

내용기반에 의한 뉴스 비디오 검색 시스템

Content-based News Video Retrieval System

배종식*, 양해솔**, 최형진***

한림성심대학 디지털문화콘텐츠과*, 호서대학교 벤츠전문대학원**, 강원대학교 컴퓨터과학과***

Jong-Sik Bae(jsbae@hsc.ac.kr)*, Hae-Sool Yang(hsyang@office.hoseo.ac.kr)**,
Hyung-Jin Choi(choihj@kangwon.ac.kr)***

요약

본 논문은 멀티미디어 정보 검색을 위한 비디오 데이터 처리 기법에 관한 연구로서 뉴스 비디오 도메인에 기반하여 비디오 정보를 효과적으로 검색할 수 있는 비디오 검색 시스템이다. 효과적인 시스템을 구축하기 위하여 비디오 데이터의 생성과 구성에 관한 사전 지식을 이용하여 의미 정보와 특징 정보를 추출한다. 이를 바탕으로 뉴스 비디오를 내용별로 인덱싱하여 신속하게 뉴스 비디오를 검색하도록 한다.

본 논문에서는 실제 KBS 방송국에서 방송 중인 뉴스에 적용하여 실험하고 시스템 평가를 위하여 프리시즌과 리콜을 사용하였다.

■ 중심어 : | 뉴스 비디오 | 내용기반 | 검색 시스템 | 키 프레임 |

Abstract

The study is content-based video retrieval system based on the news video domain as researching for the video data processing method for searching the multimedia information. For the implementation of effective system, We retrieval meaning information and special information using the knowing knowledge about formation and structure of the video data. These are possible to retrieval searching by articles fast and accurately by indexing contents the users want to search.

The news domain used in experiment of our system is the KBS news on the air nowadays and precision and recall is used to evaluate the experiment and performance.

■ keyword : | News Video | Content-based | Retrieval | Meaning Information | Key Frame |

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[1].

영상을 처음부터 보지 않고 특정 부분부터 보기를 원

하는 경우에는 단지 시간이나 텍스트 정보만으로는 사용자가 원하는 부분을 잘 선택할 수 없다. 이러한 문제점을 해결하기 위해서 키 프레임을 기반으로 한 사용자 인터페이스가 있다.

키 프레임은 내용별 비디오를 분류하기 위한 인덱싱으로 사용되며, 키 프레임을 분석하여 여러 가지 정보들을 키 프레임 특징 정보에 저장하고 내용별 비디오를

분류하기 위한 정보로 사용한다[2][3].

본 논문에서는 비디오 뉴스를 앵커 장면, 비 앵커 장면, 뉴스 아이콘, 자막으로 세분화하여 내용에 기반하여 분류하고 검색하는 접근 방법을 제안한다.

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.

본 논문에서는 앵커 장면을 추출하기 위하여 키 프레임 특징 정보에서 얼굴의 수가 두 개 이하이며, 일정한 시간동안 유지되는 등의 비디오 데이터의 생성과 구성에 관한 사전 지식을 이용한다. 의미 정보는 사전 지식에 의하여 추출된 프레임이 키 프레임인지를 판단하기 위하여 가지고 있는 정보이며, 특징 정보는 의미 정보를 이용하여 추출된 키 프레임이 가지고 있는 정보로서 사전 지식과 일치하는 정보 값이다.

실제 방송국에서 방송 중인 뉴스에 적용하여 실험하고 시스템 평가를 위하여 Precision과 Recall을 사용한다.

II. 비디오 검색 시스템

1. 비디오 데이터의 구조

프레임(frame)은 비디오 데이터에서의 최소 단위로 각각의 픽셀로 이루어진 필름 한 장에 해당하는 하나의 정지 화상을 말한다. 샷(shot)은 영상을 구분할 때의 기본 단위로써 필름이 연결된 프레임 집합의 구성이며, 일반적으로 장면 전환 추출의 기본 단위로 사용된다.

에피소드(episode)는 하나의 촬영 대상이 연속된 샷의 집합으로 이루어진 공간상, 시간상의 의미적 분리의 구성이며, 장면(scene)이라고도 한다.

컷(cut)은 샷과 샷 사이의 경계이며 장면이 전환되는 프레임이다[4].

2. 비디오 장면 전환의 종류 및 추출

비디오 데이터는 내용이 바뀔 때마다 해당 내용을 분류하고, 색인화해야만 정보를 체계적으로 데이터베이스

화하고 효율적으로 검색할 수 있다. 기본적으로 카메라 샷에 의해 구분되는 컷 추출은 여러 가지 편집 효과에 의한 페이드 인/아웃, 디졸브와 같은 점진적 변화와 주밍, 패닝과 같은 카메라의 움직임으로 발생하는 점도 고려해야 한다[5].

