비디오 자막 추출 및 이미지 향상에 관한 연구

김소명°, 최영우, 정규식*
숙명여자대학교 컴퓨터과학과, *숭실대학교 정보통신전자공학부
kimsm@cs.sookmyung.ac.kr, ywchoi@sookmyung.ac.kr, kchung@q.soongsil.ac.kr

Video Caption Extraction and Image Enhancement

Somyung Kim°, Yeongwoo Choi, Kyusik Chung*
Dept. of Computer Science, Sookmyung Women's University
*School of Electronic Engineering, Soongsil University

요 약

본 논문에서는 비디오 자막 이미지를 인식하기 위해 필요한 영상 향상의 단계로서 다중 결합을 적용한다. 또한 다중 결합을 위한 동일한 자막의 판단 및 결합된 결과를 재평가하기 위한 방법을 제안한다. 입력된 칼라 이미지로부터 RLS(Run Length Smearing)가 적용된 에지 이미지를 얻고, 수직 및 수평 히스토그램 분포를 이용하여 자막과 자막 영역에 대한 정보를 추출한다. 프레임 내의 자막 영역의 중첩 정도를 이용하여 동일 자막을 판단하고, 동일한 자막을 갖는 프레임들끼리 다중 결합을 수행함으로써 향상된 이미지를 얻는다. 끝으로 결합된 영상에 대한 평가를 수행하여 잘못 결합된 이미지들로 인한 오류를 해결하고 재평가한다. 제안한 방법을 통해, 배경 부분의 잡영이 완화된 자막 이미지를 추출하여 인식의 정확성과 신뢰성을 높일 수 있었다. 또한 동일한 자막의 시작 프레임과 끝 프레임의 위치 파악은 디지털 비디오의 색인 및 검색에 효과적으로 이용될 수 있을 것이다.

1. 서론

비디오 영상에 포함되어 있는 자막은 비디오의 내용을 함축적으로 표현하고 있기 때문에 이 자막을 정확하게 인식할 수 있다면 비디오의 색인 및 검색에 중요하게 사용될 수 있다. 또한 비디오 자막은 동/정지 영상과 음성, 음향 정보에서 표현하고 있지 않는 내용도 포함하는 경우가 있다. 그러나 비디오 자막 이미지의 해상도가 낮고 복잡한 배경을 포함할 수 있기 때문에 비디오 자막을 인식하는 것이 아주 어려운 실정이다[1-6].

비디오 자막을 정확하게 인식하기 위해서는 자막 영역의 해상도를 증가시키고, 복잡한 배경을 제거할 수 있는 영상 향상 과정이 필요하다. 기존의 영상 향상에 관한 연구를 보면, 영상의 해상도를 증대시킨 후 적응적 이진화를 적용하거나 다중 결합을 통해 복잡한 배경을 제거하는 방법 등이 제안되었다[2, 3]. 다중 결합은 동일한 자막 프레임들 중 각 위치에서 가장 작은 픽셀값을 선택하여 결과 이미지를 얻는 방법이다. [1]에서는 다중 영상의 AND 결합, 해상도 증대, 히스토그램 평활화, 문자의 획두께를 고려한 이진화와 몰포로지를 영상에 단계적으로 적용하여 자막 이미지 및 인식의 정확성을 향상시켰다[1]. 그러나, [1]의 연구에서는 다중 영상의 AND 결합을 적용하기 전에 동일한 자막을 갖는 프레임 판별이 이루어졌다는 가정 하에 연구가 수행되었다. 본 논문은 다중 결합을 이용한 이미지 향상에 필요한 동일한 자막의 판단 및 결합된 결과를 평가하여 오류를 수정하고 재평가하는 과정을 수행한다. 자막 이미지 향상을 위해 본 연구에서는 그림 1에서와 같이 두 단계의 이미지 향상 방법을 적용하였다. 1단계 향상에 관한 연구는 본 논문에서 수행된다. 2단계의 연구는 저자들의 이전 연구인 참고문헌[1]을 참고바랍니다.

