뉴스 동영상의 자동 색인화를 위한 앵커 장면 검출

신정교, 류한진, 김정림, 이현진, 설상훈
고려대학교 전자컴퓨터공학과
{jkshin, hanjin, jrkim, hjlee, sull}@mpeg.korea.ac.kr

Anchor Shot Detection for Automatic Indexing in News Video

Jung-Kyo Shin, Han-Jin Ryu, Jung-Rim Kim, Hyun Jin Lee, Sanghoon Sull
Department of Electronics and Computer Engineering, Korea University

요 약

본 논문은 뉴스 동영상의 검색을 위한 내용 기반 자동 색인화 방법을 제시한다. 제안한 방법은 장면전환 검출, 앵커 후보 장면 검출과 앵커 장면 검출의 3 단계로 구성된다. 첫 번째 단계에서, 장면전환 검출 알고리즘에 의해 검출된 대표 프레임을 대상으로 얼굴을 추출한다. 두 번째 단계에서는 추출된 얼굴의 공간적 위치와 크기를 고려하여 앵커 장면 후보들을 결정하고, 마지막 단계에서, 앵커 장면 후보에서 HSV 컬러공간의 색상정보인 Hue 값을 특징벡터로 집합론 방법(Set-theoretic approach)에 기반한 군집화 알고리즘을 적용하여 최종 앵커 장면을 검출한다. 다양한 실험영상을 통한 실험결과 제안한 방법은 앵커의 위치와 수에 상관없이 정확한 검출이 가능하였으며, 평균 검출율(Precision)은 0.97, 회수율(Recall)은 0.98 의 우수한 성능을 보였다.

I. 서론

지상과, 케이블, DMB(Digital Multimedia Broadcasting)와 인터넷 등 동영상 컨텐츠가 급격하게 증가하고 많은 양의 컨텐츠를 쉽게 접할 수 있게 됨에 따라 동영상에서 원하는 정보를 빠르게 얻기 위한 검색 서비스에 대한 관심도 점점 커지고 있다. 특히, 뉴스의 경우 각 방송사와 포털 사이트에서 인터넷을 통해 기사와 함께 볼 수 있는 서비스를 제공하고 있어, 사용자는 원하는 내용과 관련된 정보를 쉽게 찾아 볼 수 있게 되었다. 이러한 서비스를 제공하기 위해서는 동영상의 내용에 기반한 색인화 작업이 필수적이며, 자동화 된 색인화 작업은 효율적인 서비스 제공을 위해 더욱 중요하다.

뉴스 동영상은 구조적 특성상, 뉴스 기사 단위로 구성되어 있으며, 기사내용이 방송되기 전에는 항상 앵커가 등장하여 대략적인 기사내용을 소개한다. 이와 같은 뉴스 동영상의 구조적 특성으로 인해 앵커 장면검출은 뉴스의 색인화 방법에 있어서 우선적으로 고려하여야 한다.

이러한 요구로 최근 다양한 앵커 장면 검출 방법이 연구 되고 있다[1-3]. 앵커 장면에서 객체의 시각적인 특성을 이용하는 방법으로 주로 앵커의 위치를 한정시키고 앵커 옆에 위치하는 뉴스 제목이나, 변하지 않는 앵커의 배경 등과 같은 특성을 이용하였다[1-2]. 이러한 방법들은 뉴스 중에 앵커의 위치가 바뀌거나 앵커가 둘 이상인 경우, 앵커 장면에서 뉴스기사 제목이 나오지 않는 경우에는 검출하지 못하는 단점과 함께, 인터뷰나 기자회견 등 앵커 장면과 비슷한 특성을 갖는 장면이 잘 못 검출되는 단점이 있다. 또한 얼굴과 텍스트 검출알고리즘을 함께 사용함으로써 처리 과정이 복잡할 뿐만 아니라 처리 시간이 많이 걸리므로 자동화 색인 및 실시간 처리에 적합하지 않다. 다른 방법으로는 앵커장면 자체를 하나의 템플릿으로 구성하여 동일한 프로그램은 같은 구성과 배경을 갖는다는 전제하에 앵커 장면의 정보를 데이터 베이스에 미리 저장해 놓고 색인화하고자 하는 동영상과 비교하여 앵커 장면을 찾아내는 방법이 있다[3]. 그러나 이 방법은 프로그램에 대한 정보를 미리 가지고 있어야 하므로 데이터 베이스 입력을 위한 선행 작업이 필요하며, 만약 프로그램의 배경이나 구성이 바뀌게 되면 그와 같은 선행작업을 반복해야 하는 단점이 있다.

본 논문에서는 먼저 장면전환 검출 알고리즘을 이용하여 각 장면의 대표 프레임을 추출하고, 추출된 대표 프레임에 얼굴검출 알고리즘을 적용하여 검출된 앵커 위치와 크기를 기준으로 앵커 장면의 후보 프레임들을 선정한다. 이렇게 선정된 앵커 장면후보 프레임에서 HSV 컬러공간의 색상정보인 Hue 값을 이용한 히스토그램으로 집합론 방법(Set-theoretic approach)[4]을 적용하여 앵커 장면을 검출한다.

본 논문의 구성은 다음과 같다. II 장에서는 앵커 장면 검출 시스템의 일련의 과정을 설명하고 III 장에서는 본 논문에서 제안하는 앵커 장면의 특성을 적용하여 앵커 장면 후보를 구하는 방법과 앵커 장면 후보에서 집합론 방법을 적용한 앵커 장면 검출 방법을 설명한다. 그리고 IV 장에서는 제안된 방법의 실험 결과를 제시하고, V 장에서 결론 및 향후 과제를 논의한다.

II. 시스템 개요

그림 1 은 본 논문에서 제안하는 앵커 장면 검출 시스템을 나타내는 것으로써 앵커 장면의 검출 결과를 실시간으로 확인할 수 있도록 구현하였다.

먼저 히스토그램의 차이를 이용한 장면전환 검출 알고리즘을 이용하여 대표 프레임을 검출하고, 대표 프레임을 대상으로 얼굴 검출을 한다. 검출된 얼굴의 위치와 크기를 기준으로 앵커 장면 후보를 결정한다. 앵커 장면 후보로 선정된 대표 프레임에서 HSV 컬러공간의 색상정보를 특징 벡터로 추출하고 집합론 방법을 적용하여 후보 장면들 간의 군집화를 통해 최종 앵커 장면을 검출한다.