

효율적인 데이터 관리를 위한 내용기반 뉴스 비디오 검색 시스템 구현

남윤성*, 양동일*, 배종식**, 최형진*
*강원대학교 컴퓨터학과
**한림성심대학 문화컨텐츠개발과
e-mail:nyspeter@kangwon.ac.kr

Implementation of Content-based News Video Retrieval System for Efficient Video Data Management

Yun-Seong Nam*, Dong-Il Yang*, Jong-sik Bae**, Hyung-Jin Choi*
*Dept. of Computer Science, Kang-Won University
**Dep. of Culture Contents Development, Hallym College

요 약

뉴스 데이터를 구조적으로 분할하고 의미적으로 분류하여 내용별로 세분화하여 검색하는 방법을 제안한다. 구조적 분할은 공간 밝기 분포와 명암도의 불연속성 그리고 시간적인 관계 등 프레임간의 상관 정보를 이용하여 장면을 분할한다. 의미적 분류는 키 프레임에서 추출된 특징 정보를 사전 지식 정보와 비교하여 뉴스 비디오의 세부 내용을 기사별로 분류한다.

뉴스의 진행이 앵커 프레임을 중심으로 주기적으로 반복된다는 특징을 이용하여 앵커 장면과 비 앵커 장면으로 기사를 분류한다. 비 앵커 장면은 연설장면, 인터뷰장면, 일반 장면으로 세분화하고 기사별로 분류하여 검색하도록 한다. 또한 뉴스 아이콘에 의한 요약 검색 기능 그리고 자막 통합 처리에 의한 자막 검색을 하여 뉴스 비디오를 내용별로 분류하고 인덱싱하여 신속하게 뉴스 비디오를 검색할 수 있도록 설계한다.

1. 서론

초고속 인터넷 통신망 및 멀티미디어 관련 기술의 발달로 인하여 멀티미디어 정보를 보다 효과적으로 처리하기 위한 전송, 저장, 관리 및 검색이 중요한 핵심 기술로 대두되고 있다. 영상과 같은 멀티미디어 데이터는 정지 화상이나 텍스트 데이터에 비해 그 양이 방대하고 내포된 정보가 훨씬 많기 때문에 사용자가 원하는 정보를 찾는 데 많은 시간과 노력이 필요하며 컴퓨터를 이용하여 자료를 처리하는 시간도 많이 소요된다[1,2].

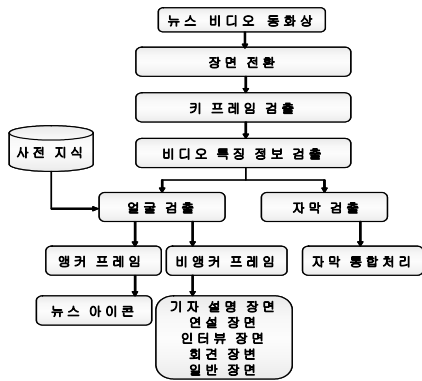
이러한 영상 중에서도 특히, 뉴스 비디오는 사회 전반의 중요한 정보를 함축적으로 짧은 시간에 전달하는 것으로 오디오와 비디오 정보가 복합된 멀티미디어의 특징을 가장 잘 나타내고 있다[3]. 비디오 데이터를 효과적으로 검색하기 위하여 장면 변화 현상

에 따라 비디오 데이터를 분할하고 색인하며 검색할 수 있어야 한다.

본 논문에서는 효과적인 시스템을 구축하기 위하여 비디오 데이터의 생성과 구성에 관한 사전 지식을 이용하여 의미 정보와 특징 정보를 추출한다. 이를 바탕으로 뉴스 비디오를 내용별로 분류하고 인덱싱하여 신속하게 뉴스 비디오를 검색한다.

2. 내용기반 분류에 의한 뉴스 비디오 검색 시스템 설계

본 논문에서는 내용기반 분류에 의한 비디오 검색을 위하여 (그림 1)과 같이 시스템 구조를 설계한다.



