

# 뉴스 비디오에서 키 프레임 추출과 앵커 인식

신성윤\*, 임정훈\*, 표성배\*\*, 이양원\*  
\* 군산대학교 컴퓨터정보과학과  
\* 인덕대학 소프트웨어개발과

## The Key Frame Extraction and Anchor Recognition in News Videos

Seong-Yoon Shin\*, Jeong-Hoon Leem\*, Sung-Bae Pyo\*\*, Yang-Won Rhee\*  
\* Dept. of Computer Information Science, Kunsan Nat'l University

E-mail : yshin@cs.kunsan.ac.kr

\*\* Dept. of Software development, Induk Institute of Technology

E-mail : pyosb@mail.induk.ac.kr

### 요약

뉴스 비디오에서 앵커가 등장하는 첫 번째 프레임은 하나의 뉴스를 시작으로 설정하는데 기준이 되는 키 프레임이라고 볼 수 있다. 본 논문에서는 뉴스 비디오의 장면 전환을 검출을 위하여 컬러 히스토그램과  $\chi^2$  히스토그램을 합성한 방법을 이용하여 키 프레임을 추출하며, 추출된 키 프레임을 대상으로 앵커 프레임의 공간적 구성과 얼굴의 특징 정보에 대한 사전 지식을 바탕으로 한 유사성 측정을 통하여 앵커를 인식하도록 한다. 앵커로 인식된 프레임은 하나의 뉴스 신에 대한 키 프레임이 되며 뉴스 비디오를 색인화 하는데 중요한 역할을 수행한다.

### 1. 서론

비디오(video) 분석과 모델링(modeling)을 이용한 내용 기반 검색은 대용량의 정보를 짧은 시간에 검색하여 함축적으로 재생하여 줌으로써 인간에게 편의성과 유용성을 제공한다.[1,2].

멀티미디어 데이터(multimedia data) 중에서 뉴스 비디오(news video)의 분석과 모델링 분야는 뉴스 비디오가 갖는 시대적인 가치와 한 시대 또는 한 국가의 정치, 경제, 사회, 문화 전반에 걸친 상황 분석 및 역사의 흐름에 따른 향후 대처 방안을 제시하는 중요한 의미를 갖기 때문에 수많은 연구가 활발히 진행되고 있다[1,2].

컴퓨터 시스템을 이용하여 인간의 얼굴을 분석하는 방법은 다양한 형태로 활발하게 연구가 수행되어 왔다. 최근에는 인간의 얼굴을 분석하기 위하여 얼굴의 구성요소 각각들 분리하여 처리하는 방법[3,4,5,6]과 얼굴의 다양한 특징들을 추출하여 사용하는 방법[7,8,9], 그리고 얼굴의 고유 공간 분석을 통한 고유 공간 모델링 방법[10,11,12] 등 많은 향상된 연구들이 진행되고 있다.

본 논문에서는 뉴스 비디오를 인덱싱(indexing) 하기 위하여 뉴스 신(scene)의 정확한 키 프레임(key frame)을 추출하는 효율적인 장면 전환 검출 방법을 이용하고 추출된 키 프레임을 바탕으로 실제 앵커가 등장하는 앵커 프레임(anchor frame)을 추출하도록 한다. 전반적인 시스템 구조는 그림 1과 같다

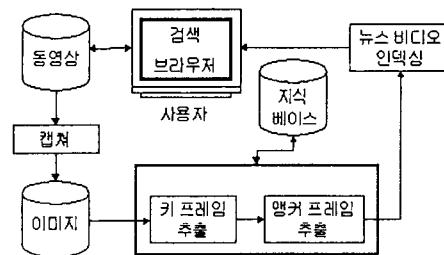


그림 1. 전반적인 시스템 구조

본 논문에서는 [1,2]에서 제안한 컬러 히스토그램(color histogram)과  $\chi^2$  히스토그램의 장점을 조합한 방법을 사용하여 먼저 비디오 샷들의 키 프레임들을 추출한다. 다음으로, 뉴스 비디오가 앵커 프레임을 중

심으로 하나의 신을 형성하는 구조적인 특성을 이용하여, 이렇게 추출된 샷들의 키 프레임들에서 앵커 프레임들을 추출하기 위하여 그림 2와 같이 앵커 인식 과정을 수행한다.

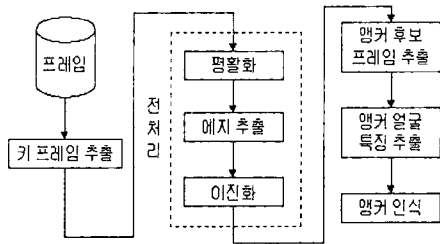


그림 2. 앵커 인식 과정

본 논문의 구성은 2장에서 장면 전환에 의한 키 프레임을 추출하는 방법에 대해 설명하고, 3장에서 앵커 후보 프레임을 추출하는 방법을 설명하며, 4장에서는 앵커의 얼굴을 인식하는 방법을 제시한다. 그리고, 5장에서는 실제 데이터를 이용한 실험을 수행하고, 6장에서는 결론을 맺도록 한다.

