영상 초록 구현을 위한 키프레임 추출 알고리즘의 설계와 성능 평가*

Design and Evaluation of the Key-Frame Extraction Algorithm for Constructing the Virtual Storyboard Surrogates

김 현 희(Hyun-Hee Kim)**

초 록

본 연구에서는 비디오의 의미를 잘 표현하고 있는 키프레임들을 추출하는 알고리즘을 설계하고 평가하였다. 구체적으로 영상 초록의 키프레임 선정을 위한 이론 체계를 수립하기 위해서 선행 연구와 이용자들의 키프레임 인식 패턴을 조사하여 분석해 보았다. 그런 다음 이러한 이론 체계를 기초로 하여 하이브리드 방식으로 비디오에서 키프레임을 추출하는 알고리즘을 설계한 후 실험을 통해서 그 효율성을 평가해 보았다. 끝으로 이러한 실험 결과를 디지털 도서관과 인터넷 환경의 비디오 검색과 브라우징에 활용할 수 있는 방안을 제안하였다.

ABSTRACT

The purposes of the study are to design a key-frame extraction algorithm for constructing the virtual storyboard surrogates and to evaluate the efficiency of the proposed algorithm. To do this, first, the theoretical framework was built by conducting two tasks. One is to investigate the previous studies on relevance and image recognition and classification. Second is to conduct an experiment in order to identify their frames recognition pattern of 20 participants. As a result, the key-frame extraction algorithm was constructed. Then the efficiency of proposed algorithm(hybrid method) was evaluated by conducting an experiment using 42 participants. In the experiment, the proposed algorithm was compared to the random method where key-frames were extracted simply at an interval of few seconds(or minutes) in terms of accuracy in summarizing or indexing a video. Finally, ways to utilize the proposed algorithm in digital libraries and Internet environment were suggested.

키워드: 키프레임 추출 알고리즘, 영상 초록, 비디오, 디지털 도서관, 키프레임, 하이브리드 방식, 랜덤 방식, 요약, 색인
image, video, storyboard, surrogate, sense making

---

* 이 논문은 2007년도 정부재원(교육인적자원부 학술연구조성사업비)으로 한국학술진흥재단의 지원을 받아 연구되었음(KRF-2007-327-H00017).

** 명지대학교 인문대학 문헌정보학과 교수(kimhh@mju.ac.kr)

■ 논문접수일자 : 2008년 11월 14일 ■ 최초심사일자 : 2008년 11월 17일 ■ 게재확정일자 : 2008년 11월 27일
■ 情報管理學會誌, 25(4): 131-148, 2008. [DOI:10.3743/KOSIM.2008.25.4.131]

www.kci.go.kr

1. 서 론

1.1 연구 필요성과 목적

최근 우리 주위의 정보 매체는 종이책에서 멀티미디어 자료로 급격하게 변화하고 있다. 이와 같이 정보 매체는 멀티미디어 형식으로 바뀌어가고 있는데 정보 매체에 접근하기 위해 사용되는 메타데이터(주로 자동 생성되고 검색을 위해서 사용)와 초록, 요약 등과 같은 서로게이트(surrogate)(주로 사람에 의해서 구성되며 원자료의 압축된 표현으로 맥락 파악(sense making)을 위해서 사용)는 여전히 텍스트 기반 자료가 대다수를 이룬다. 특히 비디오는 의미를 결정하는 다양한 특성(예, 오디오 또는 비주얼 채널)을 갖고 있는 대용량 자원이기 때문에 비디오 전체 클립을 보기 전에 적합성 판정을 위한 좀 더 세밀한 브라우징 과정이 필요하다(Kristin et al. 2006).

비디오 브라우징 단계에서 주로 사용하고 있는 텍스트 기반의 메타데이터 또는 서로게이트만으로는 비디오 영상물에 대한 정확한 정보를 얻기 힘든 상황이며 멀티미디어 자료에 익숙해진 세대에게는 의미 전달이 어려운 경우도 생겨나고 있다. 비디오 자료에 대한 적합성 판정은 이용자에 따라 많은 차이가 있으므로 비디오 자료를 검색하고 걸러내기 위해서 다양한 메타데이터와 멀티미디어 기반 초록이 필요하다는 의견도 있다(Yang 2005).

비디오 영상 초록이 전통적인 문헌에서 초록이 하는 것과 같은 기능을 하고 있고 또한 비디오의 키프레임이 전통적인 문헌의 키워드와 유사한 기능을 하고 있음을 실험을 통해서 증명하는 연구도 나오고 있다(Lyer and Lewis 2007). 이와 같이 기존 연구들이 영상 초록의 중요성을 제안하고 있지만 영상 초록의 품질을 결정하는 가장 중요한 요인 중 하나는 비디오의 의미 전달에 필요한 키프레임을 선정하는 과정에 대해서는 심도 있게 다루지 않고 있다. 이제는 텍스트 기반의 초록 및 색인 이론 처럼 비디오의 키프레임을 선정하는 과정에 대한 이론 정립과 방법론에 대한 연구가 필요한 때이다.

현재 영상 초록은 전체 비디오 클립에서 일정한 시간 간격에 따라서 키프레임들을 추출하거나 이미지 프로세싱을 이용하여 장면을 구분하고 각 장면에서 중간에 나타나는 프레임을 키프레임으로 추출하는 방식을 채택하고 있다(Browne and Smeaton 2005). 또한 비디오를 브라우징하면서 키프레임을 선정하는 수작업 방식을 채택하기도 한다. 이와 같이 영상 초록을 구성하는 모든 과정이 수작업으로 진행될 때 급속도로 증가하고 있는 비디오 자료를 처리하는데 어려움이 예상된다. 아울러 이미지 프로세싱 방법에만 의존하여 영상 초록을 구성하기에는 아직도 기술적인 어려움이 많다. 또한 지나치게 기술적인 측면을 강조하다 보니 이용자가 영상 초록의 맥락 파악 및 적합성 판정 과정에서 어떤 방식으로 수용하는지에 대한 연구가 부족한 것이 현실이다.

본 연구의 목적은 비디오 자료에 대한 적합성 기준, 이미지 인식 및 분류 체계 그리고 이용자의 키프레임 인식 패턴을 활용하여 반자동(하이브리드) 방식으로 비디오에서 키프레임을 추출하는 알고리즘을 설계, 평가하여 효율적인 영상 초록을 구현하는 방안을 제안하는데 있다.

www.kci.go.kr

1.2 연구 방법

키프레임 추출 알고리즘을 설계하고, 그 효율성을 평가하기 위해서 사용된 연구 방법은 다음과 같다. 첫째, 키프레임 추출 알고리즘을 설계하기 위한 이론적 토대를 수립하기 위해서 선행 연구를 통한 문헌 분석과 이용자의 키프레임 인식 패턴을 조사한다. 키프레임 인식 패턴의 조사를 위해서 M대학의 20명의 대학생들을 선정하여 12개의 표본 비디오의 영상 초록을 보고 요약문을 적게 하였다. 그런 다음 의미 추출에 도움을 준 키프레임(들)과 선택된 키프레임(들)의 어떤 내용이 도움을 주었는지도 기술하게 하였다. 이러한 문헌 분석 및 키프레임 인식 패턴의 분석 결과를 활용하여 키프레임 추출 알고리즘을 설계한다.