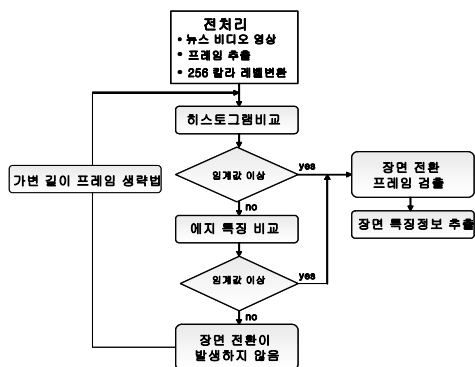
(그림 1) 뉴스 비디오 검색 시스템의 구조

(1) 장면 전환 추출 및 키 프레임 추출

뉴스 비디오의 장면 분류를 위해 내용적으로 앵커 장면, 기자 설명 장면이나 연설 장면, 회견 장면이나 인터뷰 장면, 일반 장면으로 분류한다. 각각의 장면들은 특징적으로 발생하는 색상과 에지 성분이 각각 다르기 때문에 칼라 히스토그램 정보와 에지 특징 정보를 이용하여 장면들을 분류한다.

비디오 장면 분류 후 각 샷으로부터 키 프레임을 추출한다. 키 프레임은 각 샷을 대표하는 프레임으로 장면 변화가 크게 발생하는 지점의 프레임을 추출한다. 키 프레임은 비디오 특징 정보 추출을 위한 기본 프레임으로 사용되며 브라우저에서 내용별 검색을 위한 인덱싱 프레임으로 사용된다.

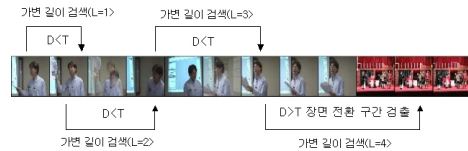
장면 전환 프레임을 추출하기 위해서 프레임의 특징을 비교하는 칼라 히스토그램 정보 비교법과 에지 특징 정보 비교법을 이용한다. 이러한 특징을 추출하기 위해서는 일련의 전처리 과정이 필요하며 이는 검색 후 데이터베이스에 저장되는 동화상의 조건을 같게하여 일반화된 특징을 추출하기 위함이다. 다음 (그림 2)는 전체적인 장면 특징 추출과정을 나타내고 있다.



(그림 2) 장면 전환 추출 과정

(2) 가변 길이 프레임 생략법

(그림 3)은 가변 길이 프레임 생략법을 나타낸 것으로 처음에는 가변 길이 초기값을 1로 하여 제일 처음 프레임과 그 다음 프레임의 히스토그램 차인 D를 구하여 임계값과 비교한다. 임계값 T보다 히스토그램 차인 D가 작으면 가중치 W를 증가시켜 생략할 프레임 수를 2, 3, 4 등으로 늘려 히스토그램 차를 비교한다.



(그림 3) 가변 길이 프레임 생략법

(3) 얼굴 프레임 추출

얼굴 프레임을 추출하기 위해 먼저 RGB 영상을 HSV 칼라 공간으로 변환시키고 색채 요소(H,S)와 명암 요소(V)를 분리하여 명암 요소를 제거함으로써 조명 변화의 영향을 줄인다. HSV 요소 중에서 색상 요소(H)와 채도 요소(S)만을 사용하여 각 화소의 칼라 값이 <식 1>의 얼굴 색상 범위에 존재하면 얼굴 영상 후보로 인식한다.

$$B(X, Y) = \begin{cases} 1 & \text{if } ((30 \leq H \leq 60) \text{ and } (20.0 \leq S \leq 35.0)) \\ 0 & \text{otherwise} \end{cases} \quad \text{<식 1>}$$

위의 식을 통해 영상에서 피부색에 해당하는 영역을 모두 찾아서 모폴로지(Morphology) 제거(opening) 연산을 수행하여 잡음 요소를 없애고 얼굴 영역을 레이블링 처리하여 얼굴 영역으로 연결되지 않는 화소를 분리한다.

프레임에서 이와 같은 과정을 통해 피부색으로 변환된 영역이 존재하면 얼굴 프레임으로 그렇지 않은 경우에는 얼굴을 포함하지 않은 프레임으로 분류한다.

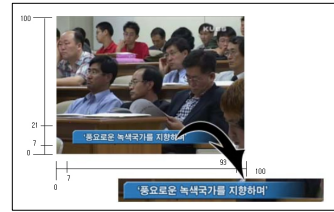
(4) 앵커 장면 추출

뉴스 비디오에서 앵커가 존재하는 프레임의 구조적인 특징을 분석하여 다음과 같이 앵커 추출에 대한 사전 지식을 설정한 후 이를 바탕으로 앵커 장면을 추출하도록 한다.

