내용 기반 영상 검색을 이용한 실시간 몽타주 시스템 디자인에 관한 연구

배성준, 최현석, 김낙우, 김태용, 최종수
중앙대학교 첨단영상대학원
email@imagelab.cau.ac.kr

A Study on Real-time Montage System Design using Contents Based Image Retrieval

SeongJoon Bae, HyeonSeok Choi, NacWoo Kim, TaeYong Kim, JongSoo Choi
Graduate School of Advanced Image Science, Multimedia & Film,
Chung-Ang University

요 약

본 논문에서는 내용 기반 영상 검색 기술을 이용하여 사용자가 원하는 영상을 쉽게 찾아내고, 이를 자동 재구성함으로써, 영화 미학의 핵심 중 하나인 몽타주 기법을 사용자 중심의 관점에서 구현하고자 한다.

본 논문에서 제안하는 실시간 몽타주 시스템은 이산 푸리에 변환(Discrete Fourier Transform)을 이용해 사용자가 선택한 영상의 특징을 찾고, 유클리디안 거리(Euclidean Distance)를 이용해 데이터베이스에 있는 영상과 유사도를 비교함으로써, 빠르고 효과적으로 사용자가 원하는 영상을 검색할 수 있다. 또한 카메라 트래킹에 의해 실시간으로 사용자의 움직임 영상을 취득하고, 취득된 영상을 검색된 사용자의 영상과 함께 자동 재구성함으로써, 손쉽게 사용자의 의도에 맞춘 영상 재구성을 하게 된다.

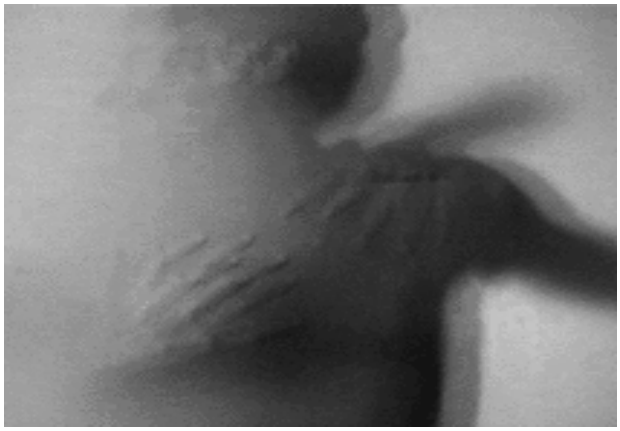
본 시스템은 사용자가 일방적으로 영상을 시청, 감상하는 소극적 영상의 소비자에서 벗어나, 기존의 영상을 이용해 임의의 영상을 조합하고 자기 자신의 영상까지 실시간으로 개입할 수 있도록 함으로써, 영상을 새롭게 구성하고 영상 재생산의 적극적 주체가 되는 사용자 중심의 새로운 영화(뉴미디어)의 토대가 될 것으로 기대된다.

