

동영상에서 이미지와 음성신호를 이용한 검색 알고리즘 구현

신인경, 이상범

단국대학교 컴퓨터공학부

gguri81@dankook.ac.kr, sbrhee@dankook.ac.kr

(Search Algorithm Embodiment which uses the Image and Speech Signal from the Vido)

In-Kyoung Shin*, Sang-Burm Rhee*

*Dept. of Computer Engineering, Dankook University

요 약

정보통신망 및 멀티미디어 기술의 발전으로 인해 정보의 형태는 단순한 텍스트 데이터에서 멀티미디어 데이터로 전환되고 있다. 멀티미디어 기술은 저장, 재생, 압축 등 관련 기술의 빠른 발전과 미디어의 사회, 문화적 역할이 계속 증가함에 따라 우리 사회 전반에 걸쳐 매우 광범위하게 사용되고 있으며, 이로 인해 동영상 검색등의 많은 검색을 요구 하고 있으나, 동영상 검색의 문제점은 생산되는 콘텐츠에서 동영상이 가지고 있는 비중은 계속해서 높아지지만 쌓아진 데이터를 검색하기엔 몇 가지 문제점이 있다. 첫 번째는 데이터의 중복성이고 두 번째는 제목, 내용 그리고 Keyword가 일치하지 않으며, 세 번째는 저작권 침해 등이 있다. 본 연구에서는 본 논문에서는 빠르게 변화되고 있는 정보화 시대에 맞게 동영상에서 음성과 얼굴영역을 검출하여, 검색 시 효율적이고 정확한 데이터의 검색이 이루어 질 수 있도록 검색 알고리즘을 제안하고 소개하며, 이중 두 번째의 문제점인 제목, 내용 그리고 Keyword의 불일치한 점에 두어 검색 시 영상의 이미지 검색과 음성을 통하여 keyword를 찾아 효율적이고 검색율의 높일 방법을 연구한다.

1. 서 론

최근 정보통신망 및 멀티미디어 기술의 발전으로 인해 정보의 형태는 단순한 텍스트 데이터에서 멀티미디어 데이터로 전환되고 있다. 멀티미디어 기술은 저장, 재생, 압축 등 관련 기술의 빠른 발전과 미디어의 사회, 문화적 역할이 계속 증가함에 따라 우리 사회 전반에 걸쳐 매우 광범위하게 사용되고 있다.

문자, 소리, 이미지, 동영상 등의 다양한 멀티미디어 데이터들이 기하급수적으로 축적되면서 빠르고 효율적인 접근, 검색, 저장 등의 기술적 중요성이 점차 더 커지고 있으며, 정보화 시대에 월드 와이드 웹의 등장으로 인터넷을 통해 전

세계의 수많은 사이트에서 제공되는 정보를 검색하고 이용할 수 있게 되었다.

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

동영상의 검색이 늘어나게 된 것은 보면 컴퓨터 하드웨어의 발달과 인터넷 전송속도의 향상에 따라 온라인에서는 UCC(User Created Contents), 뮤직비디오, 영화, 드라마, TV 오락 등 다양한 장르, 다양한 형태의 동영상 데이터들이 공유되고 있다. 많은 데이터 중에 사용자가 원하는 일치된 영상을 검색하려면 이미지나 텍스트를 가지고 많이 검색을 한다. 그러나 이 데이터들은 영상을

