자막 연결 객체만이 남겨진 영상을 생성한다.

논문의 이어질 내용은 다음과 같다. 먼저, 2장에서 관련 연구에 대해서 분석한다. 다음으로 3장에서는 제안하는 방법을 서술한다. 4장에서는 제안한 방법의 성능 평가하기 위한 실험 결과를 분석하고 5장에서 본 논문의 결론을 서술한다.

2. 관련 연구

자막은 영상의 정보를 포함하고 있다. 영상에서 자막을 추출하기 위해서 자막의 특징을 고려한 자막 추출 연구가 진행되었다. 자막의 특징으로는 자막의 높이 정렬 방향, 글자간의 가로 간격 색상, 움직임, 경계등이 존재한다[3]. 또한, [1]에서는 자막은 언어 독립적인 특징과 종속적 특징으로 분류하였다. 언어 독립적인 특징은 색상의 대비, 색상의 밝기, 글자 정렬 방향, 자막 유지성으

로 언급하였다. 반면에, 언어 종속적인 특징은 획의 밀도, 글자의 높이, 글자의 가로 세로 비율 획의 분포로 언급하였다.

기존의 자막 추출 연구들은 언어 독립적인 특징들을 고려하여 진행되었다. 자막 추출 방법들은 성능의 향상을 위해서 자막의 후보 영역을 찾고 찾아진 영역에서 글자 부분만을 분리해 내는 단계적으로 후보를 점차 축소하는 방법을 사용하였다

기존의 방법들은 중에서 [4]는 자막의 특징 중에서 색상의 대비와 글자의 정렬 방향을 고려한 지형학적인 특징 점 추출 방법과 점-선-면 확장법을 이용하여 문자 영역을 추출하였다. 이 논문은 자막 영역 추출율과 시간을 실험 결과로 제공하였으며 자막 영역 추출 정확도는 언급하지 않았다.

[5][6]에서는 자막을 추출하기 위해 라인 단위의 에지를 추출하거나 각 색상 영역으로 분리한 영상에 신경망 학습을 통하여 자막 영역을 추출하여 하나의 영상으로 병합하는 방법으로 자막을 이루는 에지를 추출하는 방법들을 제안하였다. 이러한 방법들은 검색 영역이 고정적 형태로 존재하기 때문에 고정적인 형태에 의해 영상의 크기 변화에 정상적인 동작을 보장하지 못한다

[7-9]에서는 모폴로지 방법을 기반으로 자막을 추출하는 방법들을 제안하였다. 추출한 자막의 영역 중에서 확실하게 자막이 존재하지 않는 영역을 제거하여 자막 영역의 추출 정확도를 향상시켰다. 모폴로지 방법들은 자막 영역을 추출하기 위한 블록의 크기에 따라 성능이 변동되는 단점을 보인다. 특히, 영상에 포함된 자막의 크기가 달라질 경우에는 정상적인 동작을 보장하지 못한다.

전술한 대부분의 방법들은 언어 독립적인 특징을 고려하였고, 언어 종속적인 특징들을 고려하지 않는 영어 자막 추출 연구를 진행하였다. 본 논문에서는 한국어에 특징을 고려한 자막 추출 방법을 제안하고자 한다

본 논문에서는 배경보다 밝은 한글 자막을 다양한 크기나 형태를 가지더라도 추출이 가능하도록 의사결정 트리를 이용한 연결 객체 기반의 자막 추출 방법을 제안한다. 입력으로는 한글 자막을 포함한 영상이 사용되며 출력으로는 한글 자막부분만이 남겨진 이진화 영상이 생성된다.

사각형은 주변보다 밝은 색상을 갖는 자막 픽셀을 표현한다. 약간 어두운 색상의 사각형은 자막 픽셀이 아닌 배경 픽셀을 표현한다 또한, 화살표는 각 자막 경계 픽셀이 지니는 에지의 방향성을 표현한다 자막과 배경의 경계 픽셀은 서로 반대로 되는 방향의 에지 픽셀이 존재한다. 보기 편하도록 대각 방향 성분은 표현하지 않았다