Keyword : montage, contents based image retrieval, interaction, new media

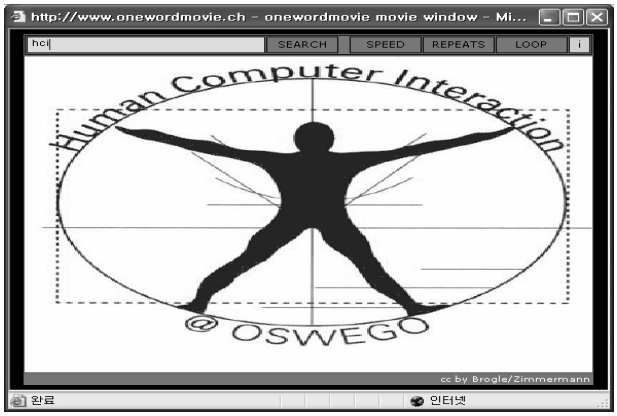
1. 서 론

방송, 영화 기술의 발달과 디스플레이 장치의 보급으로 현대 사회에서 영상은 매우 보편적인 정보 전달 매체가 되었다. 또한, 인터넷 등 네트워크 기술의 발달과 디지털 카메라, 캠코더, 컴퓨터 등 하드웨어의 발달과 보급으로 사용자는 언제 어디서나 손쉽게 영상에 접근할 수 있다. 하지만 우리가 일상생활에서 쉽게 접할 수 있는 공중파 방송, 영화 등 재래식 영상 미디어에서는 정보 전달이 일방적이고 획일적이기 때문에, 사회가 복잡해지고 가치관이 다양해짐에 따라 정보에 대한 욕구도 다양해진 현대 사회에는 적합하지 않다.

이에 따라 다양한 가치관과 욕구에 부응하는 새로운 미디어에 대한 요구가 생기게 되었고, 이러한 요구의 발산이 인터넷 등 네트워크 기술의 발달과 카메라, 각종 영상 취득 장비의 보급 및 발달과 맞물려 새로운 차원의 미디어(뉴미디어) 형태로 나타나고 있다. 이러한 뉴미디어는 사용자 개개인에게 맞춤화되고, 최적화된 환경을 제공하는 것을 가장 중요시한다. 영상을 만든 사람이 영상을 보는 사람에게 일방적으로 메시지를 전달하는 것이 아니라, 개인의 특성과 상황에 따라 다른 영상을 제공받을 수 있고, 이러한 영상을 사용자가 원하는 방향으로 쉽게 가공하고 감상할 수 있어야 한다.



< 그림 1 > 소프트 시네마



< 그림 2 > 한 단어 영화, 질의를 hci 로 한 경우

본 논문에서 제안하는 실시간 몽타주 시스템은 이러한 뉴미디어와 같이 사용자와의 상호작용이 가능하고, 쉽고 빠르게 영상 재가공을 할 수 있다. 내용 기반 영상 검색 방법을 사용하여 정확하게 영상을 검색해서 사용자의 의도대로 영상을 재구성하고, 실시간 움직임 감시를 통해 사용자의 영상 또한 개입이 가능하도록 함으로써 누구든지 새로운 영상을 만들 수 있다.

본 논문에서는 이와 같이 상호작용이 가능하고 영상 재가공이 용이한 실시간 몽타주 시스템에 대해 기술하고자 한다. 본 논문은 2 장에서 관련 연구 및 작품에 관해 설명하고, 3 장에서 구현된 시스템에 대해 살펴보며, 마지막 4 장에서 결론을 맺는 순서로 기술되어 있다.

2. 관련 연구 및 작품

2.1 소프트 시네마

소프트 시네마는 “영화는 반드시 필름을 통해 찍어야 한다거나, 디지털 카메라로 촬영된 영화가 필름영화와 똑 같은 룩(look)을 재현할 필요는 더 이상 없다”라는 관점에서 나타나게 된 뉴미디어의 대표적 형식이다. 소프트 시네마는 <그림 1>과 같이 제작자가 가지고 있는 기존의 영상을 재가공(편집, 특수효과 등)해서 새로운 작품을 만들어 내는 것으로, 세계 각지에서 많은 제작자들이 관련 연구 및 작품활동을 하고 있다[1].

하지만, 제작 방식 자체가 전문가가 아니면 접근하기 어려울 뿐만 아니라, 일단 제작된 작품은 사용자 측면에서는 감상하는 것 이외에 상호작용이 전혀 없고, 영상을 재가공하는 것이 거의 불가능 하기 때문에, 이러한 한계를 극복하기 연구가 활발히 진행되고 있다.