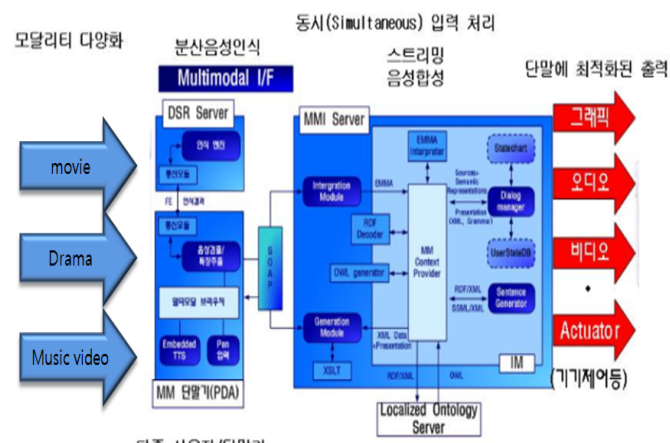
검색하는 것이 아닌 동영상을 올린 이의 Keyword 로 검색이 되어지고 있다. 사용자가 원하는 데이터임에도 불구하고 Keyword 가 등록되지 않으면 검색이 되지 않는 오류가 범해지고 있다.

동영상 검색의 문제점은 생산되는 콘텐츠에서 동영상이 가지고 있는 비중은 계속해서 높아지지만 쌓아진 데이터를 검색하기엔 몇 가지 문제점이 있다. 첫 번째는 데이터의 중복성이고 두 번째는 제목, 내용 그리고 Keyword 가 일치하지 않으며, 세 번째는 저작권 침해 등이 있다.

본 연구에서는 본 논문에서는 빠르게 변화되고 있는 정보화 시대에 맞게 동영상에서 음성과 얼굴영역을 검출하여, 검색 시 효율적이고 정확한 데이터의 검색이 이루어 질 수 있도록 검색 알고리즘을 제안하고 소개하며, 이중 두 번째의 문제점인 제목, 내용 그리고 Keyword 의 불일치한 점에 두어 검색 시 영상의 이미지 검색과 음성을 통하여 keyword 를 찾아 효율적이고 검색율의 높일 방법을 연구한다.

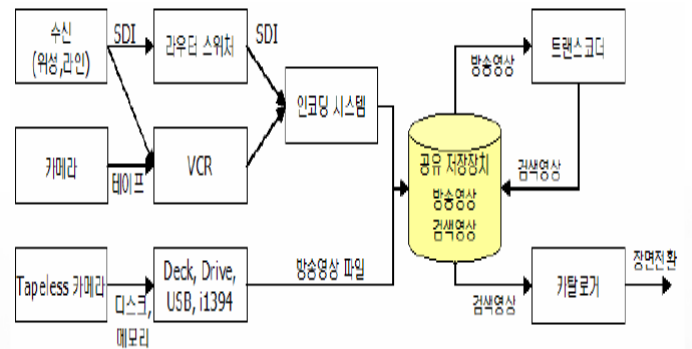
2. 연구의 목표 및 내용

본 연구에서는 동영상에 음성 데이터와 영상데이터를 추출하여 사용자가 음성 및 이미지로 검색 시 동영상 내에 검색 데이터와 많은 부분 일치하여 최대한의 검색율을 올리는 것을 최종 목표로 한다.



[그림 1] 멀티모달 아키텍처

이러한 연구 방향은 현재 많은 연구가 진행되고 있는 문자표현에 의한 내용기반 방식과는 다른 방향으로, 즉 동영상 자체내에 존재하는 음성 데이터와 영상을 캡처하여 이미지로 그대상과 얼마만큼 일치하는 지를 분석 수행한다.



[그림 2] 구성도

음성 처리부는 크게 전 처리부와 인식부로 나눌 수 있다. 전 처리부에서는 사용자가 발생한 음성으로부터 인식에 필요한 특징 벡터를 추출하고, 인식부에서 음성 데이터베이스로 부터 훈련한 기준 패턴과의 비교를 통해서 인식 결과를 얻게 된다. 보다 복잡한 구조의 음성을 인식할 때에는 언어모델을 이용한 언어 처리과정을 통해 최종 인식 결과를 출력한다.