비디오 데이터에서 장면 전환이 이루어지는 프레임을 추출하기 위해서 일반적으로 두 프레임간의 특징 차이를 이용한다. 샷은 카메라가 촬영한 연속된 장면이므로 샷 내부에 존재하는 연속적인 프레임들 간에는 전체 영상 밝기 및 영상 내에 존재하는 물체들의 분포나 개수 등의 특징이 비슷하다. 그러나 장면 전환이 일어나는 두 프레임에서는 프레임간의 밝기나 히스토그램 차이 등의 특징 차이가 어떤 임계치보다 크게 나타난다. 즉 프레임과 프레임 사이에 커다란 변화가 존재하면 장면 전환으로 판단하여 추출한다. 장면 전환이 일어나지 않는 연속된 프레임들은 인접 프레임 사이에서 유사성이 매우 강하지만, 장면 전환이 일어나는 부분의 프레임들은 상대적으로 프레임간의 유사성이 약하다.

3. 비디오 자막

비디오에서 자막은 정확한 내용 전달을 위하여 현재 장면에 대한 부연 설명을 자막으로 보여준다. 비디오의 자막은 편집 텍스트와 썸 텍스트가 있으며 편집 텍스트는 특정한 내용을 설명하기 위하여 인위적으로 화면에 삽입한 텍스트이며, 썸 텍스트는 카메라에 찍혀 나타나는 텍스트로서 아무런 의도없이 비디오 데이터에 포함된 자막이다.

4. 얼굴 추출

얼굴 추출은 영상 내에서 얼굴이 존재하는지를 검사하여, 얼굴의 위치와 범위 등의 정보를 알아내는 작업이다. 하지만, 현재까지 많은 얼굴 처리 시스템에서 사용자가 인위적으로 얼굴을 추출하는 것이 일반적이다.

얼굴 추출은 다양한 요인들로 인하여 큰 영향을 받는다. 얼굴의 크기, 위치에 따라서 결과가 달라질 수도 있으며 얼굴의 방향이나 회전된 각도, 포즈에 따라서도 큰 영향을 받는다.

이러한 얼굴 자체의 요인 뿐 아니라 영상의 빛의 밝

기, 카메라의 성능, 얼굴 주변 상황 등의 얼굴 외적인 요인들도 상당한 영향을 끼치게 된다.

III. 내용기반에 의한 뉴스 비디오 검색 시스템 설계 및 구현

내용기반에 의한 비디오 검색을 위하여 본 논문에서는 [그림 1]과 같이 시스템 구조를 설계하였다.

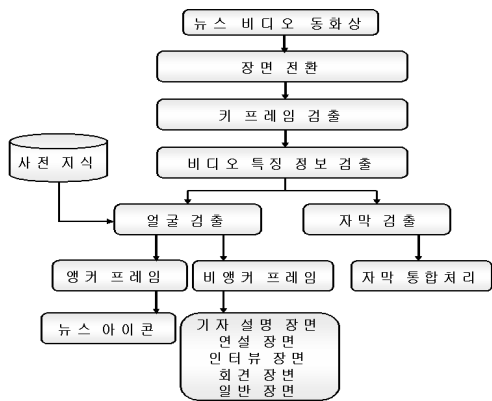


그림 1. 뉴스 비디오 검색 시스템의 설계

1. 장면 전환 추출 및 키 프레임 추출

본 논문에서는 뉴스 비디오의 장면 분류를 샷 단위로 하여 장면 전환을 추출하며, 장면 전환 추출 결과를 기반으로 하여 내용상 어떠한 분류에 속하는 장면인지를 판별하고 색인화하여 기사별 장면 검색이 가능하도록 한다.

뉴스 비디오의 장면 분류를 위해 의미적으로 동일한 장면에 대하여 내용적으로 앵커 장면과 비앵커 장면으로 분류한다. 장면들은 특징적으로 발생하는 색상과 에지 성분이 각각 다르기 때문에 칼라 히스토그램 정보와 에지 특징 정보를 이용하여 장면들을 분류한다. 키 프레임은 각 샷을 대표하는 프레임으로 장면 변화가 크게 발생하는 지점의 프레임이나 장면 전환이 완성되는 프레임은 키 프레임으로 추출한다.