## 2. 장면 전환 검출

본 논문에서는 아래의 식과 같이 컬러 히스토그램과  $\chi^2$  히스토그램의 장점을 조합한 방법을 사용하여 프레임간의 차이 값을 계산함으로써 장면 전환을 검출하고 키 프레임을 추출한다[1,2].

$$d(I_i, I_j) = \sum_{k=1}^n \left( \frac{(H_i^r(k) - H_j^r(k))^2}{H_i^r(k)} \times 0.299 + \frac{(H_i^g(k) - H_j^g(k))^2}{H_i^g(k)} \times 0.587 + \frac{(H_i^b(k) - H_j^b(k))^2}{H_i^b(k)} \times 0.114 \right) / 3$$

이렇게 추출된 키 프레임들은 뉴스 비디오의 첫 번째 프레임과 각 샷들의 첫 번째 프레임들이 되며 이 키 프레임들은 앵커 프레임과 기타 다른 프레임들로 구성되는데 앵커 프레임을 추출하기 위하여 정확한 앵커 인식 과정이 수행되어야 한다.

뉴스 비디오에서 앵커 프레임은 하나의 뉴스 신의 시작을 나타내는 중요한 키 프레임이 되기 때문에 뉴스 비디오 물리적 또는 논리적인 인덱싱을 수행하는데 매우 중요한 역할을 수행한다.

## 3. 앵커 후보 프레임 추출

효율적인 앵커 후보 프레임의 추출을 위하여 앵커 프레임들이 갖는 사전 지식을 바탕으로 평활화(equalization), 에지(edge) 추출, Negative, 이진화의 전처리 과정을 수행한 후 앵커 프레임이 픽셀 특징을 바탕으로 앵커 후보 프레임들을 추출하도록 한다.

### 3.1 전처리

#### (1) 평활화

평활화는 추출된 키 프레임들이 갖는 다양한 픽셀 분포를 완만한 픽셀 분포로 변화시키는 작업으로서 각 픽셀들이 주변 값들과 유사하게 구성되고 복잡한 에지들이 줄어들게 된다.

#### (2) 에지 추출

에지 추출은 평활화가 수행된 프레임 내의 각 객체들의 경계가 되는 부분을 추출하는 작업으로서 이웃하는 픽셀들간의 차이 값 정도에 따라서 추출한다. 3 X 3 Window를 사용하여 중앙 픽셀의 RGB 값의 평균과 이웃하는 8개의 픽셀 각각의 RGB 값의 평균과의 차이가 최대인 값에 일정한 임계치를 더하여 중앙의 픽셀로 대치한다. 이는 객체들간의 경계선을 강조하고 이진화를 용이하게 만든다.

#### (3) Negative와 이진화

처리상의 편의를 위하여 흑과 백의 픽셀값을 역으로 변환시켜 주는 Negative와, 각 픽셀 값에 임계치를 적용하여 각각의 모든 픽셀 값들이 0 또는 255를 갖도록 하는 작업이 이진화를 수행한다.

#### (4) 사전 지식 획득

이렇게 전처리 과정을 거친 프레임들의 총 픽셀 수를 계산하여 앵커 프레임이 갖는 총 픽셀 수의 사전 지식을 얻도록 한다.

### 3.2 앵커 후보프레임 추출

에지를 추출하고 이진화 하여 얻은 키 프레임들 중에서 앵커가 존재하는 프레임의 에지 픽셀 수가 다른 비 앵커 프레임들이 갖는 에지 픽셀 수와는 달리 일정 영역 내에 있다는 사전 지식을 실험을 통하여 얻을 수 있었다. 이렇게 추출된 사전 지식과 입력된 프레임의 픽셀 수를 비교하여 앵커 후보 프레임을 추출한다. 추출된 앵커 후보 프레임은 앵커 프레임, 앵커 프레임과 유사한 비 앵커 프레임, 그리고 앵커 프레임과 픽셀 수가 비슷한 비 앵커 프레임들을 포함하고 있기 때문에 보다 정확한 앵커 인식 방법이 필요하다.

#### 4. 앵커 인식

앵커 프레임의 추출은 먼저 뉴스 비디오에서 앵커가 존재하는 프레임의 구조적인 특성을 분석하여 앵커의 얼굴 영역에 대한 사전 지식을 설정한 후 이를 바탕으로 앵커의 얼굴을 인식하여 정확히 앵커로 인식된 프레임들 신의 키 프레임으로 설정한다.