둘째, 설계될 알고리즘의 성능을 평가하기 위해서 그 알고리즘을 활용하여 구성한 영상 초록(하이브리드 방식)을 일정한 시간 간격에 따라서 무작위로 키프레임을 추출하여 구성한 영상 초록(랜덤 방식)과 비교해 보고자 한다. 이 두 알고리즘들을 비교하기 위해서 두 개의 실험 시스템을 구현하였고, 이 두 실험 시스템에 사용될 비디오는 키프레임 인식 패턴의 분석을 위해서 사용한 것과 동일한 12개의 표본 비디오들이다.

피조사자 집단은 동일한 집단을 구성하기 위해서 J대학의 인문대학의 학부생을 42명 선정한 후 학과, 학년, 남녀 등을 고려하여 되도록 고르게 분포된 2개의 그룹을 구성한다. 그룹 1(21명)에게는 랜덤 방식에 의해서 구성한 영상 초록들을 보게 할 것이고, 그룹 2(21명)에게는 제안된 하이브리드 방식에 의해서 구성한 영상 초록을 보게 하여 그 차이를 요약문 및 색인어 추출의 정확도 측면에서 비교하고자 한다. 실험 결과를 분석하기 위해서 SPSS 12.0(ver.)을 이용하여 중다변량분석(MANOVA)과 t-검정을 수행한다.

2. 이론적 토대

비디오의 키프레임 추출 알고리즘 설계를 위한 사전 작업인 이론적 토대를 수립하기 위해서 먼저 멀티미디어 자료에 대한 적합성 기준과 이미지 인식 및 분류에 대한 선행 연구를 살펴본다. 그런 다음 이용자들이 비디오의 의미를 결정할 때 사용된 키프레임들을 분석해 봄으로써 키프레임 인식 패턴을 알아보고자 한다.

2.1 선행 연구

2.1.1 멀티미디어 자료에 대한 적합성 기준

Choi와 Rasmussen(2002)은 이미지 검색에서 이용자의 적합성 기준에 대하여 조사하기 위해서 38명의 피조사자들에게 미의회도서관의 “American Memory”의 이미지 자료를 검색하도록 하였다. 이들은 이용자들이 주제, 정확성, 이미지의 기술적 속성 등 9개의 적합성 기준을 활용하고, 탐색 단계에 따라서 이용자들의 적합성 기준에 대한 선호도가 달라지고 있음을 발견했다. 즉, 정보 문제를 정의하는 단계에서 중요한 기준이 주제, 정확성, 완전성, 제시성, 시간틀(time frame)순으로 나타난 반면 이미지를 본 이후에는 주제, 시간틀, 접근성, 정확성, 완전성순으로 나타났다. 또한 이들은 이

www.kci.go.kr

미지 검색을 위한 적합성 기준의 독특한 특징을 제안하였다. 예를 들어서, 전거(authority)가 텍스트 정보의 적합성 연구에서 보다 덜 중요한 것으로 나타났으나 주관성(subjectivity)과 이미지에 대한 감정적 작용(affectiveness)은 자료 선택 단계에서 매우 중요한 요인으로 나타났다. 이러한 연구 결과는 Markkula와 Sormunen(1998)의 시각 속성들이 마지막 선택 단계에서 중요한 요인이라는 주장을 재확인시켜 준다.

이러한 연구를 기초로 하여 Yang과 Marchionini(2004)는 비디오 적합성 판정을 위한 기준들을 제안하였다. 이 기준들은 세 개의 카테고리 즉, 텍스츄얼 기준(주제, 날짜, 국적, 장르 등), 비주얼 기준(촬영기술, 객체/이벤트, 기타 등) 및 함축적인 기준(흥미, 친숙성, 접근성 등)으로 구분된다. 비디오 자료에서 주제의 중요도는 텍스트 자료에서 보다 훨씬 낮았지만 주제는 여전히 가장 많이 사용하는 기준이었다. 끝으로 Lyer와 Lewis(2007)는 비디오의 영상 초록을 체크하여 적합성을 판정할 때 객체 인식이 기본적인 도구가 됨을 실험적으로 보여주었다. 따라서 확인된 객체들이 부정확하거나 비디오의 주제에 맞지 않을 때 적합성 판단을 잘못하는 경우가 종종 발생한다고 보고하였다. 이들 연구 결과를 Yang과 Marchionini(2004)가 분류한 세 가지 카테고리에 따라서 분류하면 〈표 1〉과 같다.

이들 연구를 종합해 볼 때 주제가 텍스트 자료에서 보다는 비중이 많이 작아졌지만 비디오 자료의 적합성 판단을 하기 위해서 여전히 가장 중요한 요인이었다. 더 나아가 비디오의 내용을 구성하는 핵심 객체에 대한 인식이 중요한 것으로 나타났다.

2.1.2 이미지 인식 및 분류

Panofsky(1955) 이론에 기초하여 Shatford(1986)은 영상에 대한 사람들의 이해 수준을 "일반적인 것(pre-iconography)," "특정적인 것(iconography)" 및 "추상 및 감정적인 것(iconology)"의 세 단계로 보았다. Chung과 Yoon(2008)은 Shatford(1986)의 이미지의 카테고리 스킴을 다음과 같이 정리하였다. 즉, 이미지를 크게 일반적인 것, 특정한 것, 추상적인 것 및 기타로 구분하고 이를 다시 다음과 같이 세분하였다. 즉, 일반적인 것과 특정한 것의 객체, 이벤트/활동, 장소 그리고 시간 구분은 동일하다. 다만 차이는 일반적인 것은 어떤 이벤트를

〈표 1〉 적합성에 대한 선행 연구 요약

기 준 / 선행 연구	텍스츄얼 기준	비주얼 기준	함축적인 기준
Choi & Rasmussen(2002)	주제, 시간틀	이미지의 기술적인 속성(무드, 감정, 색상 및 관점)	정확성, 완전성, 접근가능성, 흥미, 신기로움, 암시성
Yang(2005)	주제, 날짜, 국적, 저작, 장르, 재생시간, 기타(리뷰, 가격 등)	촬영기술, 객체/이벤트, 활동, 스타일, 기타(색상, 사운드, 감정 등)	흥미, 친숙성, 접근성, 적절성, 암시성
Lyer와 Lewis(2007)	-	객체, 활동	-

www.kci.go.kr

기술할 때 단순히 "전투"로 기술하는 반면 특정한 것은 전투 중에서 어떤 전투를 의미하는지 "게티스버그 전투"와 같이 자세하게 기술하는 것을 의미한다. 추상적인 것은 추상적인 객체, 감정/추상, 추상적인 장소 및 시간으로 세분화시키고 있다.