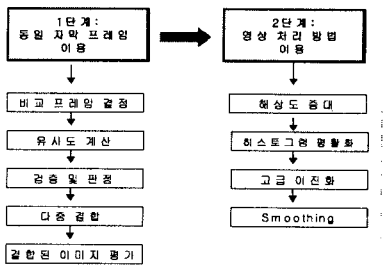

그림 1. 2단계 이미지 향상

2. 자막 추출

본 논문에서는 자막 영역을 추출하기 위해서 칼라 이미지를 명도 이미지로 변환하여 에지 이미지를 얻는다. 에지 이미지로부터 자막 영역을 강조하기 위해서 수평 및 수직 방향으로 RLS를 적용한다. RLS가 적용된 에지 이미지로부터 수직 및 수평 히스토그램 분포를 분석하여 에지 밀도가 조밀한 영역을 자막 후보 영역으로 찾아낸다. 추출된 후보 영역이 정확한 자막 영역인가를 검증하기 위해서, 본 논문에서는 비디오 자막의 사전 위치 정보와 에지 밀도를 이용한다.

자막 영역이 결정되면 각 영역의 위치, 크기, 에지 값의 밀도, 프레임 내의 자막 개수 등의 정보를 추출한다. 자막 추출 과정에서 하나의 자막이 수직으로 분할되어 나타나는 경우가 발생한다. 본 연구에서는 비디오 자막만을 대상으로 하기 때문에, 수평으로 인접한 영역들이 같은 높이에서 존재하면 두 영역의 결합을 시도한다. 영역의 결합 조건은 참고문헌[4]를 참조하여 결정하였다. 그림 2는 자막 영역 추출을 위해 제안된 방법을 수행하는 과정에서

생성된 이미지들을 보여준다.

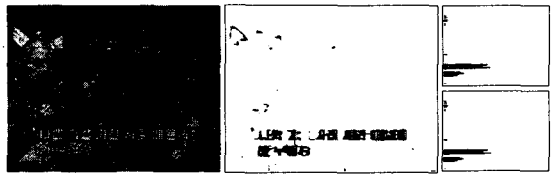

그림 2. 자막 영역 추출: 추출된 자막 영역이 표시된 원영상(좌), RLS 적용한 에지 영상(중), 수평 프로젝션 프로파일(우상), 수직 프로젝션 프로파일(우하).

3. 동일한 자막의 다중 결합을 이용한 이미지 향상

비디오 영상에서 동일한 자막은 연속적인 프레임들의 동일한 위치에 나타나며, 자막의 크기나 명도값, 폰트에 있어서 유사한 정보를 갖는다. 본 논문에서는 이러한 특징을 이용하여, 자막 영역의 중첩 정도를 계산하여 동일 자막 프레임을 판단한다. 동일 자막 프레임들간의 다중 결합을 통해 배경 부분을 제거하여 향상된 이미지를 추출한다.

3.1 비교 프레임 결정

비교 프레임 결정을 위해서, 자막이 추출된 두 프레임 사이에 자막이 추출되지 않은 프레임의 개수를 파악한다. 파악된 개수는 두 프레임으로부터 추출된 자막이 서로 동일할 가능성이 있는지를 결정하는 기준으로 사용된다. 동일 자막 프레임들 중간에 일시적으로 추출이 발생하지 않는 경우가 있기 때문에, 파악된 프레임의 개수가 임계값 이하이면, 자막이 동일할 가능성이 있다고 판단하고 비교 프레임으로 결정한다. 개수가 임계값을 초과하면, 서로 다른 자막들로 간주하고 비교하지 않는다. 임계값은 실험 대상이 되는 영상에서 동일한 자막이 나타나는 연속적인 프레임의 수를 이용하여 결정하였다.

3.2 유사도 계산

현 프레임, *t*가 비교할 이전 프레임, *t-1*과 동일한 자막을 포함하고 있는지를 판별하고 라벨링하기 위해서 두 프레임간의 유사도(similarity)를 계산한다. 유사도는 현 프레임에서 추출된 자막 영역과 비교되는 이전 프레임에서 추출된 자막 영역들이 중첩되는 영역의 비율로 정의하며, 아래의 식과 같다.

$$Similarity = \frac{1}{N}\sum_{i=0}^{N-1}\left\{\left(\frac{OS_i}{SP_i + SC_i}\right)\times 2\right\}$$

