모폴로지를 이용한 비디오 영상에서의 자동 문자 추출

장인영[0] 고병철 김길천 변혜란
연세대학교 컴퓨터과학과
{stefano, soccer1, kimkch, hrbyun}@aipiri.yonsei.ac.kr

Automatic Text Extraction in Video Images using Morphology

InYoung Jang[0] , ByoungChul Ko, KilChun Kim, HyeRan Byun
Dept. of Computer Science, Yonsei University

요 약

본 논문에서는 뉴스 비디오의 정지 영상에서 뉴스 자막과 배경 문자를 추출하기 위한 새로운 방법을 제안한다. 본 논문에서는 일차적으로 입력 컬러 영상을 그레이 영상으로 변환한 후 입력 영상의 명암 대비를 강화시키기 위해 명암 대비 스트레칭을 적용한다. 이후 명암 대비 스트레칭된 영상의 분할을 위해 적응적 임계값을 적용하고 다음 단계에서 문자와 유사한 영역들을 적당한 크기의 structuring element를 이용하여 제거하는 1차 하부 단계와 모폴로지 녹임(erosion)을 적용한 영상과 모폴로지(열림닫힘[OpenClose]+닫힘열림[CloseOpen])/2가 적용된 영상 사이의 차이 영상을 구하는 2차 하부 단계를 적용 시킨다. 마지막 단계에서 각 후보 영역들 중 실제 자막 영역을 추출해내기 위해, 후보 문자 영역의 화소수 비율과 외곽선의 화소수의 비율, 그리고 장축과 단축간의 비율 등에 대해 필터링을 적용한다. 본 논문에서는 임의의 300개의 뉴스영상을 입력 값으로 실험한 결과 93.6%의 우수한 인식률을 얻을 수 있었다. 또한 본 논문에서 제안한 방법은 structuring element의 크기 조절을 통해 크기가 다른 다양한 이미지에서도 좋은 성능을 거둘 수 있다.

1.서 론

최근 들어 인터넷 사용의 증가와 더불어 디지털 비디오의 수요 또한 급격히 증가하고 있는 추세이다. 따라서 디지털 비디오 데이터베이스의 인덱싱을 위한 자동화된 도구가 필요 하게 되었고 이에 따른 다양한 방법이 연구 되어 오고있다. 기존의 전형적인 디지털 비디오 인덱싱 방법은 관리자가 비디오를 보고 직접 분할한 후 적절한 색인어를 입력하는 방식으로 이는 상대적으로 방대한 양의 디지털 비디오를 모두 처리하기에는 상당한 양의 시간과 인력을 필요로 하기 때문에 극히 비효율적이며 또한 관리자 개개인의 주관에 의해 색인어를 입력할 수 있으므로 색인어 입력에 있어서 오류를 범할 소지가 많다[1].

현재까지 비디오 자동 인덱싱을 실현하기 위해 샷 분할 방법과 문자 추출 방법이 제안되어져 왔다. 샷 분할 방법은 그 자체만으로는 디지털 비디오 인덱싱의 실현에 한계가 있기 때문에 최근에는 비디오 문자 추출 분야로 관심이 모아 지고 있는 상황이며 디지털 비디오에 나타나는 자막 문자와 배경에 포함되어진 문자 정보들은 이러한 문제를 해결할 수 있는 중요한 단서가 된다[1-5]. 또한 최근의 논문들에서 이러한 문제를 해결하기 위한 다양한 해결책들을 찾을 수 있다. 일반적으로 이 분야에서 전형적인 문제점은 문자들의 다양한 크기와 서체, 문자들의 방향과 위치의 변화, 다른 질감, 일정치 않은 빛의 변화, 불규칙한 배경, 그리고 문자 획의 컬러 gradient등 이다[6]. 본 논문에서는 문자 추출의 이러한 문제들을 해결하기 위해 본 논문에서 제안한 모폴로지 오퍼레이션과 Geo-Correction 방법을 선 처리에 사용하고 후보 자막 영역 중 실제 자막 영역만을 추출하기 위해 필터링 알고리즘을 적용한다.