Greisdorf와 O'Connor(2002)는 Panofsky의 세 단계와 유사한 세 단계 계층에서 사람들은 이미지/비디오와 상호작용한다고 기술하고 있다. 가장 기본적인 수준에서, 이미지의 기본 특징(색상, 형태 등)이 인지되고, 두 번째 수준에서는, 객체(사람, 장소, 활동)가 인지된다. 세 번째 수준은 감정적인 단서와 분위기를 포함해 추상적인 특징에 관한 추론을 가진 이미지/비디오의 귀납적인 설명을 요구한다고 하였다.

Laine-Hermandez와 Westman(2008)은 잡지에 기술되는 이미지의 분류 카테고리로 크게 10개 즉, 기능, 씬, 주제, 스토리(story), 기술(description) 등을 사용한다. 좀 더 구체적으로 기능은 초상화, 보도 등으로 세분되고, 씬은 풍경, 자연 등으로 구분된다. 또한 주제는 음식과 음료, 작업 등으로 구분되고 스토리는 이벤트, 시간 및 활동으로 구분된다. 끝으로 기술은 속성과 수로 세분되고 비주얼은 색상, 구성 등으로 세분화된다. 이들 연구들의 내용을 논리적인 것, 추상적인 것 및 기타로 분류하여 〈표 2〉에 정리하였다.

2.2 이용자의 키프레임 인식 패턴

영상 초록에서 의미 추출에 도움을 준 키프레임(들)의 특성을 조사하기 위해서 M대학의 20명의 대학생들을 선정하였다. 피조사자들을 선정한 후 〈표 3〉에 나와 있는 12개의 교육 비디오에 대한 영상 초록(스토리보드)을 보게 한 후 요약문을 적게 한 다음 의미 추출에 도움을 준 키프레임(들)을 제시하도록 하였다. 아울러 선택된 키프레임(들)의 어떤 내용이 도움을 주었는지도 기술하게 하였다. 표본 비디오는 노스캐로리나 대학의 문헌정보학과에서 구축한 Open Video Digital Library(OVDL, www.open-video.org)에서 총 12개의 교육 비디오를 선정하였다(Marchionini and Geisler 2002).

〈표 2〉 이미지 분류 체계에 대한 선행 연구 요약

분류 기준 / 선행 연구	논리적인 것 (일반적인 것/특정적인 것)	추상적인 것	기 타
Panofsky(1955), Shatford (1986), Chung과 Yoon(2008)	객체, 이벤트/활동, 장소 및 시간	객체, 감정/추상, 장소 및 시간	색상, Flickr와 관련된 것
Greisdorf와 O'Connor (2002)	객체(사람, 장소, 활동)	귀납적 해석(상징적 가치, 감정적인 단서, 분위기)	이미지의 기본 특징(색상, 형태, 질감)
Laine-Hermandez & Westman(2008)	사람, 객체, 씬(풍경, 자연, 인테리어), 스토리(이벤트, 시간 및 활동)	감정, 주제(음식과 음료, 작업, 스포츠, 여행, 예술 및 종교 등)	비주얼(색상, 구성, 활동, 형태), 기능(초상화, 보도, 광고), 기술(속성, 수), 사진술(거리, 흑백, 스타일)

www.kci.go.kr

〈표 3〉 표본 비디오 리스트(1)

제 목	
Food Preservation	Exercise and Nutrition
Benefits of Sleep	Earthquake
First Flying Machines	Computer Rage
Ubiquitous Computing in the Living Room	Bicycle Today - Automobile Tomorrow
Flexible Wing	NASA Connect - PSA - Robots
Civil War, The	GPS for Pilots

피조사자의 답변들을 〈표 4〉에서 처럼 크게 두 가지 즉, 텍스츄얼 기준과 비주얼 기준으로 구분하였다. 텍스츄얼 기준은 자막/이미지속의 글자가 포함되고 비주얼 기준에는 객체, 이벤트/활동, 인물, 씬, 기호/도면 등이 포함되었다. 〈표 4〉에 의하면 자막/이미지속의 글자(26.9%)를 근거로 하여 가장 많이 주제를 결정하고 있다. 자막/이미지속의 글자가 있는 키프레임이 전체 키프레임의 약 11%임을 감안하면 자막/이미지속의 글자에 의존하는 실제 비율은 이 보다 훨씬 높다고 볼 수 있다. 이는 비주얼 정보가 풍부하고, 생생한 구체적인 정보를 제공하는 반면 텍스트 정보가 비디오에 대한 직접적이면서 논리적인 정확한 정보를 제공한다는 Ding 등(1999)의 주장과 텍스트 정보가 비디오의 내용에 대한 정보를 전달한다는 Hughes 등(2003)의 주장과도 일치되는 결과이다.

또한 앞의 선행 연구에서 증명된 것처럼 객체(20.2%), 이벤트/동작(16.3%)도 비디오의 의미 파악을 위한 중요한 요인들임을 보여주고 있다. 이는 달리 말하면, 확인된 자막/이미지속의 글자, 객체, 이벤트/활동이 부정확하거나 비디오의 주제와 맞지 않을 때 적합성 판단을 잘 못하는 경우가 종종 발생할 수 있음을 보여 주고 있다. 한 예로 유리병과 캔에 음식을 저장하는 방법의 발전 과정을 설명하는 비디오에서 한 키프레임에 캔이 있고 그 캔에 "chocolate"이라는 단어가 쓰여져 있는데 몇몇 피조사자들이 그 비디오의 주제를 초콜렛에 관한 이야기이라고 잘못 추측하고 있는 경우가 여기에 해당된다.

일부 피조사자들은 둘 이상의 키프레임들을 서로 연결시키셔 주제 파악을 하고 있음을 알 수 있었다. 링컨과 군인들의 이미지들을 보고 남북전쟁이 주제라고 생각하는 경우 등이 여기에 해당된다. 이는 텍스트 기반 초록에서 키워드와 키워드간의 문맥상의 관계를 통해서 의미 파악을 하는 것과 같은 원리로 볼 수 있다. 이러한 연구 결과는 비디오 유형에 따라서 키프레임을 추출할 때 쌍으로 또는 그룹으로 키프레임들을 추출하는 것이 효율적일 수 있음을 보여주고 있다. 끝으로 키프레임의 색상이 의미 파악을 하는데 도움을 주는 경우도 많지는 않지만 2건이 있었다. 이 경우는 색상이 있는 키프레임들 중에서 낡은 흑백 키프레임은 특정 객체의 초기 모형에 대한 설명 또는 특정 객체에 대한 역사적인 기록에 관한 것임을 유추할 수 있기 때문으로 생각된다. 분석 결과를 종합해 볼 때 특정 유형의 프레임이 비디오의 의미 추출에 더 효율적일 것으로 나타났다.