2.2 한 단어 영화(onewordmovie)

한 단어 영화는 텍스트 기반 영상 검색엔진을 이용해서, 사용자가 질의하는 영상을 순차적으로 보여줌으로써, 사용자에게 마치 단어에 해당되는 영상들이 이어지는 영화를 보는 듯한 감상을 주는 영상기반 몽타주 시스템이다. 한 단어 영화는 인터넷에서 가장 대중적으로 쓰이는 검색엔진(구글 등)의 이미지 검색을 이용해 데이터베이스를 구성하였고, 사용자가 텍스트로 질의를 하게 되면 DB 에서 그에 맞는 영상을 검색해서, 이에 해당하는 영상들을 미리 결정된 우선 순위에 의해 순차적으로 보여준다. <그림 2>는 질의로 “hci”를 입력했을 때의 모습으로 그림과 유사한 제목을 가지고 있는 (hci 를 포함하는)여러 정지 영상들이 순차적으로 나타난다[2].

한 단어 영화는 2004 미디어 시티 서울 및 네덜란드, 독일, 싱가폴, 일본 등 세계 각지에서 전시되어 호평을 받은 바 있다. 하지만, 텍스트 기반 검색 자체가 영상을 만든 사람의 주관성이 많이 개입되고, 미리 구성된 데이터베이스의 영상만을 이용하기 때문에 사용자의 상호작용이 불가능 하다는 근본적인 한계를 가지고 있다.

3. 실시간 몽타주 시스템

본 논문에서 제안하는 시스템은 크게 3 부분으로 구성된다. 사용자가 선택한 영상을 이용해 유사한 영상을 검색하는 내용 기반 영상 검색 부분과 사용자의 영상 삽입을 위한 실시간 움직임 감시 부분, 그리고 결과 영상을 보여주는 부분으로 구성되어 있다. 본 장에서는 내용 기반 영상 검색 부분과 실시간 움직임 감시 부분을 따로 나누어 설명하고, 결과 영상을 보여주는 부분은 실시간 몽타주 시스템에서 함께 기술하였다.

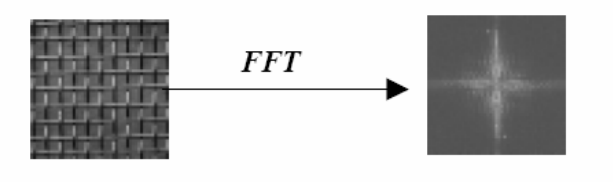
3-1. 내용 기반 영상 검색

현대의 급변하는 정보화 사회를 대변하는 가장 큰 현상은 폭발적인 양의 디지털 정보들의 범람과 계속되는 증가현상이다. 이러한 방대한 정보들 중 사용자가 원하는 정보를 쉽고, 빠르고, 정확하게 얻을 수 있는 방법을 연구하는 것이 정보 검색(IR: Information Retrieval)이다[3].

디지털 형태로 변환되고 저장되는 이미지, 즉 영상 자료들에 대한 검색 방식은 문서를 검색하는 방식과는 접근 방법이 크게 다르다. 텍스트에 기반하여 이미지를 표현하고 검색하고자 했던 과거의 이미지 검색 방식은 사용자에게 직관적이지 못하고, 영상을 텍스트로 표현한다는 것 자체가 주관성의 개입여지가 매우 많기 때문에 검색에는 적합하지 않다[4].

그렇기 때문에 이미지의 특징을 직접적으로 이용해서 검색하는 내용 기반 영상 검색(CBIR : Contents Based Image Retrieval) 기술에 대한 연구가 활발히 진행되고 있다. 이미지의 특징을 직접적으로 표현하는 방식은 크게 색상(Color), 형태(Shape), 질감(Texture) 나눌 수 있다[5].

이 중 색상과 형태를 이용한 방식은 영상의 종류(그림, 사진 등)에 따라 매우 다른 결과를 보여주는 경우가 있기 때문에, 본 시스템에서는 여러 영상에 적용이 용이하고 구현이 간단한 질감 기반 검색 방식을 이용하였다. 질감 기반 검색 방식에도 여러 가지가 있지만, 효과적인 질감 기반 검색 기법으로 널리 알려진 주파수 영역으로의 변환을 통한 비교 방식을 사용하였다.



