

음성과 이미지를 이용한 동영상 검색에 관한 연구

신인경*, 박성현*, 안효창*, 이상범*
 *단국대학교 컴퓨터응용학과 박사과정
 *단국대학교 컴퓨터응용학과 교수
 e-mail:gguri81@dankook.ac.kr

A Study Video using Image and Voice Search

In-Kyoung*, Sung-Hyun Park*, Hyo-Chang Ahn*, Sang-Burm Rhee*
 *Dept of Applied Computer Engineering, Dankook University

요 약

정보화 사회의 정보 기반 구조로서, 고속 정보망의 구축, 개인용 컴퓨터의 급속한 보급, 멀티미디어 기술의 발전 등으로 인하여 정보 서비스의 새로운 장이 열리고 있다. 동영상 데이터는 텍스트만이 아니라 영상정보, 음성정보등 각종 의미있는 다양한 멀티미디어 정보를 포함하고 있다. 본 논문에서는 동영상에서 음성과 영상을 분리하여 음성을 이용하여 음성열을 분할 및 복원하여 음성을 텍스트로 변환하여 텍스트색인파일을 만들고 영상은 이미지를 분할 및 히스토그램을 사용하여 이미지 샷을 검출하여 두 색인파일을 이용하여 인덱싱을 하여 동영상 검색에 활용한다.

1. 서론

정보화 사회의 정보 기반 구조로서, 고속 정보망의 구축, 개인용 컴퓨터의 급속한 보급, 멀티미디어 기술의 발전 등으로 인하여 정보 서비스의 새로운 장이 열리고 있다. 동영상 데이터는 텍스트만이 아니라 영상정보, 음성정보등 각종 의미있는 다양한 멀티미디어 정보를 포함하고 있다[1].

현재 동영상 데이터로부터 사용자가 원하는 정보를 검색하기까지 기존의 키워드 기반의 검색은 많은 한계에 도달한 상황이기 때문에 사용자가 원하는 정보를 내용에 기반하여 검색할 수 있는 방법이 요구 되고 있다.

따라서 대용량의 데이터를 저장하는 기술과 데이터를 효과적으로 압축, 검색 할 수 있는 방법의 필요성이 증가 되고 있으며 이에 대한 다양한 방법들이 연구되고 있다.

동영상이 늘어나는 이유는 컴퓨터의 하드웨어의 발달과 인터넷 전송속도의 향상 및 모바일 인터넷의 발달로 인하여 뮤직비디오, 영화, 드라마, TV오락 등 다양한 장르의 동영상 데이터들이 공유되고 있기 때문이다[1-2]. 또한 이런 많은 데이터 중에 사용자가 원하는 정확한 데이터를 검색하기에는 아직 텍스트 검색에 많이 의지를 하고 있다. 이러한 텍스트 검색은 사용자들이 원하는 데이터를 검색하는 것이 아닌 동영상을 올린 이의 키워드로 검색이 되어 지고 있다.

방대한 영상이 나오는가운데 데이터를 검색하기엔 몇가지 문제점이 발생되고 있다.

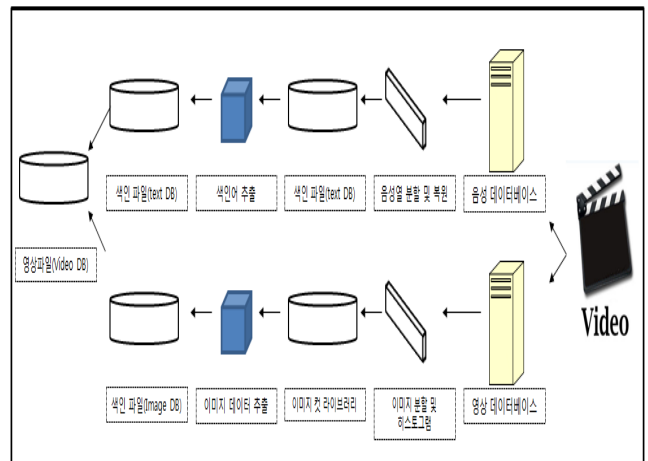
첫 번째로는 영상의 중복성이며, 두 번째로는 키워드와

영상의 일치성에 문제가 되고 있으며, 마지막으로는 저작권 침해등이 있다.