www.kci.go.kr

〈표 4〉 키프레임 인식 패턴의 분석 결과

주 카테고리	서브 카테고리	사례수 (비율)	실제(예)
텍스츄얼 기준	자막/ 이미지속의 글자	28 (26.9%)	1)키프레임 속에 "changing places"라는 문구가 집에 관한 어떤 변화를 보여 줄것이라고 생각함 2)"enough sleep"이라는 문구를 보고 충분한 수면의 중요성을 말하고자 하는 것이라고 파악했음 3)캔에 쓰여진 chocolate를 보고 초콜렛에 관한 이야기일 것이라고 추측함(잘못 인식)
비주얼 기준	객체	21 (20.2%)	1)자전거 사진으로 자전거에 대한 내용이라는 것을 알 수 있었음 2)TV가 나온 것을 보고 이 기계와 TV와 비교하여 더 낫다는 것을 알 수 있었음 3)글라이더에 달려 있는 기구가 마치 새로운 연구에 이용되는 것 같아 보였음
	이벤트/ 활동	17 (16.3%)	1)컴퓨터를 부수는 장면이 동물을 이용해 컴퓨터에 대한 실험을 하는 것 같음 2)부모가 아이를 급하게 잡는 모습을 통해서 안전사고와 관련된 내용으로 파악함 3)테이블 밑으로 숨는 것을 보고 지진임을 알 수 있었음
	이미지간의 연관성	15 (14.4%)	1)링컨과 군인들의 이미지들을 보고 남북전쟁이라고 생각함 2)풀을 보면서 우울한 표정을 짓는 것과 운동 후 표정을 보면서 심한 다이어트로 건강을 해친다고 생각함 3)키프레임속의 뇌모양과 잠자는 모습을 통해서 뇌의 활동과 수면과의 관계에 대한 주제라고 추축함
	인물	11 (10.6%)	1)링컨의 모습을 보고 남북전쟁과 관련된 것이라고 생각함 2)무중력 상태에서 우주인이 식사하는 모습을 보고 우주에서의 음식과 건강에 관련된 것이라고 생각함 3)키프레임속의 인물이 나폴레옹이므로 나폴레옹과 프랑스에 관련된 것임을 알 수 있음
	씬	6 (5.8%)	1)휘어진 시설물과 굴곡이 생긴 도로를 보고 지진에 대한 것일 수 있다고 생각함 2)무언가 생소한 물건을 들고 있는 남자와 뒷배경을 보고 첨단과학기술에 대한 것일 수 있다고 생각함
	기호/도면	4 (3.8%)	1)비행 물체의 도면을 통해서 과학적인 제작이 필요하다는 것을 알 수 있음 2)가스 밸브를 잠그는 방식이 도식으로 표현되어 가스사고나 기타 사고를 예방하기 위한 것이라고 생각함
	색상	2 (2.0%)	1)옛날 사진인 듯한 흑백사진에 사람들이 타고 있는 비행기를 보고 최초 비행기일 수 있다고 생각함 2)주제를 잡기 어려웠는데 비행기와 관련된 낡은 흑백이미지들을 보고 옛날 기구를 다시 제작하는 것이 아닌지 생각함

3. 키프레임 추출 알고리즘 설계

3.1 개념적 모형

개념적 모형은 영상에 초점을 맞춰서 구축하기로 하였다. 따라서 선행 연구 중에서 Yang과 Marchionini(2004)가 제시한 적합성 기준인 텍스츄얼 기준, 비주얼 기준 및 함축적인 기준의 세 가지 카테고리는 그대로 사용한다. 다만 텍스츄얼 기준에 속할 세부 항목은 키프레임내의 텍스트 정보로 한정하기 위해서 자막과 이미지속의 글자를 사용하기로 한다.

다음 단계로 Shatford(1986) 연구와 같은 이미지 인식 및 분류 체계에 대한 선행 연구들

www.kci.go.kr

의 내용을 정리한 이미지 분류 카테고리(표 2 참조)와 이용자의 키프레임 인식 패턴에 기초하여 〈그림 1〉과 같이 이 세 가지 기준을 다시 세분하였다. 이 세 가지 기준의 중요도순은 텍스츄얼 기준 → 비주얼 기준 → 함축적인 기준 순으로 정하였다.

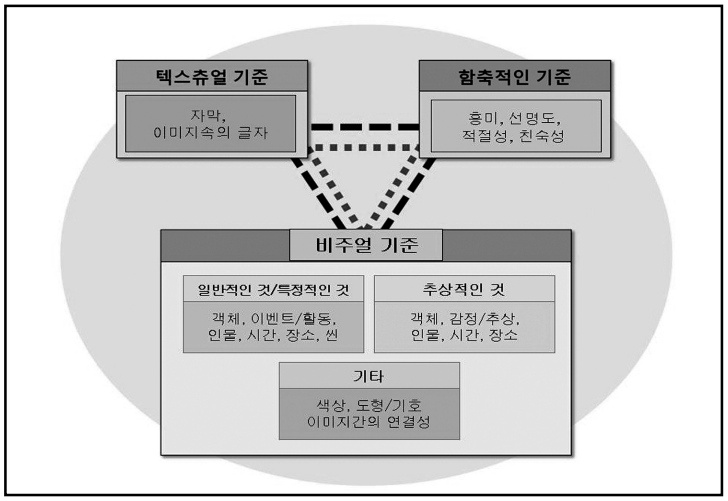
텍스츄얼 기준은 자막 또는 이미지속의 글자를 포함한다. 비주얼 기준은 일반적인 것, 특정적인 것, 추상적인 것 및 기타로 구분하고 이를 다시 다음과 같이 세분하였다. 즉, 일반적인 것과 특정적인 것의 객체, 이벤트/활동, 장소, 시간 및 씬으로 구분하였다. 예를 들어서, 일반적은 것은 어떤 객체를 표현할 때 단순히 "건물"로 표현한 반면 특정한 것은 "한국 대학교"와 같이 구체적인 건물로 표현한 것을 의미한다. 추상적인 것은 상징적으로 묘사되는 객체, 감정/추상, 추상적인 장소 및 시간으로 세분화시켰다. 기타에는 색상, 이미지간의 연결성 등을 포함시켰다. 마지막으로 함축적인 기준은 프레임에 대한 적절성, 흥미 등을 포함한다.

3.2 키프레임 추출 과정

키프레임 추출 과정은 다음과 같이 크게 두 단계로 구성한다. 제 1 단계는 이미지 프로세싱을 이용하여 키프레임 후보 그룹들을 추출한다. 제 2 단계는 제 1 단계에서 추출된 후보 키프레임 그룹 중에서 앞의 개념적 모형에 기초하여 구성한 기준에 따라서 수작업으로 키프레임들을 최종적으로 선택한다. 다음은 제 1 단계와 제 2 단계에 대해서 상술한다.

3.2.1 제 1 단계

이미지 프로세싱을 이용하여 후보 키프레임들을 추출한다. 이는 일반적으로 비디오 클립을 시공간적으로 연속적인 프레임들의 집합인 샷 단위로 분할하고 분할된 샷에서 후보 키프레임들을 추출하여 순서대로 연결하는 방식이다.



〈그림 1〉 키프레임 추출을 위한 개념적 모형

www.kci.go.kr

키프레임 추출에 많이 사용하고 있는 알고리즘에는 다음과 같은 것들이 있다.