- ① 키 프레임 특징 정보에서 얼굴의 수를 2 이하로 한다.
- ② 일정한 시간 동안 유지되며 주기적으로 반복해서 나타난다.
- ③ 배경 화면의 영상은 거의 변화가 없다.
- ④ 앵커가 두 명이면 좌우 대칭을 형성한다.

- ⑤ 앵커가 한 명이면 좌측, 중앙, 우측에 나타난다.
- ⑥ 앵커 수에 따라서 얼굴 위치가 고정되어 있다.
- ⑦ 얼굴 면적은 최소 크기 이상이다.
- ⑧ 얼굴 크기는 일정하다.
- ⑨ 얼굴 위치는 거의 고정되어 있다.
- ⑩ 자막이 나타난다.

쪽에서 7%~21% 사이의 영역을 자막 영역으로 추출한다.



(그림 4) 자막 후보 영역추출

(5) 비 앵커 장면 추출

얼굴이 포함되어 있는 장면 중에서 앵커 장면의 속성이 아닌 장면을 비 앵커 장면이라고 한다.

비 앵커 장면은 앵커 장면과 구분되는 장면 전환 특징을 가지고 있다. 비 앵커 장면은 일반적으로 앵커 장면 보다 빠르게 진행되기 때문에 배경이나 영상의 변화가 많은 특징을 가지고 있다.

비 앵커 장면 추출은 키 프레임의 특징 정보에서 얼굴 형태 정보를 중심으로 장면을 구분한다. 먼저 형태 정보에서 얼굴 수를 중심으로 비 앵커 장면을 세분화한다. 얼굴이 하나일 경우에는 설명 장면이나 연설 장면으로 분류하며, 얼굴이 두 개일 경우에는 인터뷰 장면으로 분류한다. 얼굴이 세 개 이상일 경우에는 일반 기사 설명 장면으로 분류한다.

(6) 뉴스 아이콘 추출

뉴스 아이콘 추출은 앵커 장면 추출 과정을 통해 추출된 앵커 키 프레임에 대해 횡과 종에 대한 에지 히스토그램을 구하면 뉴스 아이콘이 위치하는 곳에서 갑자기 높은 값으로 나타나며, 그 크기도 비슷하다. 그 이유는 뉴스 아이콘은 대부분 동일한 직사각형 모양을 하고 있기 때문이다.

(7) 자막 통합 처리

앵커 장면 추출 지점에서부터 다음 앵커 장면 전까지 자막이 가지는 특징을 이용하여 자막 영역을 추출한 후, 색 단순화 방법으로 자막의 유무를 판단하여 통합 처리한다.

① 자막 후보 영역 추출

자막의 유무를 판단하기 위해 (그림 4)와 같이 추출된 키 프레임에서 자막 후보 영역을 추출하여 자막 영역에 대해서만 자막의 유무를 판단한다.

자막 후보 영역은 추출된 키 프레임의 가로축과 세로축을 각각 100%로 하였을 경우, 가로축은 양쪽 끝에서 7%를 제외한 중앙 86% 부분, 세로축은 아래

② 색 단순화

자막의 특징상 자막의 배경과 자막의 색은 다른 영역에 비해 단순한 색을 가지고 있다. 따라서 자막의 특징을 두드러지게 하기 위해 자막 영역 부분의 색상을 보다 단순화 하여 자막의 특징을 높인다.

3. 실험 결과 및 성능 평가

실험에 사용한 비디오 데이터는 AVI 형식 파일이며 320×240 크기의 256 칼라 레벨이고 초당 30 프레임씩 재생되는 KBS 9시 뉴스를 사용하였다.

본 논문에서 제안한 뉴스 비디오 검색 시스템의 성능을 평가하기 위하여 <식 2>인 Recall과<식 3>인 Precision을 사용하였다.

Recall은 추출하고자 하는 내용의 개수 중에서 몇 개가 정확하게 추출되었는지를 나타내는 비율을 의미하며, Precision은 실제 추출되어진 내용 중에서 몇 개가 정확하게 추출되었는지를 나타내는 비율이다[4].

$$Recall = \frac{C_n}{C_n + M_n} \quad \text{<식 2>}$$

$$Precision = \frac{C_n}{C_n + F_n} \quad \text{<식 3>}$$

<식 3>과 <식 4>에서 C_n 은 정확히 추출된 개수, M_n 은 미추출된 개수, F_n 은 오추출된 개수를 나타낸다.