따라서 본 논문에서는 빠르게 변화하는 정보화 시대에 맞게 영상에서 음성과 대표적인 이미지를 검출하여 검색 시 효율적이고 정확한 데이터를 검색 할 수 있도록 검색 알고리즘을 제안하고 두 번째 문제인 내용과 키워드의 불일치한 점을 이미지 검출과 음성 검출을 통하여 키워드를 분리하여 효율적인 검색율을 높인다.

2. 이미지 검색 시스템

동영상에서 이미지를 검출하는 시스템은 아래의 그림1과 같다.



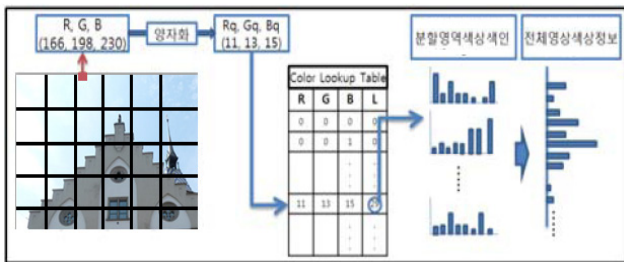
[그림 1] 전체 시스템

2.1 이미지 검색

사용자는 이미지 검색을 통하여 영상을 편집하여 인덱싱된 이미지 데이터를 만들 수 있다. 이를 위하여 처음과 마지막의 프레임을 지정하여 관련된 내용에 관한 주석과 장소에 관한 정보를 입력하여 또 다른 하나의 이미지 파일들을 만들 수 있으며 이미지를 데이터를 만드는 과정은 다음과 같다.

2.1.1 이미지 샷 만들기

영상을 재생하면서 사용자 임의대로 원하는 부분을 선택하여 이미지를 만들 수 있지만 처음부터 마지막 프레임까지 모두 하나하나 보면서 편집해야 하므로 많은 시간이 소요된다는 단점이 있다. 이러한 편집 시간을 줄이기 위해 비디오 컷 프로그램을 이용하여 자동으로 정면의 영상들을 편집하여 이미지를 만들어 줄 수 있으며 비디오 컷 프로그램에서는 이미지 컷 분할과 히스토그램을 기반으로 된 이미지를 검색한다. 이미지 컷 분할은 영상의 변화가 심할 경우 여러 장의 이미지가 생성이 될 수 있으며, 장면 변화와 무관하게 이미지 샷이 만들어 질수 있어 히스토그램을 기반으로 하여 이미지 샷을 만든다.



[그림 2] 이미지 샷 분할과 히스토그램

2.1.2. 영상 이미지 컷 라이브러리 생성

생성된 같은 내용의 이미지 샷을 모아 하나의 파일로 만들어 이미지 컷 라이브러리를 만든다. 이 파일에는 주석과 관련된 정보를 위하여 테이블을 사용하여 동영상의 맨 앞에 파일에 관한 영상No, 생성날짜, 파일의 제목등을 넣어 분류하는데 사용한다. 또한 음성 인식을 통하여 생성된 주석을 통하여 영상 이미지컷 인덱스에 넣어 효율적인 영상 검색을 위해서 파일의 생성과정은 다음과 같이 구성된다.

① 파싱

사용자에 의해 입력되어 지는 문장을 파싱하며, 공백이나 그 외의 특수 문자들은 제한을 걸어두며, 주석에 포함된 모든 단어들을 얻는다[3].

② 키워드 추출

키워드 추출은 동영상에서 가장 많이 나타나는 고유명사를 나타내면 키워드 추출은 음성 파일을 검색하여 고유

명사의 검색 추출되는 단어의 횟수 중 상위 1%의 값을 가지고 나타낸다.

③ 인덱싱

위의 과정으로 생성된 키워드와 관련된 파일의 정보를 테이블의 레코드 위치를 포함하는 인덱스 파일을 생성한다.

2.2 음성 키워드 검색

사람이 발화한 음성을 처리하여 문자 데이터로 변환하고 할 때 일반적으로 발생하는 몇 가지 문제점이 발생하는데 그 중 하나는 자연스럽게 연속 음성을 적절한 구획으로 분할해야 하는 문제이고, 다른 하나는 한국어의 특성상 연속적으로 발음 할 때 형태소 간에 음운 변동이 일어나 발음열과 자자소열이 다르게 나타나는 것이다.

음성 키워드 검색을 위한 과정은 다음과 같다.