Nagasaka와 Tanka(1992)는 분할된 샷의 경계에 있는 처음 프레임을 키프레임으로 간주하는 샷의 경계를 이용하는 방식을 제안하였다. Dufaux(2000)는 모션(motion) 분석에 기초하여 키프레임을 추출하는 방식을 제시하였다. 신성윤과 표성배(2006)는 RGB 컬러 히스토그램과 카이스퀘어 히스토그램을 합성한 장면전환 검출기법을 사용하여 키프레임을 추출하였다. 이러한 알고리즘들을 구현한 실험용 프로그램이나 상용 소프트웨어를 이용하여 후보 키프레임들을 추출한다.

3.2.2 제 2 단계

키프레임 추출을 위한 개념적 모형에 기초하여 구성한 기준들에 의해서 제 1 단계에서 추출된 후보 키프레임들 중에서 최종적으로 키프레임들을 선택한다. 단 개념적 모형에 포함된 감정/추상은 다른 요소들과 달리 또 다른 분석틀을 요구하므로 현 기준에는 포함시키지 않았다. 키프레임 선택 기준은 다음과 같다.

1) 중복되는 의미를 갖는 프레임을 제거한다. 다시 말해서 화면 내용이 비슷하거나 중복된 것 또는 같은 해설자에 대한 프레임들이 복수개인 경우가 여기에 해당된다. 중복된 프레임들을 선택할 때 함축적인 기준을 적용하여 선명도가 높고 이해하기 용이한 프레임을 선정한다.

2) 중복되지 않은 프레임 그룹에서 다음의 조건에 맞는 키프레임들을 선정한다.

(1) 텍스츄얼 기준을 제일 먼저 적용한다.

(a) 해당 장면의 내용이나 편집자의 의도를 잘 표현하고 있는 인위적으로 삽입된 자막을 포함하고 있는 키프레임을 추출한다.

(b) 주제 또는 의도를 나타내는 이미지속의 글자가 포함되어 있는 프레임을 선택한다.

(2) 비주얼 기준은 특정한 것, 일반적인 것, 기타, 추상적인 것 순으로 적용 순서를 정한다.

(a) 주제를 나타내는 특정한 객체, 이벤트/활동, 인물, 시간/장소 및 씬이 있는 프레임을 선택한다.

(b) 주제를 나타내는 일반적인 객체, 이벤트/활동, 인물, 시간/장소 및 씬이 있는 프레임을 선택한다.

(c) 이미 선택된 프레임과 연관시켜 주제를 잘 표현할 수 있는 프레임을 추가로 선택하거나 앞 단계에서 선택되지 않은 두 개 이상의 프레임들이 결합되어 비디오 의미를 명료하게 표현하는 경우 이들 프레임들을 선택한다.

(d)주제를 나타내는 기호/도면이 있는 프레임을 선택한다.

(e)천연색 이미지에서 시간을 나타내기 위해서 흑백으로 사용된 프레임을 선택한다.

(f)주제를 나타내는 추상적인 객체, 인물 및 시간/장소가 있는 프레임을 선택한다.

www.kci.go.kr

4. 제안된 키프레임 추출 알고리즘 평가

4.1 연구 가설

주제 파악을 용이하게 하는 핵심 프레임들을 선정하는 기준을 적용시켜 구성한 영상 초록(하이브리드 방식)과 일정한 시간 간격에 따라서 무작위로 키프레임을 추출하여 구성한 영상 초록(랜덤 방식)이 비디오의 요약 및 색인어 추출의 정확도에서 어떤 차이를 보이는지 알아보고자 한다. 더 나아가 이러한 두 가지 추출 방식이 순차 및 구조화 초록의 비디오의 요약 및 색인어 추출의 정확도에 어떤 영향을 미치는지 알아보고자 한다.[1] 이러한 연구 문제를 기초로 하여 다음과 같은 네 개의 연구 가설을 설정한다.

1) 비디오의 요약문 정확도는 영상 초록의 키프레임을 추출하는 방식(하이브리드 및 랜덤 방식)에 따라서 차이가 있을 것이다.

2) 비디오의 색인어 정확도는 영상 초록의 키프레임을 추출하는 방식(하이브리드 및 랜덤 방식)에 따라서 차이가 있을 것이다.

3) 비디오의 요약문 정확도에서 키프레임 추출 방식(하이브리드 및 랜덤 방식)과 영상 초록 유형(순차 및 구조화 초록)간에 상호작용 효과가 나타날 것이다.

4) 비디오의 색인어 정확도에서 키프레임 추출 방식(하이브리드 및 랜덤 방식)과 영상 초록 유형(순차 및 구조화 초록)간에 상호작용 효과가 나타날 것이다.

4.2 실험 설계

4.2.1 피조사자와 설문지

피조사자 집단은 동일한 집단을 구성하기 위해서 J대학의 인문대학의 학부생을 42명 선정하고 선정한 다음 학과, 학년, 남녀 등을 고려하여 되도록 고루게 분포된 2개의 실험 집단을 구성하였다. 21명으로 구성된 실험 집단 1은 랜덤 방식으로 구성한 12개의 표본 비디오에 대한 영상 초록을 보게 하였다. 21명으로 구성된 실험 집단 2는 본 연구에서 설계한 하이브리드 방식에 의해서 구성한 영상 초록을 보게 하여 그 차이를 비교하고자 한다.

4.2.2 표본 비디오 선정

표본 비디오로는 이용자의 키프레임 인식 패턴 분석에서 사용한 것과 동일한 비디오를 활용하였다(표 5 참조). 비디오 길이에 따라서 영상 초록의 기능이 달라질 수 있기 때문에 2분 17초에서 20분 사이에 있는 비디오를 선정하였다.

4.2.3 실험 시스템 설계

본 연구에서 제안한 영상초록 구성을 위한 키프레임 추출 알고리즘의 효율성을 평가하기 위해서 12개의 교육 비디오를 대상으로 하여

1) 순차 영상 초록(Sequential Storyboard, SEQ)은 비디오의 제작자나 편집자가 의도한 순서를 그대로 유지하여 키프레임을 배열하는 초록인데 반하여 구조화 영상 초록(Structured Storyboard, STR)은 순차 영상 초록에서 사용한 키프레임들을 그대로 사용하지만 "객체/이벤트/활동", "인물" 및 "배경"의 카테고리로 재구성하여 배열하는 초록을 지칭한다.

www.kci.go.kr

〈표 5〉 표본 비디오 리스트(2)

비디오 번호	제 목	키프레임수	초
101	Food Preservation	15개	267(04:27)
102	Exercise and Nutrition	15개	137(02:17)
103	Benefits of Sleep	15개	295(04:55)
104	Earthquake	15개	156(02:36)
109	First Flying Machines	12개	258(04:18)
113	Computer Rage	15개	171(02:51)
114	Ubiquitous Computing in the Living Room	22개	242(04:02)
115	Bicycle Today - Automobile Tomorrow	15개	600(10:00)
116	Flexible Wing	20개	200(04:40)
117	NASA Connect - PSA - Robots	15개	145(02:25)
118	Civil War, The	25개	910(15:06)
119	GPS for Pilots	15개	139(02:19)

두 종류의 영상 초록의 키프레임 추출 방식을 적용하였다. 첫 번째 방식은 두 단계에 걸쳐서 즉, 첫 단계는 이미지 프로세싱 방법에 의해서 자동으로 키프레임 후보 집단을 추출한다. 두 번째 단계는 추출된 키프레임 집단에서, 본 연구에서 제안한 키프레임 선정 기준에 따라 사람들이 최종적으로 키프레임들을 선정하는 하이브리드 방식이다.