*SP*는 비교되는 이전 프레임의 자막 영역의 크기, *SC*는 현 프레임의 자막 영역의 크기이며, *OS*는 *SP*의 자막과 *SC*의 자막이 중첩되는 영역의 크기이다. *N*은 *SP*와 *SC*중 더 높은 자막 개수이며, *i*는 한 프레임 내에서의 자막 번호이다. 유사도가 임계값 이상이면 현 프레임과 이전 프레임에 동일한 라벨을 부여한다. 그런데, 단순히 이전 프레임과의 비교만으로 동일 자막을 판별하게 되면, 동일한 자막을 갖는 프레임들인데도 불구하고 서로 다른 라벨을 부여하게 되는 문제가 발생한다. 자막 주위의 배경에 의해 일시적으로 자막 영역이 다르게 추출되기 때문이다. 따라서 먼저 *t*와 *t-1* 프레임을 비교한 후, 유사도에 차이가 있는 경우에는 *t*와 *t-2* 프레임을 비교한다. *t*와 *t-2*의 유사도가 높으면 *t, t-1, t-2* 프레임들에 대해 동일한 라벨을 부여한다.

3.3 검증 및 판정

라벨링에 대한 검증은 두 프레임에서 추출된 자막 영역에 대한 에지 값의 합의 차이를 이용한다. 서로 다른 자막이면서 자막의 길이나 추출 영역이 유사하여 유사도가 높게 나타난 경우가 있다. 이런 경우, 자막 영역내의 자막에 대한 자료인 에지 값의 합을 비교함으로써 서로 다른 자막으로 분류해 낼 수 있다. 에지 값의 합의 차이가 임계값 이상이면, 유사도가 높더라도 서로 다른 자막으로 판단하고, 이전 프레임과는 다른 라벨을 현 프레임에 부여한다.

동일한 라벨이 임계값 이상 연속되어 나타나지 않는 소수 라벨들이 있다. 이는 자막이 없는 프레임에서 자막이 아닌 영역이 일시적으로 추출되는 경우가 발생하기 때문이다. 따라서 이러한 소수 라벨들이 부여된 프레임들은 자막이 없는 프레임에서 추출이 발생한 것으로 판단하여, 다중 결합시 무시한다. 동일 자막 판정의 결과는 다음과 같이 다섯 종류로 분류된다. 1) 동일한 자막 프레임들 모두에게 하나의 라벨이 부여된 경우, 2) 동일한 자막 프레임들 가운데 일부만 라벨링된 경우, 3) 동일한 자막 프레임들이 두 개 이상의 다른 라벨을 부여받는 경우, 4) 서로 다른 두 개의 자막 그룹이 동일한 라벨을 부여받는 경우, 5) 하나의 라벨이 서로 다른 두 개의 자막 그룹의 일부 프레임들에 걸쳐있는 경우.

3.4 다중 결합

비디오 영상에서는 동일한 자막이 여러 프레임에 걸쳐 동일한 위치에 연속적으로 나타나는 반면, 배경은 변화한다. 이 특징을 이용하여 다중 결합을 수행함으로써 자막 영역에 포함된 복잡한 배경부분을 제거시킨다. 다중결합은 동일 자막을 갖는 프레임들의 동일 위치의 픽셀 값들 중에서 최소값을 결과 값으로 선택하며, 아래의 식과 같다.

$$TG_i(x, y) = MIN(TF_m(x, y), TF_{m+1}(x, y), \ldots, TF_n(x, y))$$

위의 식에서 $TG_i(x, y)$는 *i*번째 라벨을 갖는 프레임들의 동일한 *x, y* 좌표에 대해 다중결합한 픽셀 값이고, $TF_j(x, y)$는 *j* 번째 프레임의 *x, y* 좌표에 대한 픽셀 값이다. *m*과 *n* 은 각각 *i* 번째 라벨을 갖는 프레임 그룹의 시작 프레임과 마지막 프레임 번호이다.

3.5 결합된 이미지의 평가

한 라벨이 동일한 자막 프레임에게만 부여된 경우는 다중결합 후, 배경 부분의 잡영이 제거되어 향상된 이미지를 얻을 수 있었다. 그러나, 서로 다른 자막 프레임에 동일한 라벨이 부여된 경우는 다른 자막들이 겹쳐져 자막 이미지가 훼손되었다. 이런 경우, 서로 다른 자막 프레임들을 분류해 내기 위해 동일한 라벨을 갖는 프레임들을 한 프레임씩 결합하면서 자막 영역으로 설정된 부분에 대한 밝은 화소의 개수를 체크한다. 밝은 화소의 갯수가 급격히 변하는 시점을 새로운 자막의 시작 프레임으로 설정한다. 설정된 시작 프레임을 기준으로 두그룹으로 나누어 각각 다중 결합을 수행한 후 재평가한다.