특히, 한국어 자막에서는 그림 2와 같은 특징이 자주 존재한다. 그림 2는 한글 자막에서 우측으로 연속된 위쪽 방향 에지를 표현한 그림이다 그림 2에서는 검은 사각형은 배경 픽셀을 의미한다 회색 사각형은 상방향의 에지를 포함하는 자막 픽셀을 의미한다 또한, 각 사각형 안에 등장하는 숫자들은 현재 픽셀에서 왼쪽으로 등장하는 위쪽 방향의 에지와 의 거리를 나타낸다 그림 2에서 등장하는 '1'의 우측 획에서 아래쪽에 나타나는 '0111'은 자막의 밝기가 일정하지 않기 때문에 나타나는 현상이다. 각 회색 사각형들은 가로로 연속적으로 존재하며 자막 연결 객체에서는 상단에 밀집하여 존재하는 것을 확인할 수 있었다. 이러한 특징은 한글은 가로획과 세로획이 글자의 주를 이루기 때문에 나타나는 특징이라고 할 수 있다.

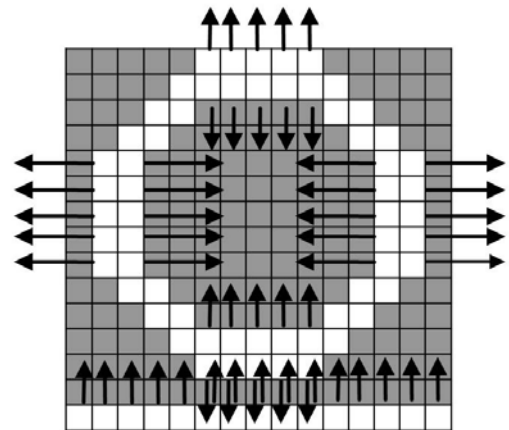


그림 1. 글자의 특징

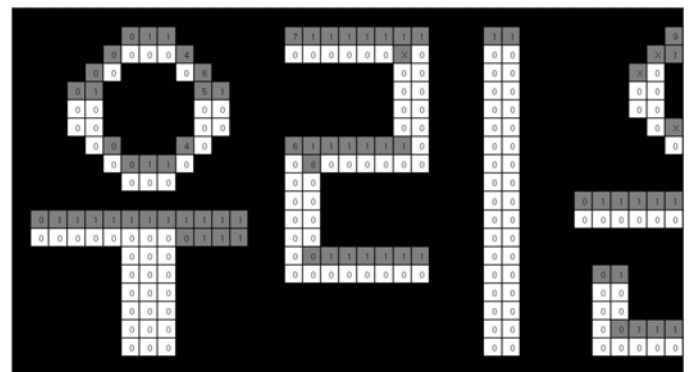


그림 2. 한글의 특징

3. 의사결정 트리를 이용한 연결 객체 기반의 한글 자막 추출 알고리즘

이 장에서는 추출 대상인 한국어 자막의 특징을 언급하고, 그러한 특징을 기반으로 하는 연결 객체 기반의 한국어 자막 추출 알고리즘을 서술한다

3.1. 한글 자막의 특징

대부분의 언어의 자막을 이루는 글자나 숫자들은 다음과 같은 특징을 갖는다. 글자를 이루는 글자들과 배경의 경계에 포함되는 픽셀들은 다음과 같은 특징을 갖는다 그림 1은 한글의 "우"를 확대한 것이다. 그림 1에서 밝은

3.2. 자막 추출 알고리즘

제안하는 알고리즘은 크게 3가지 단계로 분류가 가능하다. 먼저, 자막을 이루는 에지 정보를 추출하고 영상

에서 존재하는 자막과 유사한 연결 객체를 찾아내는 자막 찾기 단계와 찾아진 자막을 포함하는 최소 사각이 존재하는 위치를 결정하는 자막 위치 찾기 단계 찾아진 위치에서 자막과 비자막의 픽셀을 분리하여 자막을 분리하는 자막 분리 단계로 구성된다

3.2.1. 자막 찾기 단계

자막 찾기 단계에서는 자막을 이를 가능성이 있는 연결 객체를 추출한다. 입력 영상으로는 색상 영상을 사용하고, 결과 영상은 연결 객체만이 존재하는 연결 객체 영상을 생성한다.