두 번째 방식은 KMPlayer를 사용하여 일정한 시간 간격에 따라서 무작위로 키프레임 추출하고 이를 원래 순서대로 배열하여 영상초록을 만드는 랜덤 방식을 채택하였다. 본 연구는 이미지의 의미 전달의 가능성을 조사하는데 초점을 맞추었기 때문에 두 가지 방식에 의해서 선정된 키프레임 중 너무 명확하게 비디오의 표제 또는 주제가 들어나는 키프레임은 영상초록에서 제외시켰다.

각 비디오에 대하여 순차 영상 초록과 구조화 영상 초록을 구성하였고 조건을 동일하게 하기 위해서 같은 비디오에 대해서는 키프레임 추출 방식이나 영상 초록의 유형에 상관없이 같은 수의 동일한 키프레임을 사용하였다. 키프레임수는 최대 25개까지 가능하게 구성하였다(표 5 참조). 〈그림 2〉는 두 개의 실험 시스템의 초기 화면들을 결합해 놓은 것이다. 〈그림 3〉은 음식 보존에 관한 비디오의 하이브리드 및 랜덤 방식에 의해 구성한 영상 초록들이다. 랜덤 방식에 의한 영상 초록에는 중복된 이미지들이 포함되어 있다. 한편 하이브리드 방식에 의한 영상 초록에는 중복된 이미지는 거의 없고 비디오 주제를 나타내는 캔, 유리병과 같은 객체가 표현된 이미지들이 랜덤 방식 보다는 상대적으로 더 많이 포함되어 있는 것을 확인할 수 있다.

실험 집단 1에게는 하이브리드 방식에 의해서 구성된 영상 초록들을 보도록 하였고, 각 비디오에 대해서 두 가지 유형의 영상 초록 [(유형1)순차→(유형2)구조화 영상초록] 또는 [(구조화(유형2)→순차 영상초록(유형2)] 들을 연이어 보도록 하여 영상 초록의 키프레임

www.kci.go.kr

비디오코드	첫번째 브라우징	두번째 브라우징
N0.104	유형 1	유형 2
N0.101	유형 2	유형 1
N0.113	유형 1	유형 2
N0.102	유형 2	유형 1
N0.115	유형 1	유형 2
N0.103	유형 2	유형 1
N0.116	유형 1	유형 2
N0.109	유형 2	유형 1
N0.118	유형 1	유형 2
N0.114	유형 2	유형 1
N0.119	유형 1	유형 2
N0.117	유형 2	유형 1

〈그림 2〉 하이브리드/랜덤 방식으로 구성된 영상 초록 리스트

〈그림 3〉 음식 보존에 관한 비디오의 하이브리드 및 랜덤 방식에 의한 영상 초록들

추출 방식이 이 두 유형의 초록(순차 및 구조화 초록)에서 어떤 차이가 있는지 살펴보았다. 실험 집단 2에게는 랜덤 방식에 의해서 구성된 영상 초록들을 보도록 하였고, 하이브리드 방식에서처럼 각 비디오에 대해서 두 가지 유형의 영상 초록을 연이어 보도록 하였다.

www.kci.go.kr

4.2.4 실험 절차

실험 절차는 다음과 같이 세 단계로 구분하였다. 첫째, 실험 내용에 대한 소개를 10분 정도 한다. 둘째, 12개 영상 초록의 각 초록을 보고 비디오 내용에 대한 요약문을 세 문장 이상 기록하도록 하고 색인어는 4개를 기술하도록 한다. 한 비디오에 대해서 2장의 설문지, 총 12개 비디오에 대해서 24장의 설문지가 배포된다.

첫 번째 영상 초록에 대한 기술이 끝나면 설문지를 바로 수집한다. 이는 한 비디오에 대해서 순차 및 구조화 영상 초록을 각각 보여주기 때문에 바로 이전의 설문지 기술 내용에 영향을 받을 수 있기 때문이다. 그런 다음 나머지 비디오에 대해서 같은 방식으로 수행한다. 맨 처음 비디오에 대해서는 시간을 여유있게 주어서 피조사자들이 실험 내용을 이해할 수 있도록 한다. 셋째, 두 번째 비디오부터는 각 영상 초록(순차 또는 구조화)에 대해서 8분 정도 시간을 준다. 따라서 한 비디오에 대해서 대략 16분이 소용되고 12개 비디오에 대해서는 3시간 20분이 소요된다.

4.3 실험 결과

4.3.1 데이터 분석

중다변량분석(MANOVA)은 변량분석(ANOVA)의 일반화된 형태로 독립변인의 주효과와 상호작용 효과를 검증하는 일차적인 목적을 가지고 있으며, 여러 종속변인의 선형조합에 대한 독립변인의 효과 검증이 가능하다. 본 연구에서는 독립 변인으로 키프레임 추출 알고리즘 방식(하이브리드 및 랜덤 방식)과 영상 초록의 유형[순차 초록(SEQ) 및 구조화 초록(STR)] 그리고 이 독립변인들의 상호작용으로 구성된다. 종속변인은 요약문 정확률과 색인어 정확률이 된다. 분석 결과, 요약문 정확률에서 키프레임 추출 알고리즘 방식과 두 독립변인의 상호작용(키프레임 추출 알고리즘 방식 * 영상 초록 유형)이 종속 변인에 영향을 미치는 것으로 나타났다. 색인어 정확률에서는 종속변인에 영향을 미치는 변인 또는 상호작용 효과가 없었다. 분석 결과는 가설 검증에서 상술된다.

4.3.2 가설 검증

1)연구 가설 1과 2의 검증: “비디오의 요약문 및 색인어 정확도는 영상 초록의 키프레임을 추출하는 방식(하이브리드 및 랜덤 방식)에 따라서 차이가 있을 것이다.” 앞의 중다변량 및 t-검증 결과, 하이브리드 방식에 의한 영상 초록의 요약문 정확률 평균(0.61)이 랜덤 방식의 평균(0.55) 보다 높고 그 차이가 통계적으로 유의미한 것으로 나타나 가설 1은 검증되었다(표 6과 표 7 참조). 그러나 〈표 8〉에 의하면 하이브리드 방식에 의한 영상 초록의 색인어 정확률 평균(0.52)이 랜덤 방식의 평균(0.51)과 전혀 차이를 보이지 못해서 가설 2는 검증되지 못하였다.