4. 실험 및 결과

실험 환경은 Pentimum III 600MHz PC로 Visual C++ 6.0을 이용하여 제안한 방법들을 구현하였다. 실험에 사용된 영상 데이터는 일반 영화 비디오 30 *frame/sec*를 8 *frame/sec*로 샘플링하여, 크기 640 x 480 해상도의 AVI 포맷으로 변환하여 사용하였다.

표 1은 동일 자막 판별을 위한 실험 결과 데이타이다. 동일 자막 여부를 구별하기 위해 프레임에 부여된 총 라벨 개수와 소수 라벨을 제외한 최종적인 라벨의 개수를 보여주고 있다.

표 2와 3은 각각 한글자막 영상과 영어자막 영상에 대한 동일 자막 판단 실험 결과 중에서 표 1의 소수라벨을 제외한 최종적인 라벨 유형과 라벨 개수에 대해 분석한 것이다. 표 2와 3의 각 항목은 다음과 같다. (I) 동일한 자막 프레임들 모두에게 동일한 하나의 라벨이 부여된 경우, (II) 동일한 자막 프레임들 가운데 일부만 라벨링된 경우, (III) 동일한 자막프레임들이 두 개 이상의 다른 라벨을 부여 받는 경우, (IV) 서로 다른 두 개의 자막 그룹이 동일한 라벨을 부여 받는 경우, (V) 하나의 라벨이 서로 다른 두 개의 자막 그룹의 일부 프레임들에 걸쳐 있는 경우. 결과 유형에

서 표 2와 3에서 III, IV, V의 경우는 동일한 자막을 다른 자막으로, 또는 다른 자막을 동일한 자막으로 잘못 판단하여 라벨링한 것이다. 이러한 예는 자막 주위의 배경 변화가 에지 값에 변화를 주게 되어 자막 영역이 정확하게 추출되지 않기 때문에 발생한다.

표 1. 동일 자막 판별을 위해 프레임에 사용된 라벨 개수

구분	총 자막 개수	부여된 라벨개수	최종 라벨개수 (소수라벨 제외)
한글영상 1	63	84	65
한글영상 2	72	72	66
영어영상	128	331	115

표 2. 실험 영상에 대한 동일자막 판별 결과

구분	I	II	III	IV	V
한글영상 1	44	9	6	3	3
한글영상 2	51	6	2	0	7
영어영상	57	32	14	10	2

다중 결합의 결과는 크게 표 2에서의 I, II, III에 해당하는 동일한 자막들만 결합한 경우와 IV와 V에 해당하는 다른 자막들을 결합한 경우로 분류할 수 있다. 동일한 자막들을 결합한 경우는 결합 이미지에서 자막 영역이 잘 보존된다. 그러나 다른 자막들을 결합한 경우는 그림 3과 같이 자막이 훼손되는 형태로 나타났다.

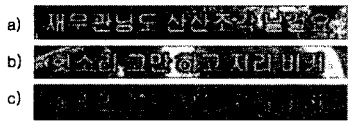

그림 3. 다른 자막들에 대해 동일한 라벨을 부여한 경우의 다중결합: a) 자막 1; b) 자막 2; c) 자막 1과 2에 해당하는 프레임들을 다중결합.

또한 배경의 변화에 따라, 자막 영역의 배경이 급격히 변하는 경우, 배경이 천천히 변하는 경우와 배경의 변화가 거의 없는 경우로 분류할 수 있다. 그림 4과 같이 배경의 변화가 있는 경우는 배경으로 인한 잡영이 많이 제거 되었으나, 그림 5와 같이 배경의 변화가 거의 없는 경우는 다중 결합으로는 향상된 결과를 얻을 수 없었다.

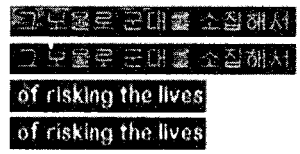

그림 4. 배경의 변화가 급격한 경우: a), c) 다중 결합하지 않고 추출; b), d) 다중 결합한 후 추출.