실험에 사용된 뉴스 비디오 데이터는 전체 프레임 수가 35,900 여 프레임이며, 장면 전환 프레임 수는 203 프레임이며, 앵커 프레임 수는 16 프레임이다. 비 앵커 프레임 수는 131 프레임이며, 그 중에서 기자 설명 프레임 수는 7 프레임이며, 인터뷰 프레임 수는 32 프레임이며, 연설 프레임 수는 5 프레임이며 일반 장면 프레임 수는 87 프레임이다. 뉴스

아이콘 수는 14 프레임이며 자막 프레임 수는 42 프레임으로 구성되어 있다. 다음 (그림 5)는 추출된 장면들이다.



[표 1] 실험 평가

분 류		Recall	Precision
장면 전환 프레임 수		93%	97%
앵커 프레임 수	1 명	100%	88%
	2 명	100%	100%
기자 설명 프레임 수		86%	75%
인터뷰 프레임 수		87%	87%
연설 프레임 수		100%	71%
일반장면 프레임 수		95%	94%
뉴스 아이콘 수		100%	88%
자막 프레임 수		95%	85%

실험 결과에서 장면 전환 추출과 앵커 프레임 추출은 우수한 추출률을 보였다. 이러한 실험 결과는 정형적인 비디오 뉴스의 생성 구조 및 사전 지식 설정이 잘 되었음을 나타낸다. 기자 설명 장면 추출과 인터뷰 장면 추출은 검출률이 낮았다. 그 이유는 비디오 뉴스 데이터에 잡음이 포함되어 있어 제대로 얼굴을 추출하지 못했기 때문이다. 연설 프레임 장면은 얼굴 추출 알고리즘이 완벽하지 않았기 때문에 정확성이 떨어졌으며, 자막 추출은 자막의 색이 점진적으로 변화하여 색 단순화 알고리즘이 자막인식을 제대로 하지 못했기 때문에 발생하였다.

4. 결론 및 향후연구 방향

본 논문에서는 뉴스 비디오의 주요 내용을 신속하고 정확하게 검색하기 위하여 뉴스 데이터를 의미

적으로 분류하고 구조적으로 분할하였다. 의미적 분류는 비디오 데이터의 생성과 구성에 관한 사전 지식을 설정하고 조건을 부가하여 검색하도록 하였으며 구조적 분할은 공간 밝기 분포, 명암도의 불연속성 및 유사성 그리고 시간적인 관계 등의 프레임간의 상관 정보를 이용하여 분할하였다. 이와 같은 뉴스 비디오 검색 개념을 바탕으로 사용자에게 내용에 기반한 뉴스 검색이 가능하도록 하였으며 방대한 분량의 뉴스 비디오에서 특정한 정보를 신속하고 정확하게 검색할 수 있도록 하였다.

실험은 현재 방송되고 있는 KBS 뉴스를 대상으로 하였으며 제안된 검색 방법에서 세부 항목별 Recall과 Precision이 85% 이상의 결과를 얻을 수 있었다.

향후 효율적인 뉴스 비디오 검색을 위해서는 비디오 뉴스를 대상으로 얼굴 추출 기법 개선과 기사 내용을 보다 세분화하여 검색할 수 있는 방법에 대한 연구가 필요하다. 그리고 더 많은 의미 정보를 추출하기 위하여 음성 인식에 의한 내용 추출 방법도 필요하며, 여러 뉴스 비디오 구조에 적용하여 신뢰성 있는 모델로 개선해야 할 것이다.

참고문헌

- [1] Ediz Saykol, Ali Kemal Sinop, Ugur Gudukbay, Ozgur Ulusoy, "Content-Based Retrieval of Historical Ottoman Documents Stored as Textual Images," IEEE transactions on images processing, Vol. 13, No 3, Mar. 2004
- [2] Gaurav Aggarwal, Ashwin T V., Sugata Ghosal, "An Image Retrieval System with Automatic Query Modification," IEEE transactions on Multimedia, Vol. 4, No. 2, July 2002.
- [3] W. Qi, L. Gu. et al, "Integration Visual, Audio and Text Analysis for News Video," Proc. of International Conference on Image Processing, Vol. 3, pp. 520-523, 2000.
- [4] Rainer Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithm," Conf. of Storage and Retrieval for Image and Video Databases VII, pp 290-301, 1999.