앵커 프레임에 대하여 다음과 같은 사전 지식을 얻을 수 있으며 그림 3은 앵커 프레임이 갖는 공간적 구성을 나타낸다.

- (1) 뉴스 비디오의 첫 번째 프레임은 항상 앵커로 시작한다.
- (2) 뉴스 비디오에서 앵커가 한 명으로 시작했으면 끝날 때도 한 명이다.
- (3) 뉴스 비디오에서 앵커가 두 명으로 시작했으면 끝날 때도 두 명이다.
- (4) 뉴스 비디오에서 앵커가 한 명이고 좌측에 있으면 반드시 우측엔 뉴스 아이콘이 존재한다.
- (5) 뉴스 비디오에서 앵커가 한 명이고 우측에 있으면 반드시 좌측엔 뉴스 아이콘이 존재한다.
- (6) 뉴스 비디오에서 앵커가 한 명이고 중앙에 있으면 뉴스 아이콘은 없다.
- (7) 뉴스 비디오에서 앵커가 두 명이면 좌우대칭을 이룬다.

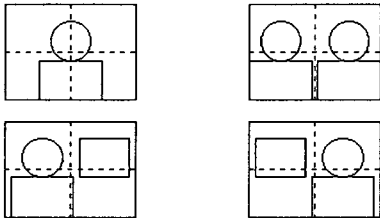


그림 3. 앵커 프레임들의 공간적 구성

앵커의 인식 과정은 앵커의 특징 추출, 앵커 얼굴 영역 추출, 그리고 추출된 특징들을 바탕으로 앵커의 얼굴을 인식하는 단계로 구성된다.

앵커의 특징 추출은 앵커 얼굴의 영역에 대한 사전 지식과 앵커 얼굴 특징들의 위치 정보와 기하학적인 특성을 이용하여 앵커 후보 프레임으로 추출된 프레임들을 대상으로 추출한다.

먼저, 사전 지식에 의하여 앵커의 얼굴 영역에 대한 히스토그램을 작성하여 이를 바탕으로 양 눈, 코, 그리고 입의 위치를 추출한다. 그리고 이렇게 추출된 특징 정보들의 기하학적인 특성을 이용하여 특징 정보들 사이의 거리와 각을 계산한다. 앵커 인식을 위해

측정되는 값들은 양 눈 사이의 거리, 왼쪽 눈과 코의 거리, 왼쪽 눈과 입의 거리, 코와 입의 거리, 그리고 양 눈과 코의 각도이다. 계산된 값들은 사전 지식베이스에 저장된 각각의 거리와 각의 값들과의 비교를 통하여 각각의 앵커를 인식한다.

앵커로 인식된 프레임들은 뉴스 신의 키 프레임으로 설정하고 앵커로 인식되지 않은 프레임들은 이전에 키 프레임 추출 과정을 통해 얻어진 것과 같이 샷들의 키 프레임이 된다. 앵커로 인식되어 추출된 앵커 프레임은 뉴스 비디오가 앵커 프레임을 시작으로 하여 하나의 뉴스 신을 구성한다는 구조적인 특성을 이용하여 물리적으로 또는 다양한 논리적 형태로 인덱싱 되어 효율적인 검색을 수행할 수 있다.

#### 5. 실험

실험에는 펜티엄 800MHz PC와 비디오 플레이어 시스템에서 윈도우 98 환경의 MS Visual C++ 6.0을 사용하였으며 실험 데이터는 한국방송공사에서 방영한 저녁 9시 뉴스를 녹화하여 비디오 플레이어로 구동시킨 후, 오스카II 캡처 보드를 이용하여 AVI 파일로 데이터를 획득하였다. 이 데이터에서 초당 5 프레임을 실험 데이터로 캡처 한 후 프레임 크기를 460 X 370으로 정규화 하여 실험에 사용하였다.

본 논문에서 이용한 키 프레임 추출 방법을 사용하는 키 프레임을 추출하는 과정은 그림 4와 같으며 그림 5는 추출된 얼굴 영역을 나타낸다.

전체적인 실험 결과로서, 제안된 키 프레임 추출 방법을 통하여 얻은 200개의 키 프레임들 중에서 앵커 후보 프레임으로 12개의 프레임을 추출하였고 앵커 프레임으로 7개의 프레임을 얻을 수 있었는데 이는 실제 뉴스 비디오에서 정확히 앵커가 존재하는 신의 수와 일치함을 알 수 있었다.