2.2.1 어절 분리

어절 분리는 발음열 분할을 위해 나타낸 것으로 연속된 발음열을 받아 앞 음절의 종성에 따라 어떠한 음운 변동이 일어나는지 예측하여 발생가능한 모든 조건을 고려하여 정보를 추출하며 이렇게 분열된 발음열을 앞 음성의 종성과 다음 음절 초성 사이의 음운 변동을 통하여 음성의 어절을 분할하여 정보를 갖는다[3-4]. 이렇게 발음열은 두 개의 분할 과정을 거쳐 음운변동과 어절을 분할 한 다음 음절의 분리를 위하여 사용되는데 한 음절을 사이에 두고 분리 위치가 추정된 경우에 올바른 위치를 잡기 위해 이용된다.

위의 방식으로 음절을 분리하는데 이때 음절 분리 가능도를 $B(Ei)$ 는 음절 Ei 가 끝 음절로 사용되는 가능성이라고 하고 다음과 같은 식을 적용한다.

$$B(Ei) = \log \frac{He(Ei)}{Hf(Ei)} \quad (식1)$$

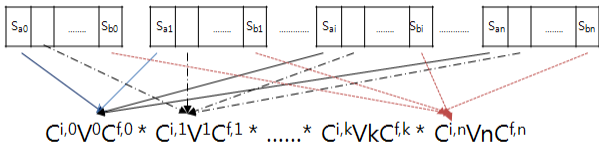
식(1)에서 Ei 가 끝 어절로 사용되는 가능성이라고 정의하였고 각 영상의 어절을 추출하여 다음 식(2)에 의해 계산된다.

$$He(Ei) = \frac{Eihi}{E1h1 + E2h2 + \dots + Eih_i + Enh_n} \quad (식2)$$

식(2)에서 $E1, E2, \dots, Ei, En$ 은 끝음절로 사용되는 음절을 나타냈으며 $h1, h2, \dots, hi, hn$ 은 끝음절로 사용된 음절의 빈도값을 나타냈다.

어절 분리는 위의 단계를 거쳐 연속된 음성 발음열을 어절이라 추정된 후 적절한 크기로 분할 시켜주며, 분할크기와 위치가 결정되면 음절 복원기에서 분할된 음원을 받아 복원을 수행 한다.

분할된 음절의 크기와 음절의 복원 패턴의 대응 단계는 다음과 그림3과 같다.



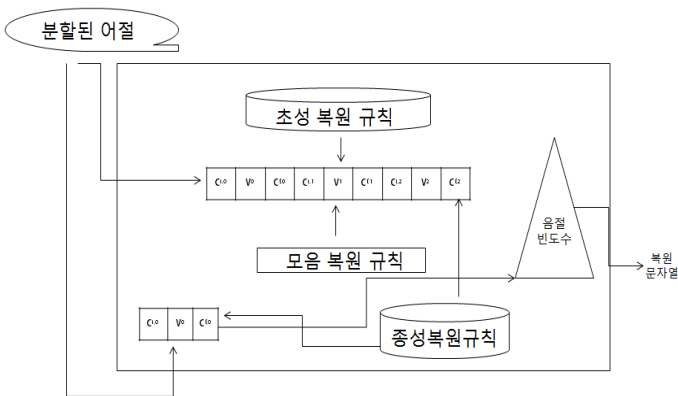
[그림3] 분할된 음절과 음절 복원

분할된 음절과 음절 복원의 식은 다음 식(3)과 같다.

$$S_a \dots S_b = C_{i,0}V^0C_{f,0}^* C_{i,1}V^1C_{f,1}^* \dots C_{i,k}V^kC_{f,k}^* \dots C_{i,n}V^nC_{f,n}^* \quad \text{식(3)}$$

2.2.2 음절 복원

어절 분리에서 생성된 어절들을 입력으로 받아 복원을 실행한다. 음절 복원의 흐름도는 그림4와 같다. 그림 4에서는 위의 식3의 형태로 입력된다.



[그림 4] 음절 복원의 흐름도

2.2.3 형태소 분석

어절 분할과 음절 복원의 처리 과정을 통해 생성된 텍스트데이터베이스(Text DataBase)에 대해 문장 단위로 형태소 분석을 한다. 형태소를 분석 후 교정 단계에서 교정을 수행한다.

2.2.4 교정

분할된 어절은 규칙에 의하여 복원 후 형태소를 분석한다. 어떤 어절에 대하여 분석에 성공한 값이 존재하지 않을 때는 위의 과정에서 오류가 있다고 판단 적절한 교정을 수행하며, 이 역할을 교정기가 담당한다. 교정의 대상은 오류로 판단되는 어절의 경계를 추정하여 위의 분할 및 복원을 다시 수행한다.