2.선처리 단계

비디오 자막 추출 시스템은 디지털 비디오의 인덱싱을 위한 자동 문자 추출 시스템이다. 이러한 문자 추출 시스템의 공통적인 처리순서는 기본적으로 세가지 정도로 분류 되어진다. 첫번째는 선처리 단계로 입력 영상을 컬러 영상 자체로 사용할 것인지 그레이 영상으로 사용할 것인지를 알고리즘에 따라 결정하는 것이다. 두 번째는 후보 문자 영역들은 유지한 채 비 문자 요소들을 제거하기 위해 기존의 다양한 영상 처리 알고리즘을 적용하여 세 번째 단계인 필터링 단계에서 사용할 입력 값으로 처리한다. 선처리 단계는 모든 문자 추출 시스템에 있어서 가장 중요한 단계로 이 과정의 성능에 따라 문자 추출의 인식률이 좌우 되어진다. 본 논문에서는 이 선처리 단계에서 모폴로지 오퍼레이션을 기반으로 한 (OpenClose+CloseOpen)/2 연산자와 Geo-Correction 방법을 제안한다. 본 논문에서 제안한 알고리즘의 가장 기본적인 개념은 먼저 모폴로지 오퍼레이션을 사용하여 후보 문자 요소들을 제거한 영상을 구한 후 (OpenClose+CloseOpen)/2와 Geo-Correction이 적용된 영상과의 차이 영상을 얻어냄으로써 비 문자 요소는 제거하고 후보 문자 요소들만 추출하는 선처리 단계이다. 이 선처리 단계의 결과를 입력으로 받아 임계값을 기준으로 후보 문자 영역들을 여러 가지 필터링 기법을 통해 비 문자 요소들을 제거해 최종적으로 문자만을 추출한다.

2.1 단계 1: 명암 대비 스트레칭과 적응적 임계값

본 논문에서는 입력 받은 컬러 영상을 그레이 영상으로 변환한 후 명암 대비 스트레칭을 적용한다. 이후 영상을 (Width/30)*(Height/30)의 블록으로 분할한 후 적응적 임계값을 적용하여 이진화 된 영상을 만든다. 이때 적응적 임계값은 각 블록 화소 값의 평균과 분산으로 구해진다.

2.2 단계 2: 모폴로지와 Geo-Correction

단계 1의 출력 값은 단계 2 과정에서 두개의 하부 과정으로 나뉘어진다. 첫번째 하부 과정은 이진화 영상에 모폴로지 녹임(erosion), 모폴로지 불림(dilation), 그리고 모폴로지(OpenClose+CloseOpen)/2가 순서대로 적용되어진다. 이 과정에서 모폴로지 녹임(erosion)과 불림(dilation)은 3*3의 structuring element를 사용하는데 이는 비디오 영상에서 나타나는 후보 문자 요소를 제거하기 위한 자막 문자 획의 폭을 조사한 결과에 따른 것이다(그림 2의 c 참조). 두 번째 하부 과정에서는 (OpenClose+CloseOpen)/2와 Geo-Correction이 적용된다(그림 2의 d 참조). 하부 과정 1과 2에서 적용된 (OpenClose+CloseOpen)/2 방법에서 structuring element는 오직 수평방향으로만 적용하는데 이것은 대부분의 자막이 수평방향으로 존재하기 때문이다. 본 논문에서 제안한 Geo-Correction은 화소들간의 간격을 임계값에 의해 연결해주거나 유지시켜주는 역할을 한다. 그림 3에서 해당 화소의 값이 255일 경우 연속된 화소를 따라 가다가 새롭게 0인 화소를 만나게 되면 다시 255의 화소를 만날 때까지 0인 화소수를 세게 된다. 그리고 중간에 존재하는 0의 연속적인 화소수가 임계값 이하이면 0을 255로 만들어 주어 두개의 단절된 선을 이어주고 임계값 이상이라면 원래 0의 값으로 유지 시킨다. 이 방법은 모폴로지 오퍼레이션의 적용에 의해 손상되어진 후보 문자 요소들을 보존하거나 보상해주는데 있어서 매우 효과적이다. 다음 과정은 위의 두 하부단계로부터의 결과 영상간의 차이 영상을 얻어냄으로써 후보 문자 요소들만 남기게 된다.(그림 2의 e 참조)