2. 1. 전처리 과정

2.1.1 끝점 검출(End-point detection)

입력되는 신호로부터 음성구간만을 검출하는 처리 과정이다. 음성 인식에 사용하게 되는 신호를 전달하게 되는 가장 처음 과정이므로 매우 중요한 과정이다. 하지만, 전체 인식 처리속도에 영향을 주지 않기 위해서는 실시간에 가까운 처리 속도를 요구하므로 간단한 알고리즘으로 구현되어야 한다. 입력 신호로부터 계산과정이 비교적 간단한 통계적 특징을 추출하여 음성구간과 묵음구간을 구분하는 방식을 사용한다.

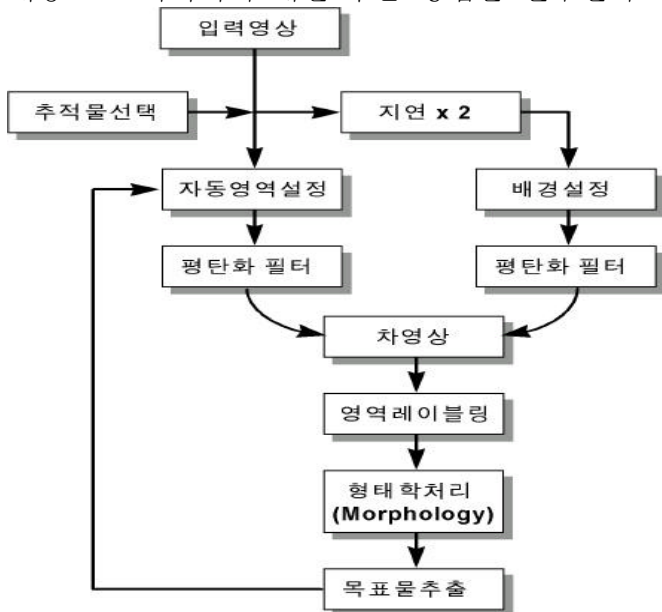
2.1.2 특징 추출(Feature Extraction)

입력된 음성 신호로부터 인식에 유효한 특징 파라미터를 뽑아내는 과정이다. 동일한 단어를 여러 사람이 발음하였을 경우 단어의 의미가 동일하더라도 음성과형은 동일하지 않으며, 동일한 사람이 동일한 단어를 동일한 시간에 연속으로 발음하였다고 하여도 음성 파형은 동일하지 않다. 이와 같은 현상의 이유는 음성 파형에서는 음성의 의미 정보 이외에도 화자의음색, 감정상태 등과 같은 정보도 포함하고 있기 때문이다. 성문신호(Glottal signal)의 영향을 최소화하여 음성과 잡음의 구분을 뚜렷하게 하기 위해 고주파 성분을 강조해준다. 하나의 차수를 갖는 전달 함수로 표현되며 시간상에서는 1 차의 미분 형태로 구현한다.

2.1.3 창 함수(Windowing)

특징 추출 과정은 음성신호를 짧은 길이의 구간으로 나누어 분석하는 것이 일반적이다. 구간으로 나누는 과정에서 구간사이의 신호의 연속성이 반영되지 못하여 이로 인한 주파수 왜곡현상이 발생한다. 이와 같은 구간 앞뒤에서의 신호의 단절로 인한 왜곡 현상을 최소화하기 위해서 구간 앞뒤를 감쇠 시킨 형태의 창 함수 값을 신호 성분 곱하여 분석한다.

이미지 처리부는 움직이는 물체는 제한된 위치에서 이루어지기 때문에 보다 정교한 물체 추출 알고리즘을 요구한다. 따라서 다음과 같은 내용으로 이미지에 대한 추출 방법을 연구한다.