< 그림 3 > Fast Fourier Transform

사용자가 선택한 영상과 데이터베이스에 있는 영상들을 주파수 영역으로 변화시키기 위해서는 일반적으로 식 (1)과 같이 2 차원 푸리에 변환(Fourier Transform)을 이용하게 된다.

$$F(u,v)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,e^{-2\pi i\,(ux+uy)}\,dx\,dy \quad (1)$$

본 시스템에서 사용하는 영상은 디지털화 된 영상이기 때문에 식 (2)와 같이 푸리에 변환의 이산적 형태인 이산 푸리에 변환(Discrete Fourier Transform)을 사용해야 하지만 계산의 효율성을 위해서 이산 푸리에 변환의 반복 연산을 효율적으로 개선한 고속 푸리에 변환(Fast Fourier Transform)을 이용해 영상을 주파수 영역으로 변환한다.

$$F(u,v)=\frac{1}{N^2}\sum_{-\infty}^{N-1}\sum_{-\infty}^{N-1} f(x,y)\,e^{-2\pi i\,(\frac{ux}{N}+\frac{vy}{N})} \quad (2)$$

<그림 3>은 원 영상을 FFT 를 이용해 주파수 영역으로 변환한 영상이다.

이와 같이 주파수 영역으로 변환시킨 영상들을 다시 N x N 크기의 영상으로 변환시킨 후, 식 (3)을 이용해 b x b 크기의 영역으로 나누고, 각 영역의 명암 값의 합(B_n)을 구한다.

$$B_n=\sqrt{\sum_{n=0}^{\frac{N^2}{b}-1}\sum_{u=nb}^{nb+b-1}\sum_{v=nb}^{nb+b-1} F(u,v)} \quad (3)$$

각 영역의 명암 값의 합을 다른 영상과의 비교를 정확히 하기 위해 식 (4)를 이용하여 정규화(Normalization)하면, 이로부터 영상 전체에서 해당 영역이 차지하는 비중(C_n)을 구할 수 있다.

$$A_n = \sum_{n=0}^{\frac{N^2}{b}-1} B_n, \quad C_n = \frac{B_n}{A_n} \tag{4}$$

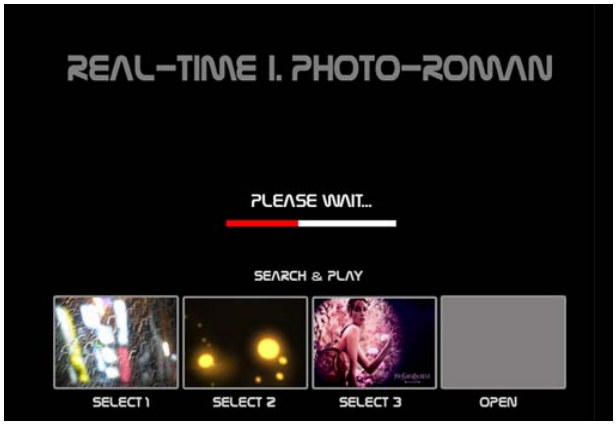
이와 같은 과정을 사용자가 선택한 기준 영상과 데이터베이스에 있는 모든 영상에 적용한 후, 기준 영상과 데이터베이스의 영상간의 유사도 비교를 한다. 유사도는 식 (5)와 같이 유클리디안 거리(euclidean distance) 를 이용해 빠르고 효과적으로 비교하게 된다. 비교 결과 유사한 영상일수록 유클리디안 거리가 작게 되고, 이 순서대로 영상 순서를 결정하게 된다.

$$D_n = \sum_{n=0}^{\frac{N^2}{b}-1} C_n{}^q - C_n{}^d \tag{5}$$

3-2. 실시간 움직임 감시

컴퓨터 비전(Computer Vision) 분야에서 가장 활발하게 연구되고 있는 분야 중 하나가 영상 트래킹(tracking) 부분이다. 물체 인식, 움직임 추출, 영상 검색 등 다양한 분야에서 연구, 활용하고 있다[6]. 본 논문에서는 사용자의 영상의 삽입 시기를 결정하는 방법으로 차분영상 기반 영상 트래킹을 이용한다.