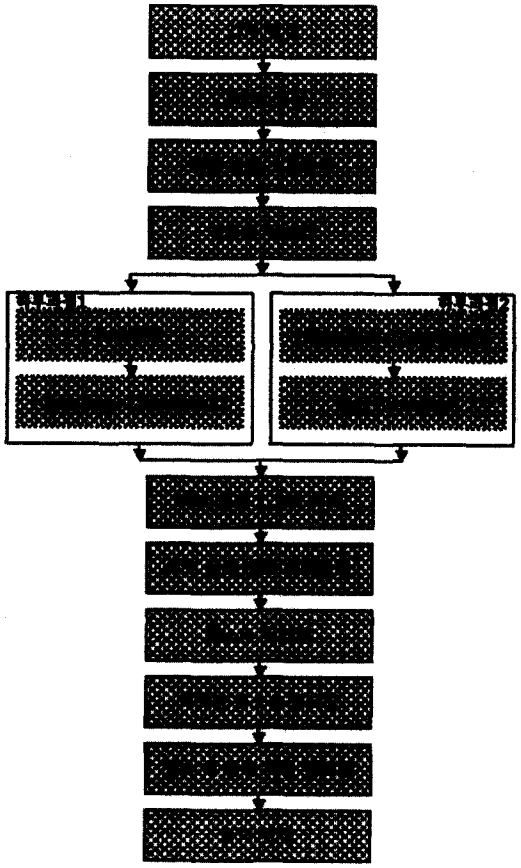
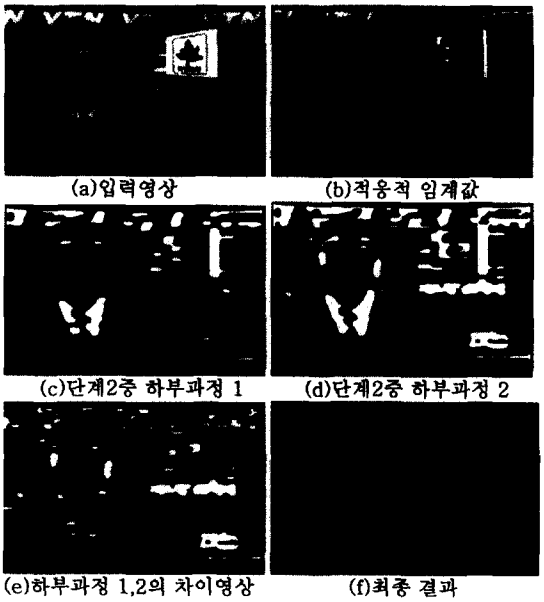


그림 1.비디오 영상에서 자동 문자 추출 알고리즘



(a)입력영상 (b)적응적 임계값
(c)단계2중 하부과정 1 (d)단계2중 하부과정 2
(e)하부과정 1,2의 차이영상 (f)최종 결과

그림 2. 선처리 단계의 과정

if 0 의 개수 <= threshold ; 연결
0 의 개수 > threshold ; 유지

■: 255 □ : 0

그림 3. Geo-Correction

2.3 단계 3: 후보 문자 영역의 필터링

본 필터링 단계에서는, 후보 문자 영역에서의 비 문자 요소들과 잡음 요소들을 제거한다. 본 논문에서는 먼저 후보 문자 영역 레이블링을 수행한 후 각각의 레이블 된 영상으로 다음 단계의 필터링을 수행한다.

2.3.1 후보 문자 영역의 화소수를 이용한 필터링 방법

전체 영상 픽셀 수 대비 4%이하의 화소수를 가진 후보 문자 영역은 제거된다.

2.3.2 후보 문자 영역의 외곽선의 화소수를 이용한 필터링 방법

후보 문자 영역과 그 경계선(edge) 사이의 화소수의 비율이 0.23이하인 요소들은 제거된다.

2.3.3 후보 문자 영역의 장 축대 단축의 비율을 이용한 필터링 방법

후보 문자 영역에 bounding box를 만들고bounding box의 장축대 단축의 비율이 0.66이하인 요소들은 제거된다.

3. 실험 결과 및 분석

본 논문에서는 웹(web)으로부터 획득한 한국과 CNN뉴스 300개의 컬러 영상을 320*210크기의 영상으로 변환하여 제안한 알고리즘을 평가하였다. 이 실험 데이터는 연속적인 프레임들로부터 임의로 추출되어졌으며, 결과를 평가하기 위하여 세 가지의 성능 기준을 정하였다. 첫 번째는 정확 인식

률 (PRR1: precise recognition rate)로써 문자를 정확하게 추출해낸 경우이다. 두 번째는 실용 인식률(PRR2: practical recognition rate)로써 허용될 수 있는 약간의 오차를 포함한 경우이다. 마지막 세 번째는 실패율(ERR: erroneous recognition rate)로써 문자 요소가 아님에도 문자로 인식한 경우와 존재하는 문자를 추출해내지 못한 경우 모두이다.