[그림 3] 이미지 추출 순서도

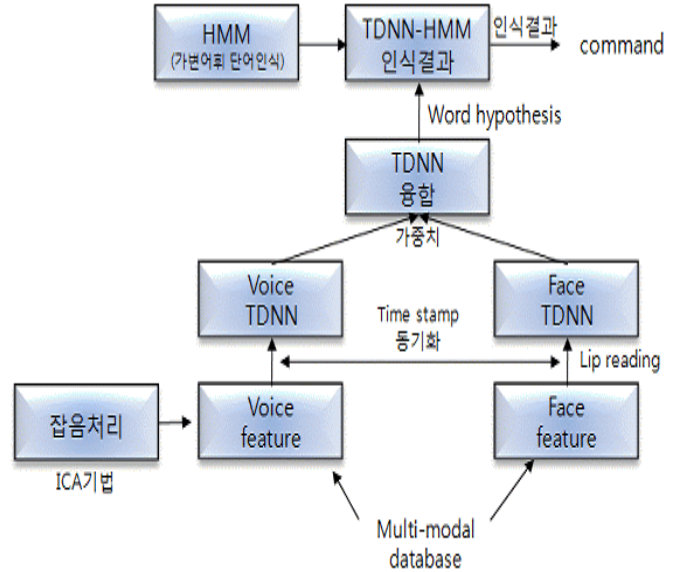
가) 추적물 선택 : 추적하고자 하는 물체를 마우스로 클릭한다. 이때 공간적 좌표가 결정되고 그 점을 중심으로 물체가 다음 프레임에서 영역을 벗어나지 못할 크기의 추적창이 설정된다.

나) 배경설정 : 현재 입력 영상의 2 프레임 뒤 영상을 배경으로 설정한다.

2.1.4 차영상 : 2 프레임 뒤의 배경영상과 현재 프레임의 차영상을 추적창 내에서 구할 때 그 차영상은 2 개의 블록을 가지게 된다. 이때 골프공의 칼라 성분을 이용하여 차영상의 양의 영역을 물체 영역으로 선택한다.

다) 목표물 추출 : 형태학 처리후 이진화된 물체 블록에 대해 무게중심점을 구한다.

라) 자동영역설정 : 무게중심점을 중심으로 새로운 추적창을 설정한다



[그림 4] 최종 연구를 위한 시스템도

2.2 멀티모달 데이터베이스 구축

멀티모달의 동기화된 데이터베이스와 그것을 음소로 레이블링된 데이터베이스를 제작하고, 이를 멀티모달 융합기의 훈련용으로 활용한다. 또한, 데이터베이스 제작 프로그램을 제작하여 추가적인 데이터 수집을 용이하게 한다. 부수적으로 레이블링 검증 프로그램을 제작한다.

2.3 음성 끝점 검출 알고리즘 개발

다중 임계값을 가지며 프레임별 에너지의 변화율을 이용하여 안정적이고 환경에 적응 가능한 음성 경계점 검출을 수행하게 하고 음성기를 구동시킨다.

2.3.1 음성 특징벡터 추출

일반적으로 널리 쓰이는 MFCC 특징벡터와 청각신경 모델링에 기반한 잡음에 강인한 특성을 가지는 특징벡터 추출법을 구현한다.

2.3.2 음성신호 향상법

입력을 받은 잡음 신호를 제거하고 또한 반향효과까지 제거할 수 있는 음질향상법을 독립요소기법을 적용하여 구현한다.

2.4 영상특징벡터 추출

이미지영역 정확히 검출할 수 있도록 컬러 분포도를 활용하는 알고리즘을 개발하며 검출된 영역으로부터 영상특징벡터를 추출하기 위해 SVM 이미지 분류기와 특징벡터차원을 줄이기 위한 PCA 기법을 구현한다. 또한 실시간 이미지를 처리위해 프로그램을 최적화한다.

2.5 음성/영상 특징벡터 동기화

입력으로 들어가는 음성특징벡터와 영상특징벡터의 동기화를 위한 방법을 구현한다. 프레임율은 음성특징벡터에 맞춘다.

2.6 이중모달 융합기 개발

동기화된 음성/영상 특징벡터를 입력으로 받아 자체 성능 및 잡음에서 향상된 성능을 가지도록 Neural Network 과 HMM 접근법을 구현한다. 또한 명령어 인터페이스의 개인화자의 특화된 특성을 반영할 수 있는 문맥정보 활용법을 구현한다.