먼저 입력으로 들어온 영상을 연산 량의 감소를 위해 보편적으로 사용되는 명암도 영상으로 변환한다 변환은 아래의 수식(1)을 사용하였다.

$$G(x,y) = 0.3 * C_R(x,y) + 0.59 * C_G(x,y) + 0.11 * C_B(x,y) \quad (1)$$

수식 (1)에서 G는 명암도 영상을 의미하며 각 C는 색상 영상을 의미한다. R, G, B는 각 색상 영역을 의미하며, 차례대로 붉은색, 녹색, 파란색을 의미한다. x와 y는 각 영상에서 존재하는 픽셀들의 좌표 값을 의미한다

이렇게 생성된 G에는 잡음 성분이나 영상의 압축에 의한 번짐 현상이 존재한다 이러한 성분은 에지 기반의 방법에서는 영향력이 크므로 이러한 성분을 제거하기 위해 보편적으로 사용되는 산술 평균 필터를 사용한다 산술 평균 필터는 대상 점과 주변 8개의 점의 밝기 값의 총 합을 9로 나누어 대상 점의 값을 대체하는 방법이다 이렇게 생성된 영상은 잡음 성분이 많이 감소하게 된다

잡음 성분이 많이 제거된 필터링된 영상(G_F)은 자막의 연결 객체를 추출하기 위해 에지를 추출하는 작업을 수행한다. 영상에 포함된 에지는 자막의 크기가 작으면 에지의 강도가 약해진다. 특히, 한글에 경우에는 글자의 획이 많이 포함되어 있기 때문에 에지의 강도가 더욱 약해진다. 제안하는 방법에서는 라인 기반 동적 이진화 방법을 제안하여 에지를 추출하였다

라인 기반 동적 이진화 방법은 작은 자막에 포함된 에지를 추출하기 위해서 낮은 임계치를 적용하고 큰 글자의 경우엔 높은 임계치를 적용하여 자막의 에지 정보는 최대한 유지하면서 자막이 아닌 비자막의 에지를 제거하는 방법이다. 이 방법은 한글 자막의 상 하방향 에지 픽셀들은 상방향, 하방향으로 밀집되어 연속으로 존재하는 특징을 고려하였다.

알고리즘은 다음과 같다. 먼저 필터링된 영상에 12시 방향으로 방향성을 지닌 에지와 6시 방향으로 방향성을 지닌 에지를 추출한다. 각 12시 6시 에지 영상은 각 수평 라인마다 그림 2에서와 같이 이전 에지 픽셀로부터 현재 에지 픽셀까지의 간격을 조사한다 이때, 조사된 에지 픽셀의 간격이 연속이면 1, 아니면 그 이외로 구분한다. 전체 라인에서 1을 지닌 에지의 비율이 일정 수준이상이라면 그 수평 라인은 이진화를 위한 기준 라인으로 선택된다. 본 논문에서는 0.85으로 정하였다. 선택된 기

준라인은 각 기준 라인마다 간격을 구하며 간격 간에 존재하는 에지 픽셀은 (2)에 의해 라인 단위 동적 이진화의 임계치 값이 결정된다

$$adaptive_{thres} = min_{thres} + LW_{thres} * (y_n - y_p) \quad (2)$$

min_{thres} 은 최소 이진화 값이다. LW_{thres} 은 각 라인의 간격에 곱해지는 가중치 상수 값이다 y_n, y_p 은 현재 검색중인 기준 수평 라인과 이전에 검색된 기준 수평 라인의 y 좌표 값이다. 실험에서는 최소 이진화 값은 10, 라인 간격 상수값은 0.5로 결정하였다. 연구 도중 조사에 의하면 최소 작은 자막은 10 이상의 높이와 최소 15정도의 에지 크기를 지니고 있었기 때문에 모든 경우를 추출하기 위해서 최소값을 15로 지정하기 위해 정한 수치이다

또한, 동시에 필터링된 영상에서 수식 (3)을 사용하여 각 방향의 에지 픽셀의 에지 크기를 추출한다

$$\begin{aligned} E_{tl}(x,y) &= G_F(x,y) - G_F(x-1, y+1) \\ E_t(x,y) &= G_F(x,y) - G_F(x, y+1) \\ E_{tr}(x,y) &= G_F(x,y) - G_F(x+1, y+1) \\ E_l(x,y) &= G_F(x,y) - G_F(x-1, y) \\ E_r(x,y) &= G_F(x,y) - G_F(x+1, y) \\ E_{bl}(x,y) &= G_F(x,y) - G_F(x-1, y-1) \\ E_b(x,y) &= G_F(x,y) - G_F(x, y-1) \\ E_{br}(x,y) &= G_F(x,y) - G_F(x+1, y-1) \end{aligned} \quad (3)$$

각 E는 위에서부터 차례대로 좌상 상, 우상, 좌, 우, 좌하, 하, 우하를 의미한다. x, y는 각 픽셀의 위치를 의미한다.