이때 어절 분할은 띄어 쓰기 오류로 보고 교정을 한다. 띄어 쓰기 오류는 여러 어절이 하나의 어절로 결합된 경우

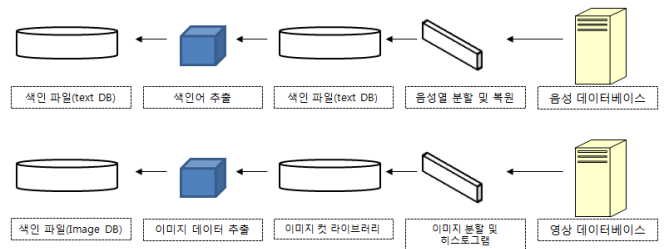
로 함 음절씩 잘라서 적절한 위치를 찾아 넣는다. 이 과정이 완료되면 음성정보에서 변환된 텍스트데이터베이스가 만들어지게 된다.

이렇게 만들어진 텍스트데이터베이스에는 효율적인 색인어를 추출하기 위해 상위 10%내외의 출현 고빈도어를 대상으로 색인 데이터베이스를 작성한다.

3. 정보 검색

정보 검색이란 사용자의 정보 요구가 얼마나 근접한가를 계산하여 그 값이 높은 순서로 문서를 검색하는 것이다. 보통 문서의 내용은 사용자 질의의 유사도를 계산하는 정합(matching)을 통하여 효율적인 문서를 분석하여 내부적인 표현을 하는 것인데 이를 인덱싱이라고 한다[6-7].

본 논문에서는 이미지 컷 라이브러리와 함께 음성파일에서 검출된 분할된 음절과 복원된 음절을 통하여 얻어낸 텍스트데이터베이스에서 각각의 정보를 얻어 그 값을 인덱싱하여 데이터베이스에 넣는다.



[그림 5] 색인 데이터 베이스를 활용한 정보검색 모델

4. 결론

복합적인 정보를 갖는 동영상 데이터로부터 사용자에게 원하는 정보를 제공하기 위해서는 내용기반의 검색이 필요로 한다. 멀티미디어 정보의 효율적인 검색을 지원하며, 정보 검색 기술의 연구 동향을 비추어 보 면 디지털 영상 신호처리, 컴퓨터 비전, 자연어 처리, 멀티미디어, 음성 인식 및 처리, 데이터베이스 분야를 포함한 다양한 영역의 지식들이 어울어져 활용 될 것으로 보이기에 이를 이용한 연구가 필요하다.

본 연구에서는 비디오 데이터의 내용을 기반으로 검색을 위한 영상 편집 및 음성데이터베이스를 텍스트데이터베이스로 변환하고 변환된 텍스트데이터베이스로부터 색인 데이터베이스에 대하여 설명하였다. 국내의 동영상 검색은 주로 영상 정보와 문자 정보만을 사용한 반면 본 연구에서는 오디오 정보를 이용하여 텍스트데이터베이스로 변환하고 변환된 값으로 색인데이터베이스를 구축함으로써 텍스트와 음성, 영상 정보를 활용하였다는 점이 차이라고 할 수 있다.

이후 제안한 정보검색모델의 사용자 인터페이스부분 모

들 부분과 제안한 기능들을 조금씩 개선해 나가는 것이다.

5. 참고문헌

[1] Veltkamp, R. and Tanase, M., "Content Based Image Retrieval Systems : A Survey", 'Department of Computing Science, Utrecht University, working material(2000)

[2] Zhang Lei, Lin Fuzong, and Zhang Bo, "A CBIR Method Based On Color-Spatial Feature", TENCON 99, Proceedings of the IEEE Region 10 Conference publication Date: 1999.

[3] RGB채널 색상 정보와 HSV 채널 색상 정보
<http://explanation-guide.info/meaning/HSV-color-space.htm>

[4] Microsoft Research Cambridge Object Recognition Image Database ver1.0

[5] 양방향 최장일치법에 의한 한국어 형태소 분석기의 구현, 최재혁, 1993.

[6] Caltech categories dataset
<http://www.cision.caltech.edu/html-files/archive.html>

[7] David Gavilanm Hiroki Takahashi and Masayuki Nakajima. "Contents Classification in a Mobile Environment: Image Categorization using Color Blobs", In IEICE, Sendai, Japan, March 2003.