실험을 통한 결과로서, 뉴스 비디오에서 뉴스를 진행하는 앵커의 인식은 앵커가 갖는 특징 정보와 이들 사이의 거리와 각의 측정값을 이용하여 완벽하게 인식할 수 있었다. 이는 뉴스 비디오의 특성상 앵커의 위치가 거의 일정하며 앵커의 수가 국한되어 있다는 사실로 인하여 얻은 결과이다.

이와 같이 추출된 앵커 프레임은 물리적으로 키 프레임 번호를 부여하여 인덱싱이 가능하며, 사용자의 요구에 따라 다양한 분야별 또는 일자별로 인덱싱이 가능하므로 사용자의 요구 사항에 맞게 검색 할 수 있으며, 또한 편집도 가능하다.

## 6. 결 론

본 논문에서는 뉴스 비디오의 장면 전환 검출을 위하여 컬러 히스토그램과  $\chi^2$  히스토그램을 조합한 방법을 사용하여 보다 효율적으로 키 프레임들을 추출하였고, 앵커 프레임에 대한 사전 지식을 바탕으로 보다 간편하고 정확한 앵커 프레임의 특징들을 측정하여 앵커프레임 추출하였다.

향후 보다 효율적인 장면 전환 검출 방법과 캡션을 분석하는 방법의 이용과 다른 분야의 장면 전환 검출에 효율적인 새로운 알고리즘의 개발이 필요하다고 사료된다.

## [참 고 문 헌]

- [1] 이동섭, 김재원, 배석찬, 이양원, "비디오 장면변환 검출 알고리즘", 『한국퍼지 및 지능시스템학회 '98 춘계학술대회논문집』, 제8권, 제1호, 1998.6.
- [2] 이동섭, 이양원, "뉴스 비디오의 장면분할을 위한 키 프레임 추출기법", 『한국정보과학회 춘계 학술 발표논문집』, Vol. 25, No. 1, pp. 613-615, 1998.
- [3] B. S. Manjunath, R. Chellappa and C. v.d. Malsburg, "A Feature Based Approach to face Recognition", *Proc. IEEE Computer Society Conf on Computer Vision and Pattern Recognition*, pp. 373-378, 1992
- [4] R. Brunelli and T. Poggio, "Face Recognition through Geometrical Features", *Proc. ECCV'92*, pp. 792-800, <http://hera.itc.it:3003/~brunelli/Papers.html>, 1992
- [5] Y. Cheng, K. Liu, J. Yang and H. Wang, "A Robust Algebraic Method for Human Face recognition", *ICPR'92*, pp. 221-224, 1992
- [6] Y. Yacoob and L. S. Davis, "Computing Suptio-Temporal representation of Human Faces", in *Proc. IEEE Computer Society Conf on Computer Vision and Pattern recognition*, pp. 70-75, 1994
- [7] Z. Hong, "Algebraic Feature Extraction of Image for Recognition", *Pattern Recognition*, Vol. 24, pp. 211-219, 1991
- [8] Alan L. Yuille, Peter W. Hallinan and David S. Cohen, "Feature Extraction from Face Using Deformable Templates", *ICCV'92*, pp. 99-111, 1992
- [9] I. Craw, H. Ellis and J. Lishman, "Automatic

Extraction of Face Features", *Pattern Recognition Letters*, Vol. 5, pp. 183-187, 1987

- [10] Matthew Turk and Alex Oentland, "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, Vol. 3, No. 1, 1991
- [11] Pentland A., Starner T., Etcoff N., Masoiu A., Oliyide O. and Turk M."Experiments with Eigenfaces", *IJCAI'93*, MIT Media Lab. Tr#-194, 1992
- [12] 김경수, 최형원, "효과적인 얼굴 영상 분류를 위한 컬러 고유 공간 분석", 『한국퍼지 및 지능시스템학회 춘계학술대회논문집』, pp. 195-200, 1997

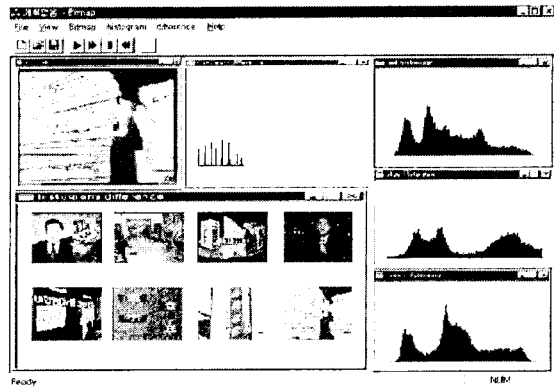


그림 4. 키 프레임 추출 장면

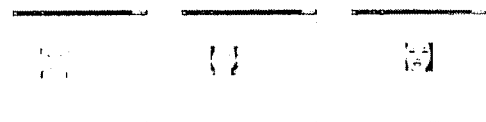


그림 5. 얼굴 영역 추출