추출한 각 에지 영상들은 각 라인마다 동적 이진화 임계값으로 이진화 영상으로 대체된다 그 후에 추출된 각 8 방향의 에지 성분들이 하나라도 좌표에 존재한다면 모두 하나의 영상에 표현하는 하나의 전 방향 에지 영상(E_{all})을 생성한다.

이렇게 생성된 전 방향 에지 영상에 이웃 연결 객체 방법을 사용하여 연결 객체 영상 CC_{first} 를 생성한다. 8-이웃 연결 객체 방법 보다는 4-이웃 연결 객체 방법이 연결 객체가 불필요하게 많이 접촉되는 경우를 방지하기 때문에 사용하였다. 생성된 연결 객체는 수식(4)를 만족하며, 자막을 표현하는 연결객체의 후보라 할 수 있다

$$CC_{first}(x,y) = \begin{cases} E_{all}(x,y) \neq 0 \wedge \\ CC_{first}(x,y-1) \neq 0, & CC_{first}(x,y-1) \\ E_{all}(x,y) \neq 0 \wedge \\ CC_{first}(x-1,y) \neq 0, & CC_{first}(x-1,y) \\ else & 0 \end{cases} \quad (4)$$

식 (4)에서 E_{all} 은 전 방향 에지 영상을 의미하며 x,y는 영상의 각 좌표를 의미한다 CC_{first} 는 연결 객체로 표현된 영상을 의미한다

이렇게 생성된 연결 객체는 자막과 비자막 연결 객체

를 모두 포함한다. 또한, 자막의 연결 객체가 아니거나 각 연결 객체 간을 연결되는 에지를 포함한다 이러한 에지들은 자막을 이루는 경계의 에지가 아니기 때문에 그림1에서 표현된 것과 같이 연결 객체 내의 에지 중에서 서로 반대 방향으로 존재하는 대응되는 에지 픽셀이 존재하지 않을 확률이 높다 이러한 픽셀들을 제거하기 위해, 각 에지 픽셀들은 자신과 에지 방향과 반대에 존재하는 에지 픽셀에 자신과 반대되는 방향의 에지 픽셀이 존재하는지 확인하는 작업을 거친다 만약, 반대되는 방향의 에지 픽셀이 존재하지 않을 경우에는 자막을 이루는 에지 픽셀이 아니라 판단하고 연결 객체에서 제거한다.

이렇게 자막의 경계가 아닌 에지 픽셀을 제거한 연결 객체들은 4-이웃 연결 객체 방법을 재사용하여 CC_{second} 를 만든다. 이렇게 생성된 연결 객체 영상은 의사 결정 트리를 생성하기 위해서 그림 3의 마스크를 이용하여 각각의 불확실한 방향성을 제거한다 각 마스크는 위에서 좌상, 상, 우상, 좌, 우, 좌하, 하, 우하로 구성되며, 각각 자신이 검사할 에지 영상 E에 적용한다 예를 들면, E_t 은 mE_t 을 적용하며, mE_t 을 E_t 의 각 픽셀에 대응한다. mE_t 의 각 픽셀에 mE_t 의 중심점을 위치하고 X 사각형에 대응되는 픽셀에 에지가 존재하지 않고 O 사각형에 대응되는 픽셀에 에지가 존재한다면 그 픽셀은 정제된 에지 픽셀이라고 기록한다 이런 식으로 모든 E를 정제하면 방향의 불확실성이 비교적 제거된 에지 정보를 획득할 수 있다.

