韓國IT서비스學會誌
第12卷 第1號
2013年 3月, pp.271-287
http://dx.doi.org/10.9716/KITS.2013.12.1.271

워키태깅 : 스마트폰 환경에서 음성기반의 효과적인 영상 콘텐츠 어노테이션 방법에 관한 연구*

박준영** · 이수빈*** · 강동엽** · 석영태****

WalkieTagging : Efficient Speech-Based Video Annotation Method for Smart Devices*

Joon Young Park** · Soobin Lee*** · Dongyeop Kang** · YoungTae Seok****

■ Abstract ■

The rapid growth and dissemination of touch-based mobile devices such as smart phones and tablet PCs, gives numerous benefits to people using a variety of multimedia contents. Due to its portability, it enables users to watch a soccer game, search video from YouTube, and sometimes tag on contents on the road. However, the limited screen size of mobile devices and touch-based character input methods based on this, are still major problems of searching and tagging multimedia contents. In this paper, we propose WalkieTagging, which provides a much more intuitive way than that of previous one. Just like any other previous video tagging services, WalkieTagging, as a voice-based annotation service, supports inserting detailed annotation data including start time, duration, tags, with little effort of users. To evaluate our methods, we developed the Android-based WalkieTagging application and performed user study via a two-week. Through our experiments by a total of 46 people, we observed that experiment participator think our system is more convenient and useful than that of touch-based one. Consequently, we found out that voice-based annotation methods can provide users with much convenience and satisfaction than that of touch-based methods in the mobile environments.

Keyword : Media Annotation Service, Mobile Multimedia Contents, Usability Testing

논문투고일 : 2013년 01월 26일 논문수정완료일 : 2013년 03월 12일 논문게재확정일 : 2013년 03월 17일

* 본 연구는 지식경제부 산업원천기술개발사업[10035166, 창의적 인재육성을 위한 지능형 튜터링 시스템 기술 개발 연구]의 일환으로 수행하였음.

** 한국과학기술원 IT융합연구소 지식융합팀

*** 한국과학기술원 IT융합연구소 지식융합팀, 교신저자

**** SK텔레콤 솔루션 사업본부 솔루션 개발팀

1. 서 론

스마트폰 및 태블릿 PC와 같은 모바일 기기의 빠른 보급과 성장은 동영상 콘텐츠 사용자들에게 수많은 혜택을 제공하고 있다. 사용자들은 언제, 어디서든 자신의 모바일 기기를 통해 축구경기를 생중계로 시청할 수 있으며, 달리는 버스 안에서 YouTube과 같은 영상 콘텐츠 제공업자의 서비스를 이용하여 동영상을 검색하거나 시청할 수 있게 되었다[14]. 휴대성, 이동성, 그리고 편재성 등과 같은 모바일 기기의 특성이 동영상 콘텐츠 서비스에 결합됨에 따라 이를 이용하는 사용자들에게 전례 없는 편의를 제공한 결과이다. 진화하는 모바일 기기 및 서비스 환경 속에서 동영상 콘텐츠 사용자의 편리성은 증대되고 있으며, 모바일 시장의 콘텐츠 소비량도 증가하고 있다. 또한 모바일 동영상 콘텐츠 소비 증대는 양적인 차원의 시청 증대를 넘어서 개인화된 콘텐츠 검색 및 태깅 연구에 대한 욕구를 함께 증대시키고 있다[2, 5, 8].

그러나 모바일 기기의 제한적인 스크린화면 및 입력방식은 여전히 사용자들의 활발한 콘텐츠 소비활동에 장애요소가 되고 있고, 콘텐츠에 주석이나 태깅을 함에 있어서도 불편함을 주고 있다. 수많은 동영상 콘텐츠 공급업체에서 비디오 어노테이션을 지원하는 툴 및 서비스를 제공하고 있지만 대부분 PC 환경의 입력모드(마우스 또는 키보드)를 기반으로 인터페이스가 설계되어 모바일 기기 상에서 활용하기에는 다소 한계가 있는 상황이다. 모바일 기기의 화면이나 입력방식이 전체 비디오 영상 어노테이션 서비스를 지원할 만큼 충분하지 않기 때문이다. 현존하는 비디오 어노테이션 솔루션들의 대부분은 영상에 추가적인 정보를 덧붙이기 위해 터치화면 상에서 텍스트 입력을 통해 어노테이션을 수행하는 전통적인 방식을 따르고 있다[6]. 이에 따라 사용자들은 그들이 비디오 영상을 시청하고 있는 동안, 어떠한 주석이나 태깅도 영상에 덧붙일 수 없는 상황이다. 모바일 상에서 기존의 비디오 어노테이션 시스템을 이용하는 사용자가 비디오에 어노테이션을 수행하기 위해서는 다음의 4가지 작업이 요구된다. 첫째, 사용자는 어노테이션 정보가 들어갈 시작 시간을 정해야 한다. 둘째, 마찬가지로 어노테이션 정보가 들어갈 끝 시간을 정해야 한다. 셋째, 시작 시간과 끝 시간을 정한 후, 사용자는 터치를 이용하여 어노테이션 할 텍스트를 입력해야한다. 넷째, 사용자는 전송버튼과 같은 인터페이스를 통해 어노테이션을 수행한 데이터를 해당서버에 전송해야 한다. 이러한 과정들은 제한적 화면크기의 모바일 기기 상에서 수행하기에는 다소 복잡하고 어려운 작업들로써 모바일 기기를 이용하여 비디오를 감상하는 사용자에게 비디오 어노테이션 작업을 포기하게 만드는 원인이 된다. 모바일 기기가 동영상 시청에 있어 중요 수단이 됨에 따라, 제한적인 화면상에서 기존의 입력방식으로 영상에 어노테이션 하는 작업이 도전을 맞이하고 있는 것이다.