2)연구 가설 3과 4의 검증: “비디오의 요약문 정확도에서 키프레임 추출 방식(하이브리드 및 랜덤 방식)과 영상 초록 유형(순차 및 구조화 초록)간에 상호작용 효과가 나타날 것이다.” 〈표 6〉의 분석 결과에 의하면 “키프레임 추출 방식”과 “영상 초록 유형”간에 상호작용 효과가 있는 것으로 나타났다. 이를 좀 더 자세히 살펴보면 하이브리드 방식과 랜덤 방식간의 요약문 정확률 평균 차이는 순차 영상초록(하이브리드=0.63, 랜덤=0.50, p=0.00)에서만 통계적으로 유의미하게 나타났지 구조화 영상초

www.kci.go.kr

〈표 6〉 중다변량 분석 결과

독립 변인 \ 종속변인	요약문 정확률 f(p)	색인어 정확률 f(p)
키프레임 추출 알고리즘 방식(하이브리드 vs. 랜덤)	4.64(0.03)*	0.14(0.71)
영상 초록 유형(SEQ vs. STR)	0.94(0.33)	0.00(0.99)
키프레임 추출 알고리즘 방식 * 영상 초록 유형	7.01(0.01)**	0.06(0.81)

**는 0.01 수준에서 유의미하고 *는 0.05 수준에서 유의미함

〈표 7〉 요약문 정확률 분석 결과

추출 방식 \ 초록 유형	요약문 정확률 평균(표준편차)		
	순차 영상초록(SEQ)	구조 영상초록(STR)	SEQ+STR
하이브리드 방식	0.63(0.27)	0.58(0.25)	0.61(0.26)
랜덤 방식	0.50(0.29)	0.60(0.33)	0.55(0.31)
하이브리드+랜덤	0.56(0.28)	0.59(0.29)	-

t-검증 결과

대응	유의확률
하이브리드(SEQ+STR)(0.61) vs. 랜덤(SEQ+STR)(0.55)	0.03*
하이브리드(SEQ)(0.63) vs. 랜덤(SEQ)(0.50)	0.00**
하이브리드(STR)(0.58) vs. 랜덤(STR)(0.60)	0.74

〈표 8〉 색인어 정확률 분석 결과

추출 방식 \ 초록 유형	색인어 정확률 평균(표준편차)		
	순차 영상초록(SEQ)	구조화 영상초록(STR)	SEQ+STR
하이브리드 방식	0.52(0.25)	0.52(0.24)	0.52(0.25)
랜덤 방식	0.51(0.26)	0.52(0.30)	0.51(0.28)
하이브리드+랜덤	0.51(0.26)	0.52(0.27)	-

록(하이브리드=0.58, 랜덤=0.60, p=0.74)에서는 나타나지 않았다(표 7 참조). 따라서 비디오의 의미 파악에서 키프레임 추출 방식과 영상 초록 유형간에 상호작용 효과가 나타날 것이라는 가설 3은 검증되었다. 또한 비디오의 색인어 정확도에서 키프레임 추출 방식과 영상 초록 유형간에 상호작용 효과가 나타날 것이라는 가설 4는 검증되지 못했다.

4.4 논 의

실험 결과에 의하면 하이브리드 방식으로 구성된 영상 초록의 요약문 정확률 평균(0.61)이 랜덤 방식의 평균(0.55) 보다 높고 그 차이가 통계적으로 유의미한 것으로 나타났으나 그 차이가 예상한 것 보다는 크지 않았다. 순차 영상 초록에서 두 가지 키프레임 추출 방식의 평균 정확도(0.63, 0.50)의 차이가 나타났지만, 구조

www.kci.go.kr

화 영상 초록에서 이 두 방식의 평균 정확도(0.58, 0.60)의 차이가 거의 나타나지 않았기 때문이다(표 7 참조). 이런 결과가 나온 것은 구조화 영상 초록의 경우 랜덤으로 추출된 키프레임들이라도 키프레임간의 관계를 설정해 줌으로써 비디오의 맥락 파악을 용이하게 해 주었기 때문으로 생각된다.

이러한 연구 결과는 현재 주로 순차 영상 초록을 사용하고 있는 비디오 검색 환경에 활용될 수 있을 것으로 생각된다. 첫째, 순차 초록에서 하이브리드 방식에 의한 영상 초록의 질이 랜덤 방식에 의한 영상 초록의 질 보다 훨씬 높게 나타났다. 따라서 비디오의 정확한 의미 파악을 위해서 비디오 색인자의 의미 부여가 첨가된 하이브리드 방식이 대학, 학교 및 방송사 디지털 도서관의 비디오 자료 표현에 활용되는 것이 바람직 할 것으로 보인다.

장기적으로는 이미지 프로세싱과 패턴인식에 대한 앞으로의 연구 성과를 활용하여 본 연구에서 제안한 키프레임 추출 알고리즘(제 2단계에서 적용한 기준)에 따라 키프레임을 자동으로 추출하는 방안에 대한 지속적인 연구가 필요하다고 생각한다. 현재 비디오의 내용 검색을 위해서 해당 장면의 내용이나 편집자의 의도를 잘 표현하고 있는 인위적으로 삽입된 자막을 자동으로 추출하는 알고리즘을 제안하는 연구(김원익, 김창익 2008)와 같이 이미지 프로세싱에 대한 연구가 수행되고 있지만 당장 실용화하기에는 아직도 가야할 길이 멀다.

둘째, 키프레임들을 무작위로 추출하여 구성한 영상 초록을 많이 활용하고 있는 인터넷 검색 환경에는 영상 초록의 효율성을 높이기 위해서 자동으로 추출된 키프레임들을 문맥적 관계에 기반하여 재구성한 구조화 영상 초록을 구현하는 방안을 적용해 볼 수 있겠다. 물론 이 부분도 장기적으로는 키프레임들을 객체/이벤트/활동, 인물 및 배경의 세 가지 카테고리로 자동으로 분류하는 방안이 모색되어져야 할 것이다.

〈표 8〉에 의하면 하이브리드 방식에 의한 영상 초록의 색인어 정확률 평균(0.52)이 랜덤 방식의 평균(0.51)과 전혀 차이를 보이지 못해서 가설 2는 검증되지 못하였다. 이러한 결과가 나온 것은 피조사자들이 색인어를 추출할 때 키프레임간의 관계를 파악하기 보다는 개별적인 키프레임을 보는 경향 때문이 아닌가 생각된다. 본 실험에서 4개의 색인어만 선정하도록 하고 있는데 이는 랜덤 방식으로 구성된 영상 초록에 하이브리드 방식으로 구성된 영상 초록에 있는 최소 4개의 핵심 키프레임들이 포함될 가능성이 높다고 예측해 볼 수 있다. 다시 말해서 색인어 수를 5개 이상으로 늘인다면 품질이 높은 키프레임을 상대적으로 많이 포함하고 있는 하이브리드 방식에 의한 영상 초록의 색인어 추출 정확도가 랜덤 방식의 정확도 보다 더 높아질 가능성이 있다고 생각한다.

5. 결 론

영상 초록의 품질을 결정하는 가장 중요한 요인 중 하나는 비디오의 의미 전달에 필요한 키프레임을 선정하는 작업이다. 본 연구는 키프레임 선정을 위한 이론 체계를 수립하기 위해서 멀티미디어 자료에 대한 적합성 기준, 이미지 인식 및 분류 체계 그리고 이용자의 키프

www.kci.go.kr

레임 인식 패턴을 조사하여 분석해 보았다. 그런 다음 이러한 이론 체계를 기초로 하여 하이브리드 방식으로 비디오에서 키프레임을 추출하는 알고리즘을 설계해 보았다.