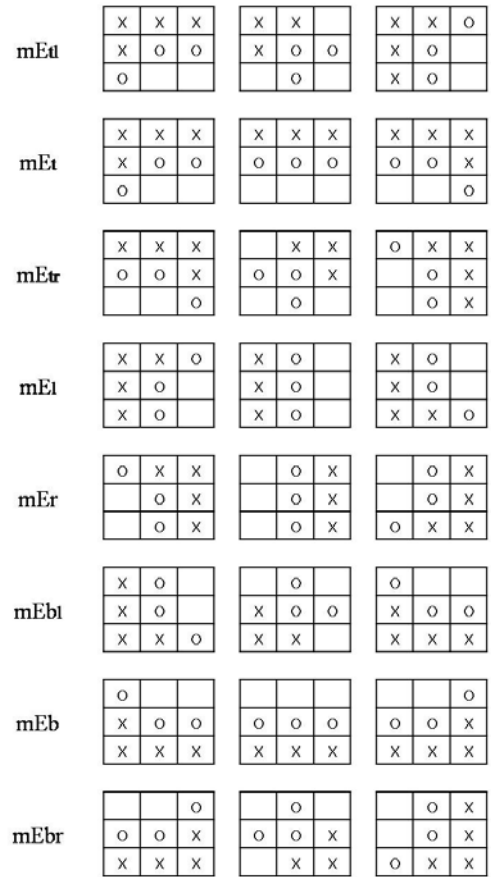


그림 3. 방향 정제 마스크

이렇게 정제한 에지 영상(RE)들은 각 대응되는 연결 객체 영상 CC_{second} 을 사용하여, 아래와 같은 특징 값들을 생성한다.

1. RE_t 의 픽셀의 수 / CC_{second} 의 픽셀의 수
2. RE_t 의 픽셀의 수 / CC_{second} 의 픽셀의 수
3. RE_{tr} 의 픽셀의 수 / CC_{second} 의 픽셀의 수
4. RE_l 의 픽셀의 수 / CC_{second} 의 픽셀의 수
5. RE_r 의 픽셀의 수 / CC_{second} 의 픽셀의 수
6. RE_{bt} 의 픽셀의 수 / CC_{second} 의 픽셀의 수
7. RE_b 의 픽셀의 수 / CC_{second} 의 픽셀의 수
8. RE_{br} 의 픽셀의 수 / CC_{second} 의 픽셀의 수
9. 각 수평으로 대응되는 에지의 중앙 픽셀의 밝기들의 표준 편차
10. 각 수직으로 대응되는 에지의 중앙 픽셀의 밝기들의 표준 편차

각 10개의 특징 값들은 글자의 크기에 크게 영향을 받지 않는다. 이러한 특징 값들로 의사결정 트리를 사전에 생성하고, 입력된 특징 값들로 각 연결 객체를 자막 연결 객체와 비자막 연결 객체로 분류한다

비자막 연결 객체로 분류된 연결 객체들은 모두 제거하고, 자막 연결 객체로 분류된 연결 객체들만을 포함한 결과 영상이 생성한다

3.2.2. 자막 위치 단계

이 과정에서는 자막과 유사한 객체를 포함하는 최소 사각 범위를 생성한다 제안하는 방법에서는 이 부분은 다른 논문에서 제안한 방법을 사용하였다 자막을 포함하는 최소 사각형을 생성하기 위해서 전체 영상에서 서서히 자막의 영역으로 좁혀가는 coarse-to-fine 방법을 사용한다[1].

초기에 자막 존재 범위는 전체 화면으로 설정된다 그 후에 수평 프로젝션 방법을 사용해서 자막 연결 객체가 존재하는 지역만을 수평라인으로 분할한다 다음에는 분할된 수평라인 내부에서 수직 프로젝션 방법을 사용하여 수직라인으로 분할한다 수직으로 분할된 내부에선 다시 수평으로 분할을 시도한다 더 이상 수직, 수평 분할이 되지 않을 때까지 수행하면 최종적으로 자막으로 판단되는 연결객체를 포함하는 최소 사각 영역을 구할 수 있다. 한글 자막의 경우엔 글자가 다수의 획으로 분리되어 존재하기 때문에, 이러한 경우에도 하나의 구역으로 잡기 위해서, 간격이 3이하인 경우에는 사각 영역을 분할하지 않았다.

3.2.3. 자막 분리 단계

이 과정에서는 자막에 포함된 픽셀들을 그 외의 픽셀에서 분리하는 작업을 수행한다 분리된 영상은 OCR의 입력으로 사용되어 자막의 글자의 인식이 가능하도록 한다. 제안하는 방법에서는 [2]에서 제안한 획 기반 이진화 방법을 사용한다

자막을 이루는 경계는 자막과 비자막의 중간색을 떨 확률이 높다. 이러한 특징을 이용하는 획 기반 이진화 방법은 다음과 같다. 먼저 자막 영역에 존재하는 연결 객체의 경계의 밝기 값을 모두 합산하고 합산한 값을 경계에 포함된 에지 픽셀의 개수로 나누어준다 이렇게 나온 값은 자막과 비자막의 경계에 존재하는 밝기 값의 평균 값이며, 자막과 비자막을 분리하는 임계치로 사용이 가능하다. 이 임계치를 사용하여 자막과 비자막 픽셀을 분리를 수행한다