본 논문에서는 장면 전환 추출을 위하여 [그림 2]와 같이 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 영상의 조건을 같게하여 일반화된 특징을 추출하기 위함이다.

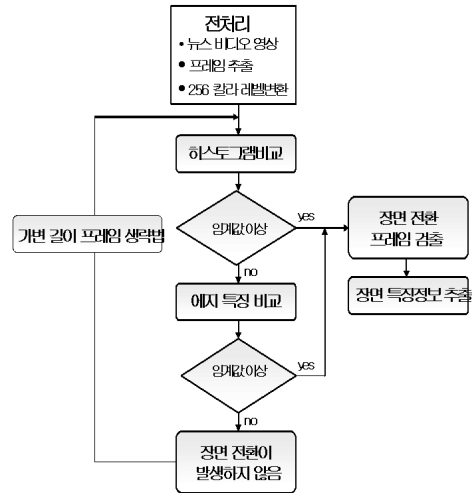


그림 2. 장면 전환 추출 과정

가변 길이 프레임 생략법은 가중치를 부여하여 장면 전환이 자주 일어나는 경우에는 생략하는 프레임 수를 줄이고, 장면 전환이 거의 발생하지 않는 경우에는 생략하는 프레임 수를 늘려서 히스토그램 비교 횟수를 줄이는 방법이다.

2. 앵커/비앵커 장면 추출

앵커 장면은 [그림 3]과 같이 얼굴 추출 알고리즘을 통해 키 프레임 특징 정보에 저장되어 있는 얼굴 정보를 이용하여 앵커 장면인지를 판별한다. 앵커 장면을 추출하기 위한 사전 지식으로 키 프레임 특징 정보에서 얼굴의 수가 두 개 이하, 일정한 시간 동안 유지되며 주기적으로 출현, 배경화면의 무 변화, 앵커가 두명이면 좌우 대칭형, 앵커가 한 명이면 좌측, 중앙, 우측에 위치, 자막출현, 최소 2초 이상 앵커 프레임 지속 등을 설정하여 이를 바탕으로 앵커 장면을 추출한다.

비 앵커 장면은 앵커 장면 추출 후 다음 앵커 장면이 추출되기 전까지 나타나는 장면으로서 앵커 장면 이후

시간 순으로 전개되며 사건 담당 기자가 나와 앵커의 기사 설명을 보충하는 형식으로 구성된다. 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.

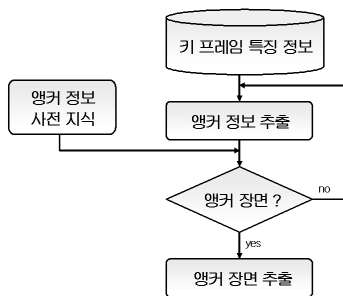


그림 3. 앵커장면 추출 과정

3. 자막 통합 처리

뉴스 비디오는 사건과 관련된 중요한 내용을 요약하여 자막으로 표현하며 정형화된 특징을 가진다. 자막 통합 처리는 [그림 4]와 같이 뉴스 비디오에 나타나는 자막을 모아서 문서 형식으로 구성하여 하나의 프레임으로 볼 수 있도록 한다. 앵커 장면 추출 지점에서부터 다음 앵커 장면 전까지 자막이 가지는 특징을 이용하여 자막 영역을 추출한 후, 색 단순화 방법으로 자막의 유무를 판단하여 통합 처리한다. 자막의 특징으로는 대부분 화면 아래쪽에 나타나며, 배경을 가지고 있으며 글자가 선명하고 흰색을 띤다.

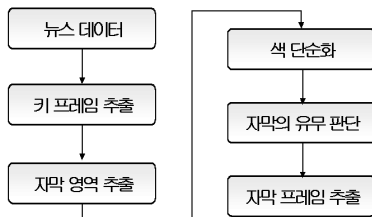


그림 4. 자막 추출 구조도

IV. 구현 결과 및 성능 평가

1. 실험 결과

1.1 검색 시스템 구현

본 논문에서 구현에 사용한 비디오 데이터는 AVI 형식의 320×240 크기의 256 칼라 레벨이고 초당 30 프레임씩 재생되는 KBS 9시 뉴스를 사용하였으며 비주얼 C++ 6.0으로 구현하였다. [그림 5]는 사용자 인터페이스를 위한 초기 화면이다. 비디오 파일을 열면 동영상 정보 영역에 가로크기, 세로크기, 전체 프레임 수를 표시하며 분석 속도는 동영상의 화면 출력 속도를 위하여 사용자로부터 초 단위의 시간을 입력 받는 항목이다.