본 논문에서는 사용자가 움직일 때마다 사용자의 영상을 삽입하게 되는데, 움직임을 파악하기 위해서는 카메라로 입력되는 영상의 변화를 감지해야 한다. 식 (6)과 같이 현재 영상과 이전 영상 각 화소의 명암 값을 비교해서 그 차이가 일정 한계치 이상이면 해당 화소에 변화가 있는 것으로 감지하도록 하였다.



< 그림 4 > 실시간 몽타주 시스템

< 그림 5 > 실시간 몽타주 시스템 : 검색화면

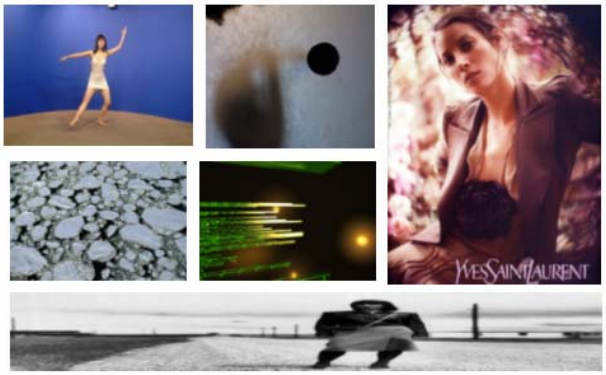
$$\begin{aligned} &if \quad |f(x,y) - F(x,y)| > 30, \quad P(x,y) = 1 \\ &else \quad P(x,y) = 0 \end{aligned} \tag{6}$$

식 (7)을 이용해서 한계치를 넘는 픽셀의 수(T)가 사용자가 설정한 수 보다 많게 되면, 사용자가 움직인 것으로 추정해 이 영상을 실시간으로 저장하고, 저장된 영상을 화면에 보여준다.

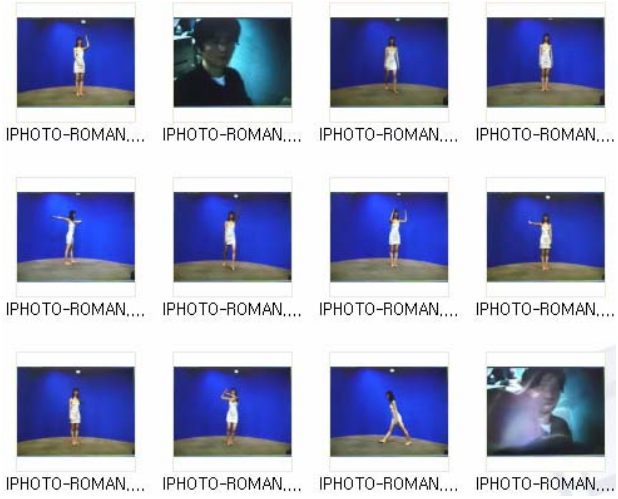
$$T = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} P(x,y) \tag{7}$$

3-3. 실시간 몽타주 시스템

본 논문에서 제안하는 실시간 몽타주 시스템은 < 그림 4 >와 같이 구성되어 있다. 사용자가 검색을 하고자 하는 기준 영상을 선택하고, 검색을 원하는 데이터베이스를 선택하게 되면, 자동으로 선택된 영상과 유사한 영상을 데이터베이스에



< 그림 6 > DB 를 구성하고 있는 영상들



< 그림 7 > 시스템 실행화면

서 검색해서 유사한 영상일수록 우선순위를 주게 된다. 질의 영상과 데이터베이스의 영상은 256 x 256 크기로 일정하게 변화시켰고, 비교를 위해 분할하는 영역의 크기는 16x16 으로 하였다.