구현된 알고리즘은 개념적 모형을 기반으로 하여 키프레임 추출 과정을 두 단계 즉, 제 1 단계는 이미지 프로세싱을 이용하여 키프레임 후보 그룹들을 추출하고, 제 2 단계는 제 1 단계에서 추출된 후보 키프레임 그룹 중에서 정해진 기준에 따라서 수작업으로 키프레임들을 최종적으로 추출하였다. 제안된 알고리즘의 성능을 평가하기 위해서 그 알고리즘을 활용하여 구성한 영상 초록(하이브리드 방식)을 일정한 시간 간격에 따라서 무작위로 키프레임을 추출하여 구성한 영상 초록(랜덤 방식)과 비교해 보았다.

실험 결과에 의하면 하이브리드 방식으로 구성된 영상 초록의 요약문 정확률 평균(0.61)이 랜덤 방식의 평균(0.55) 보다 높고 그 차이가 통계적으로 유의미한 것으로 나타났으나 그 차이는 크지 않았다. 순차 영상 초록에서 두 가지 키프레임 추출 방식의 평균 정확도(0.63, 0.50)의 차이가 나타났지만 구조화 영상 초록에서 이 두 방식의 평균 정확도(0.58, 0.60)의 차이가 거의 나타나지 않았기 때문이다. 이는 구조화 영상 초록의 경우 랜덤으로 추출된 키프레임들이라도 키프레임간의 문맥적 관계를 설정해 줌으로써 비디오의 맥락 파악을 용이하게 해 주었기 때문으로 생각된다.

이러한 연구 결과를 디지털 도서관 및 인터넷 환경에서 비디오 자료의 검색 효율성을 높이기 위해서 응용해 볼 수 있을 것이다. 첫째, 순차 초록에서 하이브리드 방식에 의한 영상 초록의 질이 랜덤 방식에 의한 영상 초록의 질보다 훨씬 높기 때문에 하이브리드 방식이 대학, 학교 및 방송사 디지털 도서관의 비디오 자료 표현에 활용되는 것이 바람직 할 것으로 보인다. 둘째, 이미 키프레임들을 무작위로 추출하여 구성한 영상 초록을 많이 활용하고 있는 인터넷 검색 환경에는 영상 초록의 효율성을 높이기 위해서 자동으로 추출된 키프레임들을 문맥적 관계에 기반하여 재구성한 구조화 영상 초록을 구현하는 방안을 적용해 볼 수 있을 것이다.

참 고 문 헌

김원익, 김창익. 2008. 새로운 비디오 자막 영역 검출 기법. 『방송공학회논문지』, 13(4): 544-553.

김종성, 이순탁, 백종환. 2005. 내용기반 비디오 요약을 위한 효율적인 얼굴 객체 검출. 한국통신학회논문지』, 제30권 7C호: 675-686.

김현희, 김용호, 고수현. 2007. 비디오 자료의 의미 추출을 위한 영상 초록의 효용성에 관한 실험적 연구. 『정보관리학회지』, 24(4): 53-72.

신성윤, 표성배. 2006. 텔레매틱스에서 효율적인 장면전환 검출기법을 이용한 비디오 브

www.kci.go.kr

라우징. 『한국컴퓨터정보처리논문지』, 11(4): 147-154.

이준용, 문영식. 2003. 샷 기여도와 왜곡률을 고려한 키 프레임 추출 알고리즘. 『전자공학회논문지』, 제40권 CI편 제3호: 11-17.

Browne, P. and A. F. Smeaton. 2005. "Video Retrieval Using Dialogue, Keyframe Similarity and Video Objects." *ICIP 2005 - International Conference on Image Processing* Genova, Italy: 11-14.

Choi, Y. and E. M. Rasmussen. 2002. "User's Relevance Criteria in Image Retrieval in American History." *Information Processing and Management*, 38(5): 695-726.

Chung, E. K. and J. W. Yoon. 2008. "A Categorical Comparison between User-supplied Tags and Web Search Queries for Images." *In Proceedings of the ASIST Annual Meeting* Silver Spring, MD: American Society for Information Science and Technology.

Ding, W. et al. 1999. "Multimodal Surrogates for Video Browsing." *In Proceedings of the fourth ACM Conference on Digital Libraries* (August, Berkeley CA, USA), ACM: 85-93.

Dufaux, F. 2000. "Key Frame Selection to Represent a Video." *In IEEE Proceedings of International Conference on Image Processing* vol.2: 275-278.

Greisdorf, H. and B. O'Connor. 2002. "Modelling What Users See When They Look at Images: a Cognitive Viewpoint." *Journal of Documentation*, 58(1): 6-29.

Hughes, A., et. al. 2003. "Text or Pictures? an Eye-tracking Study of How People View Digital Video Surrogates." *In Proceedings of CIVR 2003* 271-280.

Kristin, B. et al., 2006. *Audio Surrogation for Digital Video: a Design Framework* UNC School of Information and Library Science(SILS) Technical Report TR 2006-21.

Laine-Hermandez, M. and S. Westman. 2008. "Multifaceted Image Similarity Criteria as Revealed by Sorting Tasks." *In Proceedings of the ASIST Annual Meeting* Silver Spring, MD: American Society for Information Science and Technology.

Lyer, H. and C. D. Lewis. 2007. "Prioritization Strategies for Video Storyboard Keyframes." *Journal of American Society for Information Science and Technology* 58(5): 629-644.

Marchionini, G. and G. Geisler. 2002. "The Open Video Digital Library." *D-Lib Magazine*. 8(12). [cited 2008.9.25].

Markkula, M. and E. Sormunen. 1998. "Searching for Photos - Journalistic Practices in Pictorial IR." In: Eakins, J.P., et al. (Eds), *The Challenge of Image Retrieval*. Newcastle upon Tyne, 5-6 Feb 1998. British Computer Society(BCS), Electronic Workshops in Computing.

Mu, X. and G. Marchionini. 2003. "Enriched

www.kci.go.kr

Video Semantic Metadata: Authorization, Integration, and Presentation." *In Proceedings of the ASIST Annual Meeting*: 316-322. Silver Spring, MD: American Society for Information Science and Technology.

Nagasaka, A. and Y. Tanka. 1992. "Automatic Video Indexing. and Full-Video Search for Object Appearances," *Visual Database Systems*, Vol.2: 113-127.

Panofsky, E. 1955. *Meaning in the Visual Arts: Meaning in and on Art History* Doubleday.

Shatford, S. 1986. "Analyzing the Subject of a Picture: a Theoretical Approach." *Cataloging & Classification Quarterly*, 6(3): 39-62.

Yang, M. 2005. *An Exploration of Users' Video Relevance Criteria*. Unpublished Doctoral Dissertation, University of North Carolina at Chapel Hill.

Yang, M. and G. Marchionini. 2004. "Exploring Users' Video Relevance Criteria - A Pilot Study." *In Proceedings of the ASIST Annual Meeting*. 229-238. Silver Spring, MD: American Society for Information Science and Technology.

www.kci.go.kr