구현된 시스템은 사용자에게 내용 분류에 의한 검색 과정을 화면으로 보여주는 기능과 화면 출력 없이 내부적으로 처리하는 기능을 가진다.



그림 5. 구현한 검색 시스템의 초기 화면

1.2 장면 전환 추출 결과

[그림 6]은 히스토그램 분석에 의한 장면 전환 추출 결과를 보여주는 화면으로서 숫자는 장면 전환이 발생한 키 프레임의 번호이다. 번호를 선택하면 현재 프레임에서부터 다음 장면 전환 지점까지 연속적으로 재생된다. 에지 분석은 히스토그램 분석과 에지 분석을 실행하여 장면 전환으로 추출된 키 프레임의 번호를 표시한다.



그림 6. 장면 전환 추출 결과

1.3 앵커 장면 추출 결과

[그림 7]은 앵커가 한 명일 때 추출된 결과이다. 앵커 검출 버튼을 누르면 현재 앵커 장면에서부터 다음 앵커 장면까지 연속적으로 재생된다.



그림 7. 앵커 장면 추출 결과

1.4 비 앵커 장면 추출 결과

비 앵커 장면은 일반 장면, 기자 설명 장면, 연설 장면, 회견 장면이나 인터뷰 장면으로 구분되며 각 장면의 버튼을 선택하면 검색된 장면이 연속적으로 재생된다. [그림 8]은 일반 장면을 추출한 결과로서 일반 장면이 연속적으로 재생된다.



그림 8. 비앵커 장면추출 결과

1.5 그 외 장면 추출결과

앵커/비앵커 장면의 뉴스 아이콘 장면, 자막통합처리 결과장면, 비앵커 장면에서 기자설명 장면, 인터뷰 장면, 연설장면, 일반장면 등을 추출하였다.

2. 성능 평가

본 논문에서는 구현한 시스템의 성능 평가를 위하여 (식 1)인 Recall과 (식 2)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[6].

$$Recall = \frac{C_n}{C_n + M_n} \quad (1)$$

$$Precision = \frac{C_n}{C_n + F_n} \quad (2)$$

(식 1)과 (식 2)에서 C_n 은 정확히 추출된 개수, M_n 은 미추출된 개수, F_n 은 오추출된 개수를 나타낸다.

실험에 사용된 뉴스 비디오를 내용별로 분류하면 [표 1]과 같다.

표 1. 실험에 사용된 뉴스 비디오 데이터

분류		갯수
전체 프레임 수		35,900
장면 전환 프레임 수		203
앵커 프레임 수	1 명	14
	2 명	2

기자 설명 프레임 수	7
인터뷰 프레임 수	32
연설 프레임 수	5
일반 장면 프레임 수	87
뉴스 아이콘 수	14
자막 프레임 수	42

[표 2]는 실험 데이터의 추출 결과를 나타내는 자료로서 N_{total} 은 $N_s + N_m$ 이며 실험 데이터에서 나타난 각 항목의 전체 개수를 의미하고, N_s 는 정확히 추출된 개수, N_m 은 미추출된 개수 그리고 N_f 는 오추출된 개수를 의미한다.

표 2. 실험 결과 데이터

분 류	N_{total}	N_s	N_m	N_f
장면 전환 프레임 수	203	190	13	5
앵커 프레임 수	1 명	14	14	0
	2 명	2	2	0
기자 설명 프레임 수	7	6	1	2
인터뷰 프레임 수	32	28	4	4
연설 프레임 수	5	5	0	2
일반장면 프레임 수	87	83	4	5
뉴스 아이콘 수	14	14	0	2
자막 프레임 수	42	40	2	5

[표 3]은 실험 결과의 데이터를 기초로 Recall과 Precision을 한 결과이다.

표 3. 실험 평가

분 류	Recall	Precision
장면 전환 프레임 수	93%	97%
앵커 프레임 수	1 명	100%
	2 명	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 변화여 색 단순화 알고리즘이 자막인식을 제대로 하지 못했기 때문에 발생하였다.

V. 결론

뉴스 비디오의 검색 작업은 많은 사용자들에게 편리한 검색과 정보를 제공하는 주문형 뉴스 서비스에 있어서 필수적인 기술이다. 사용자가 원하는 정보를 보다 간단하고 효율적으로 관리하며, 주요 정보를 빠르게 검색하기 위한 여러 검색 기법들이 제안되고 있으며 일반적으로 영상이 가지는 내용 정보, 즉 영상 자체로부터 추출된 정보를 이용하여 검색 시스템을 개발하고 있는 상황이다.