PRR1 = Tct / Tat
PRR2 = (Tct + Tpl) / Tat
ERR = (Tnt + Tfe) / Tat

Tct: 정확하게 추출된 문자 영역의 총수
Tat: 실제 문자 영역의 총수
Tpl: 허용 오차를 포함하여 추출된 문자 영역의 총수
Tnt: 비문자 영역을 문자로 추출한 총수
Tfe: 문자 영역의 추출에 실패한 총수

본 논문에 사용된 실험 데이터에서 실제 문자열의 수는 총 687개이다. 정확한 인식률은 77.5%(533 문자열)로 나타났다. 실용적 인식률은 93.6%(643 문자열)로 나타났으며 실패율은 12.8%(비문자를 문자로 추출한 수 46문자열, 실제 문자를 추출하지 못한 수 42문자열)로 나타났다.(표1 참조)

Tct	Tpl	Tnt	Tfe
533	110	46	42
77.5%	93.6%	12.8%	

표1. 실험 데이터에 대한 문자 추출의 인식률

4. 결론 및 향후 연구 과제

본 논문에서는 디지털 비디오의 영상으로부터 문자를 추출하기 위한 새로운 방법을 제안하였다. 본 연구 분야에서 우수한 결과를 얻기 위해서는 선처리 단계가 무엇보다 중요하다. 따라서 본 논문에서 제안한 알고리즘의 가장 중요한 점은 모폴로지 오퍼레이션에 기반을 둔 선처리 과정에 있다. 선처리 과정에서 우수한 결과를 얻기 위해서 먼저 문자의 폭이 고려된 적당한 크기의 structuring element를 사용하여 비 문자 요소들을 제거하고 이후, 이 녹임(erosion) 영상과 (OpenClose+ CloseOpen)/2가 적용된 영상간의 차이 영상을 얻어내는데 이는 후보 문자 요소들만을 남기는데 매우 효과적이다. 열림 닫힘(OpenClose) 과 닫힘 열림(CloseOpen) 처리는 잡음을 감소 시키고 영역 내에서 존재하는 빈 공간을 채우는데 효과적이다. 또한 본 논문에서 제안한 Geo-Correction 알고리즘은 모폴로지 오퍼레이션에 의해 손상을 받은 후보 문자 요소들을 보존하고 보상하는데 좋은 효과를 준다. 본 논문의 실험결과는 제안한 문자 추출 알고리즘이 디지털 비디오 영상으로부터 문자의 추출과 탐지에 우수한 성능을 나타내고 있음을 보여주며, 또한 제안한 방법은 디지털 비디오 색인에 충분히 사용되어 질 수 있음을 보여준다.

본 논문에서 제안한 방법은 영화, 광고 방송, 다큐멘터리 그리고 그 외의 다양한 영상에서도 우수한 성능을 나타냈다. 앞으로의 연구는 본 알고리즘을 책 표지, 컴팩트 디스크 표지, 그리고 거리의 풍경 영상등에서도 유연한 결과를 보여 줄 수 있도록 지속적으로 연구 할 것이다.

5.참고 문헌

[1] Jae-Chang Shim, Chitra Dorai, Ruud Bolle, " Automatic Text Extraction from Video for Content-Based Annotation and Retrieval," Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, On page(s): 618 - 620 vol.1 16-20 Aug. 1998

[2] S. Antani, D. Crandall, R. Kasturi, " Robust Extraction of Text in Video," Pattern Recognition, 2000. Proceedings. 15th International Conference on, On page(s): 831 - 834 vol.1 3-7 Sept.

[3] H.Kuwano, Y.Taniguchi, H.Arai, M.Mori, S.Kurakake, H.Kojima, " Telop-on-demand: video structuring and retrieval based on text recognition," Multimedia and Expo, 2000. ICME 2000. 2000 IEEE International Conference on, On page(s): 759 - 762 vol.2 30 July-2 Aug. 2000

[4] U. Gargi, S. Antani, R. Kasturi, " Indexing text events in digital video databases," Pattern Recognition, 1998. Proceedings. Fourteenth International Conference on, On page(s): 916 - 918 vol.1 16-20 Aug. 1998

[5] Sameer Antani, Ullas Gargi, David Crandall, Tarak Gandhi and Rangachar Kasturi, " Extraction of Text in Video," Dept. of Comput. Sci. & Eng., Pennsylvania State Univ., Technical Report, CSE-99-016, August 30, 1999

[6] S. Messelodi and C.M. Modena, " Automatic identification and skew estimation of text lines in real scene images," Pattern Recognition, Vol. 32 (5) (1999) pp. 791-810

(a)일반 뉴스 영상(YTN) (b)추출된 자막영상
(c)스포츠 뉴스 영상(MBC) (d)추출된 자막(배경포함)영상
(e)일반 뉴스 영상(CNN) (f)추출된 자막영상
(g)LeBourgeois 결과[5] (h)본 논문 결과

그림4. 문자 추출의 결과