4. 실험 결과

실험 데이터는 뉴스나 다큐멘터리 영상에서 수집하였다. 수집한 영상의 수는 총 668장을 사용하였으며, 12개의 방송 프로그램 영상 캡처로 수집하였다

의사결정 트리를 성성을 위해서 영상에서 약4500개의 연결 객체와 40000개의 비자막 연결 객체를 수집하고, 의사 결정 트리를 생성하였다 생성된 의사 결정 트리의 검증하기 위해서 사용된 패턴은 자막 연결 객체 800개와 비자막 연결 객체로 약 5000개가 사용하여 검증하였다. 일정 높이 이상의 의사 결정 트리는 에러율이 0.045에서 크게 변하지 않아서 트리의 높이 20에서 pruning을 수행하였다.

자막 추출의 성능의 평가를 위해서 사용된 영상은 약 180장이었으며, 4개의 방송 프로그램에서 추출하였다 영상에 포함된 자막들은 배경보다 밝은 색상으로 존재하였으며, 자막 추출의 성능을 평가하기 위해 실제로 자막

이면서 검출된 연결 객체와 제안하는 방법이 찾아낸 연결 객체의 비율인 precision과 실제로 자막이면서 검출된 연결 객체와 실제 연결 객체의 비율인recall을 사용하였고, 표 1과 같이 결과가 나타났다.

표1. 자막 추출 결과(caption extraction result)

	Recall(%)	Precision(%)
자막 추출	76%	84%

Precision은 영상에서 찾아진 연결 객체가 얼마나 정확하게 찾아졌는지를 보여준다 recall에 비해 높은 수치를 나타내었다. 본 논문은 한글을 대상으로 하였으나 그림 5의 'm'과 같이 한글 자막과 같이 존재하는 영문도 추출되는 경우가 존재했다. 이런 경우에는 오검출로 판단하였으며, 자막 주변에 존재하는 배경이 자막보다 밝거나 유사한 경우에는 잘못된 프로젝션이 수행되어 배경 픽셀이 자막 픽셀처럼 검출되었다

Recall이 낮은 경우는 주로 작은 글자의 추출에서 발생하고 있다. 그림 특히 그림 7과 같은 자막은 약 12픽셀 정도로, 획의 밀도가 높아서 에지의 강도가 약한 '흠', '월', '일'과 같은 경우에는 제대로 추출이 되지 않았다 하지만, 일반적으로 작은 글자들은 많은 정보를 표현하기 위해 사용되기 때문에 한 라인에 큰 글자에 비해 많이 등장한다. 이러한 특징을 이용하여 작은 글자들을 다시 복구하는 모폴로지 방법의 연구를 하고 있다 또한, 그림 7에서 좌측에 보이는 넓은 지역의 연결 객체는 다른 자막 연결 객체에 비해서 크기가 거대하므로 이러한 연결 객체를 제거하는 연구가 필요하다

5. 결론

영상은 다른 콘텐츠보다 직관적으로 정보의 전달이 가능하다. 하지만, 다른 콘텐츠에 비해서 규모가 크며 자동적으로 검색이나 저장을 연구가 필요하다 이러한 영상의 검색 및 저장을 위해서 영상의 내용과 관련이 있는 영상의 자막을 추출하는 연구가 진행되어 왔다

기존의 자막 추출 연구들은 자막을 추출하기 위해서 자막을 이루는 에지의 정보를 추출하기 위한 연구가 있었다. 하지만, 이러한 연구들은 언어에 독립적인 특징만을 고려하여 추출하였으며, 그렇기 때문에 특정한 형태학적인 조건을 만족하지 못하면 정상적인 작동을 보장하지 못하였다.