또한 사용자의 움직임을 실시간으로 감시하는 부분에서는 변화된 화소의 양이 전체의 20%가 되면 움직임이 있는 것으로 판단하도록 하였고, 사용자가 언제든지 수치를 조절할 수 있도록 하여 사용자 움직임 모습의 입력 정도를 변경할 수 있도록 하였다. <그림 5>의 왼쪽 모니터 영상은 검색 화면이고, 오른쪽 화면은 사용자의 움직임을 감시하는 화면이다.

검색이 끝나고 몽타주 시스템을 실행시키면 <그림 6>과 같이 구성된 8 개 카테고리 1500 개의 데이터베이스 영상 중 검색된 영상을 우선 순위에 따라 순차적으로 보여주고, 이 때 사용자가 움직이기 되면 사용자의 영상을 실시간으로 검색된 영상 사이에 삽입하게 된다.

<그림 7>은 시스템 실행시 순차적으로(약 0.3 초 간격) 보여주는 영상을 나열한 것이다. 가장 처음 영상이 질의로 사용된 영상이고, 이후의 영상은 검색된 영상이다. 사용자의 움직임이 있을 때마다 검색된 영상 사이에 사용자의 모습이 삽입된 것을 알 수 있다. 이를 통해 사용자는 검색된 영상과 사용자의 영상이 마치 하나의 시퀀스로 연결된 것처럼 보이게 되는 몽타주 효과를 느낄 수 있다. 사용자는 자신의 의도에 따라 영상을 선택해 유사한 영상의 시퀀스를 구성할 수 있게 되고, 자신의 영상까지 원하는 시기에 삽입시킴으로써 상호작용을 극대화할 수 있으며, 또한 사용자가 소유하고 있는 영상을 언제든지 다른 영상으로 재가공할 수 있다.

4. 결 론

본 논문에서는 내용기반 영상검색과 실시간 움직임 감시를 통해 기존의 영상들을 새로운 영상으로 재구성하는 실시간 몽타주 시스템에 대해 기술하였다. 기존의 재래식 미디어와는 사용자의 선택에 의해 다른 영상들을 쉽게 만들 수 있고, 사용자의 영상까지 쉽게 개입시킬 수 있도록 하였으며, 내용 기반 영상 검색 방법을 이용해 보다 정확하게 원하는 영상을 찾아 재구성이 가능하도록 하였다.

실시간 몽타주 시스템은 사용자가 일방적으로 영상을 시청, 감상하는 소극적 영상의 소비자에서 벗어나, 영상 재생산의 적극적 주체가 되는 사용자 중심의 새로운 영화(뉴미디어)의 토대가 될 것으로 기대된다.

※ 본 과제(결과물)는 교육인적자원부, 산업자원부, 노동부의 출연자금으로 수행한 최우수실험실지원사업의 연구 결과입니다.

참고 문헌

[1] Lev Manovich, Soft Cinema, http://www. Softcinema. net

[2] Beat Brogle, Philippe Zemmermann, One Word Movie(onewordmovie), http://www.onewordmove.ch

[3] Glenn Becker. Information in Images, chap 6: Content-based Query of Image Databases. Thomson Technology Labs, 1997. Online book. http://www.thomtech.com:80/mmedia/tmr97/tmr97.htm.

[4] Kjersti Aas and Line Eikvil. A survey on: Content-based access to image and video databases. Report 915, Norwegian Computing Center, March 1997. http://www.nr.no/home/kjersti/video.html.

[5] Atsuo Yoshitaka and Tadao Ichikawa. A survey on content-based retrieval for multimedia databases. IEEE Transactions on Knowledge and Data Engineering, 11(1):81–93, January/February 1999.

[6] J.K. Aggarwal and Q. Cai. Human motion analysis: a review. Computer Vision and Image Understanding, 73(3):295–304, 1999.