의미적 분류는 내용기반에 의한 뉴스 비디오의 세부 내용을 기사별로 검색하기 위한 방법으로 얼굴을 중심으로 분류를 하였다. 얼굴의 다양한 형태 정보인 크기, 개수, 위치 등에 대한 사전 지식과 조건을 설정하여 검색함으로써 내용별 기사 분류에 대한 추출 효율을 향상시켰다. 앵커 프레임의 구조적 특징에 대한 사전 지식을 이용하여 뉴스 아이콘 및 자막을 추출하였기 때문에 85% 이상 검색되었다. 제안된 검색 방법은 Recall과 Precision이 85% 이상의 결과를 얻을 수 있었다.

향후 효율적인 뉴스 비디오 검색을 위해서는 음성 인식에 의한 내용 추출 방법도 필요하며, 여러 뉴스 비디오 구조에 적용하여 신뢰성 있는 모델로 개선해야 할 것이다. 그리고 텍스트 정보를 이용하는 방법에 대한 연구도 필요하다.

참고 문헌

[1] Yi Li, Linda and G. Shapiro, "Object and Concept Recognition for Content-Based Image Retrieval," in Lecture Note in Computer Science, Springer-Verlag, to Appear, 2004.

[2] Anita Komlodi and Gray Marchionini, "Key frame preview techniques for video browsing," in Proc. ACM conference on Digital Libraries(Digital Libraries '98), pp.18-125, 1998.

[3] B. T. Truong, C. Dorai, and S. Venkatesh, "Improved fade and dissolve detection for reliable video segmentation," Proceedings of 2000 IEEE International Conference on Image Processing, Vol.3, pp.961-964, 2000.

[4] 장석우, "퍼지추론에 의한 비디오 데이터의 샷 경계 추출", 한국정보처리학회논문지B, 제10-B권, 제6호, pp.611-618. 2003.

[5] 이종희, 오해석, "의미기반 비디오 검색을 위한 인택싱 에이전트의 설계", 한국정보처리학회논문지 B, 제10-B권, 제6호, pp.687-694. 2003.

[6] Rainer Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithm," Conf. of Storage and Retrieval for Image and Video Databases VII, pp.290-301, 1999.

[7] R. Brunelli and O. Mich, "Histograms analysis for image retrieval," Pattern Recognition Vol.34, Issue8, pp.1625-1637, 2001(8).

저자 소개

배종식(Jong-Sik Bae) 정회원



- 1987년 2월 : 중앙대학교 전자계산학과(공학사)
- 1989년 2월 : 중앙대학교 전자계산학과(공학석사)
- 2005년 8월 : 강원대학교 컴퓨터과학과(이학박사)

- 1996년 ~ 현재 : 한림성심대학 디지털문화콘텐츠과 교수

<관심분야> : 멀티미디어, 콘텐츠 개발, 컴퓨터 비전

양해술(Hae-Sool Yang)

정회원



- 1975년 2월 : 홍익대학교 전기공학과 졸업(학사)
- 1978년 8월 : 성균관대학교 정보처리학과 졸업(석사)
- 1991년 3월 : 日本 오사카대학 정보공학과 S/W공학 전공(공학박사)

• 1975년 5월 ~ 1979년 6월 : 육군중앙경리단 전자계산실 시스템분석장교

• 1980년 3월 ~ 1995년 5월: 강원대학교 전자계산학과 교수

• 1986년 12월 ~ 1987년 12월: 日本 오사카대학 객원연구원

• 1995년 6월 ~ 2002년 12월: 한국소프트웨어품질연구소 소장

• 1999년 11월 ~ 현재 : 호서대학교 벤처전문대학원 교수

• 2010년 3월 ~ 현재 : 호서대학교 글로벌창업대학원 원장

<관심분야> : S/W공학(특히, S/W 품질보증과 품질평가, 품질감리 및 컨설팅, OOA/OOD/OOP, SI), S/W 프로젝트관리, 품질경영

최형진(Hyung-Jin Choi)

정회원



• 1982년 2월 : 영남대학교 물리학과(이학석사)

• 1987년 3월 : 동경공업대학 정보공학과(공학석사)

• 1990년 3월 : 동경공업대학 정보공학과(공학박사)

• 1991년 8월 ~ 현재 : 강원대학교 컴퓨터과학과 교수

<관심분야> : 인공지능, 컴퓨터 그래픽스, 영상처리, 컴퓨터비전