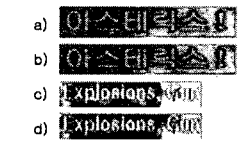

그림 5. 배경의 변화가 거의 없는 경우: a), c) 다중 결합하지 않고 추출; b), d) 다중 결합한 후 추출.

표 2에서 문제가 되는 IV와 V의 경우를 수정하기 위해 서로 다른 자막 프레임을 분류하는 작업을 수행하면 그림 6, 7과 같이 분류되어 각각 다중 결합된다. 한글영상 1과 영어영상은 IV와 V에 해당하는 자막들 모두가 성공적으로 분류되었고, 한글영상 2는 IV와 V에 해당하는 자막들 중 단 한 개를 제외하고는 모두 성공적으로 분류되었다. 그리고, 영어 자막에서 동일한 자막 프레임들인데도 불구하고 배경의 급격한 변화로 인해, 두 그룹으로 나뉘는 잘못된 경우가 한 번 발생하였다.

그림 6. 한글영상: a) 자막 프레임 분류를 수행하기 이전 다중 결합한 이미지; b), c) 분류된 자막 프레임들을 각각 다중 결합한 이미지

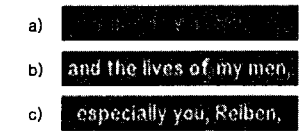

그림 7. 영어영상: a) 자막 프레임 분류를 수행하기 이전 다중 결합한 이미지; b), c) 분류된 자막 프레임들을 각각 다중 결합한 이미지

5. 결론

본 논문에서는 다중 결합을 이용한 이미지 향상에 필요한 동일 자막 프레임을 판별하고, 결합 결과를 재평가하는 방법을 제안하였다. 제안된 방법을 적용하여, 동일한 자막 프레임이면서 추출 영역이 일정치 않은 경우라도 동일한 자막들을 판정해낼 수 있었다. 또한 동일 자막 프레임들의 다중 결합을 통해 자막 이미지들이 가지고 있던 복잡한 배경들이 제거되는 결과를 얻을 수 있었다. 동일자막들의 처음과 끝 프레임의 위치를 판정함으로써 비디오 검색 및 색인을 위해 활용될 수 있을 것이다. 향후 연구 과제로는 하나의 자막이 분리 추출된 경우에 자막 영역을 결합하는 방법을 보완하고 추출된 자막 이미지의 인식 실험을 통해 제안한 방법의 성능을 평가하는 것이다. 또한 제안한 방법을 다양한 종류의 영상들에 대해서도 적용할 수 있도록 일반화하는 연구가 필요하다.

참고 문헌

[1] 곽상신, 김소명, 최영우, 정규식, "효율적인 비디오 자막 인식을 위한 영상 향상 방법", 제 12회 영상처리 및 이해에 관한 워크샵 발표 논문집, pp. 342-346, 2000.

[2] H. Li, O. Kia, D. Doermann, "Text Enhancement in Digital Video", Part of the IS&T/SPIE Conference on Document Recognition and Retrieval VI, SPIE Vol. 3651, pp. 2-9, 1999.

[3] T. Sato, T. Kanade, E. K. Hughes, M. A. Smith, "Video OCR for Digital News Archives", IEEE Workshop on Content-Based Access of Image and Video Databases(CAIVE '98), Bombay, India, January, 1998.

[4] Y. Shong, H. Shang, A. K. Jain, "Automatic Caption Localization in Compressed Video", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 4, pp. 385-392, 2000.

[5] 전병태, 정세윤, 이재연, 배영태, "뉴스 아이콘 자막 및 내용 자막 추출", 한국 정보과학회 가을 학술 발표집, Vol. 24, No. 4, pp.385-392, 2000

[5] 이미숙, 방건, 임영규, 홍영기, 김두식, 이성환, "내용기반 색인 및 검색을 위한 실시간 뉴스 비디오 파서의 설계 및 구현", 한국정보과학회 가을 학술발표집, Vol. 24, No.1, pp. 365-268, 1997.

[6] H. Li, D. Doermann, "Automatic Identification of Text In Digital Video Key Frames", In Proceedings of ICPR, Toshino, pp. 129-13 2, 1998.