3. 결론

본 연구에서 제안하고 연구하는 동영상에서의 음성신호와 이미지 영역을 분석하고 동영상에서 제공되는 음성과 이미지를 이용하여 보안성과 편리성이라는 장점을 갖는것이 가능할 것이다. 또한 이용자에 대한 사후 추적이 가능하며 관리면에서도 안전한 시스템을 구축할 수 있으며 향후에도 널리 이용되기에 충분할 것이다.

따라서 본 연구에서 얻어지는 동영상에서의 이미지와 음성의 분할 및 각각의 검색은 서로의 단점을 충분히 보충 해 검색률이 보다 높아 질 것이다.

나. 기대성과

동영상 공유 서비스 시장은 초기에는 저작권에 자유로운 글로벌 서비스 유튜브가 크게 인기를 끌면서 동영상 붐을 일으켰는데 이후 국내의 다양한 동영상 서비스가 등장하였고, 그중에서도 동영상에 댓글을 달 수 있어, 동영상을 매개로 커뮤니티를 즐길 수 있는 독특한 서비스를 들고 나온 일본발 글로벌 서비스가 나오면서 동영상의 관심이 늘어나게 되었다.

동영상 서비스의 성장은 자연스럽게 동영상 검색에 대한 수요도 촉발하여, 기존 거대 검색 사업자도 동영상 검색 서비스를 선보이고 있지만, 동영상 전문 검색 서비스 벤처기업이 속속 등장, 국내 동영상 콘텐츠에 국한되지 않은 다양한 나라의 콘텐츠에 대한 검색은 물론이고 세계 각국의 언어로도 검색할 수 있는 서비스를 선보이며 경쟁력을 높여갈 수 있다.

6. 인용문헌

[1] W.Niblack et al., "Query by Content for Large On-Line Image Collections", IEEE COMPUTER SOCIETY PRESS, pp.357-378, 1995.
 [2] Nagasaka A & Tanaka Y, "Automatic Video Indexing and Full-Video Search for Object Appearance, Visual Database System", IFIP

Elservier Sciece Publishers e.v.,pp.113-127, Oct.1992

[3] Atsushi Ono, Masashi Amano, Mituhiro Hakaridani, Takashi Satou, Masao Sakauchi, "A Flexible Content-based Image Retrieval System with Combined Scene Description Keyword", IEEE, pp.201-208, 1996.
 [4] Kyoji Hirata, Toshikazu Kato, "Query by Visual Example", Extending database Technology ' 92, pp.56-71,1992.
 [5] Myron Flickner, Harpreet Sawhney, Wayne Niblack, Janathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, Peter Yanker, "Query by Image and Video Content : The QBIC System", IEEE, pp.23-32, 1995.
 [6] Jacopo M. Corridoni, Alberto Del Bimbo, Silvio De Magistris, "Querying Retrieving Pictorial Data Using Semantics Induced by Colour Quality and Arrangement", IEEE, pp.219-222, 1996.
 [7] W.Niblack, R.Barber, W.Equitz, M.Flickner, E.Giasman, D.Petkovic, P.Yanker, "The QBIC Project : Querying Image by Content Using Color, texture and Shape", SPIE vol.1908, pp.173-187,1993.
 [8] John. R. Smith, Shih-Fu Chang, "Visual SEEK: a fully automated content-based image query system", ACM Multimeoa ' 96, November, 1996.
 [9] A.Desai Narasimhalu, "Special section on content-based retrieval", ACM Multimedia Systems, 1995, No. 3, pp.1-2.
 [10] Micheal j Swain, "Interactive Indexing into Image Database", SPIE vol.1908, pp.95-103,1993.
 [11] Forsyth, D.A. 1990. "A novel algorithm for color constancy", Intern. J. Computer. Vision 5:5-35.