제안하는 논문에서는 입력 영상으로부터 에지 픽셀들을 추출하여 연결 객체를 구성한 후에 의사결정 트리로 분석한 연결 객체 패턴을 이용하여 연결 객체를 자막과 비자막 연결 객체로 분류하였다 이렇게 분류된 연결 객체는 대부분의 자막 연결 객체를 포함하였다

향후에는 모폴로지 연산을 이용해서 의사결정 트리에 의한 100%가 아닌 분류를 보완하기 위해 작은 자막 연결 객체들을 복구하는 연구와 자막과 연결 객체 속성이

유사한 비자막 연결 객체를 제거하는 연구를 할 것이다



그림 4 입력 영상 1

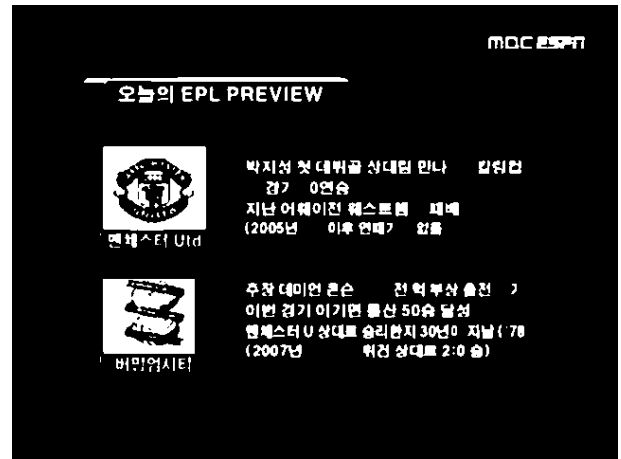


그림 7 결과 영상 2



그림 5 결과 영상 1

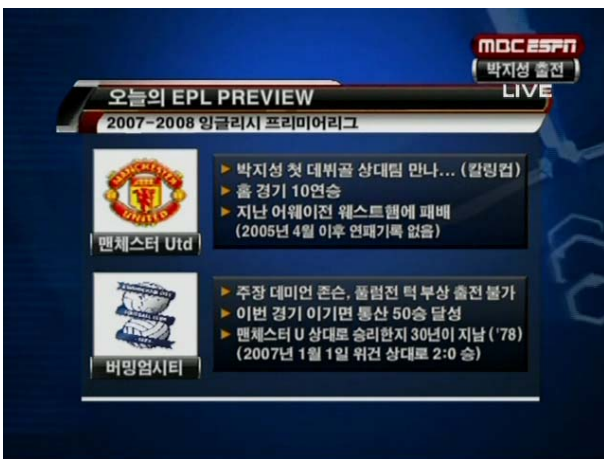


그림 6 입력 영상 2

6. 참고 문헌

- [1] R. Lyu, J. Song, M. Cai, "A Comprehensive Method for Multilingual Video Text Detection, Localization and Extraction", IEEE Transaction on Circuits and Systems for Video Technology, Vol. 15, No. 2, pp. 243-255, 2005.
- [2] 정종면, 차지훈, 김규훈, "디지털 비디오를 위한 획기적인 자막 추출 알고리즘, 퍼지 및 지능시스템학회 논문지, Vol. 17, No. 3, pp. 297-303, 2007.
- [3] K.C. Jung, K.I. Kim, A.K. Jain, "Text Information Extraction in Images and Video: A Survey", Journal on Pattern Recognition, Vol. 37, No. 5, pp. 977-997, 2004.
- [4] 전병태, 배영래, 김태윤, "일반화된 문자 및 비디오 자막 영역 추출 방법", 정보과학회논문지:소프트웨어 및 응용, Vol. 27, No. 6, pp. 632-641
- [5] E.K. Wong, M. Chen, "A Robust Algorithm for Text Extraction in Color Video", IEEE Int'l Proc. Multimedia and Expo 2000(ICME 2000), Vol. 2, pp. 797-800, 2000.
- [6] K.C. Jung, E.Y. Kim, "Automatic Text Extraction for Content-Based Image Indexing", Lecture notes in Computer Science, Proc. 8th Pacific -Asia Conf. (PAKDD 2004), Vol. 3056, pp. 497-507, 2004.
- [7] Y.M.Y. Hasan, L. J. Karam "Morphological Text Extraction from Images", IEEE Trans. Image Processing, Vol. 9, No. 11, pp. 1978-1983, 2000.
- [8] H.E. Jiaying, L.I. Shaofa, "Hybrid Chinese/English Text Identification in Web Images", Proc. 3rd Int'l Conf. Image and Graphics(ICIG '04), pp. 361-364, 2004.
- [9] R. C. Gonzalez, Digital Image Processing, 2nd edition, Prentice Hall, New Jersey, 2001.