본 연구는 이러한 상황에서 효과적인 비디오 어노테이션 작업을 지원하기 위해 워키태깅(Walkie Tagging) 서비스를 제안하고자 한다. 워키태깅 서비스는 기존의 워키토키(무전기)의 단순한 인터페이스에서 착안 한 것으로 기존의 터치기반의 입력방식보다 훨씬 더 직감적인 원터치 및 음성입력 기반의 어노테이션 방식을 사용한다. 워키태깅은 음성기반의 입력방식을 활용하여 기존의 video annotation 방식이 제공하는 것과 마찬가지로 태깅 시작 시간, 기간, 태그단어 등의 메타데이터 입력을 지원한다.

본 연구에서 제안하는 워키태깅은 기존의 사용자 기반의(사용자에 의해 수행된) 영상 어노테이션 관련 서비스에 대한 분석에서부터 시작되었으며, 유용성 검증을 위한 사용자 실험을 진행하였다. 구체적인 설명을 위해 제 2장에서 기존의 영상 어노테이션 서비스와 관련된 연구를 소개하고 있으며, 제 3장에서는 시스템 설계 면에서 워키태깅 서비스 구현을 위한 시스템 아키텍처 및 어노테이션 방식에 대해 설명한다. 제 4장에서는 사용자 실험을 위한 설계 및 방법론에 대해 서술하며, 제 5장

에서 사용자 실험에 대한 결과를 분석하였다. 마지막으로 6장에서는 결론 및 향후 계획을 통해 워키태깅 서비스의 의의 및 추후계획에 대해 정리하였다.

2. 관련 연구

2.1 VideoAnnEx

VideoAnnEx는 IBM에서 배포한 영상 어노테이션 툴이다[20]. VideoAnnEx는 모바일 기기 및 중앙 서버간의 데이터 통신 및 MPEG-7 메타데이터 정보를 동원하여 비디오 시퀀스에 대한 어노테이션 작업을 돕는다. 구체적으로 VideoAnnEx는 사용자의 비디오 어노테이션 작업을 지원하기 위해 모바일 인터페이스 상에서 MPEG 비디오 시퀀스와 샷 세그멘테이션 파일을 입력으로 받는다. 여기서 샷 세그멘테이션 파일이란 비디오 시퀀스를 비디오 샷 단위로 분할해놓은 파일을 말하며, 시퀀스 내에서 디졸브나 페이드와 같은 효과가 발생하는 구간을 찾아내는 샷 세그먼트 알고리즘에 의해 감지되어 만들어 진다. 만약 입력과정에서 해당 샷 세그멘테이션 파일이 없을 경우에 Video AnnEx 툴은 IBM CueVideo Shot Detection 툴킷을 이용하여 이를 발생시킨다. 다시 말해, Video AnnEx는 비디오 시퀀스를 이루고 있는 비디오 샷을 어노테이션의 단위로 하여 동작하는 시스템으로써 비디오 시퀀스와 샷 세그멘테이션 파일을 함께 입력으로 받아 사용자 어노테이션 작업을 수행한다. 사용자는 VideoAnnEx 툴이 제공하는 정적인 화면과 관련된 단어, 키 객체에 대한 묘사, 그리고 이벤트 관련 단어 등의 어휘목록을 활용하여 비디오 샷에 어노테이션을 남길 수 있다. 이 과정에서 선택된 어휘 및 사용자 입력어는 MPEG-7 xml 형태의 파일로써 저장된다. 한 번 저장된 MPEG-7 xml 파일은 추후에 해당 비디오 시퀀스가 열람될 때 함께 사용되어 시각화된다.

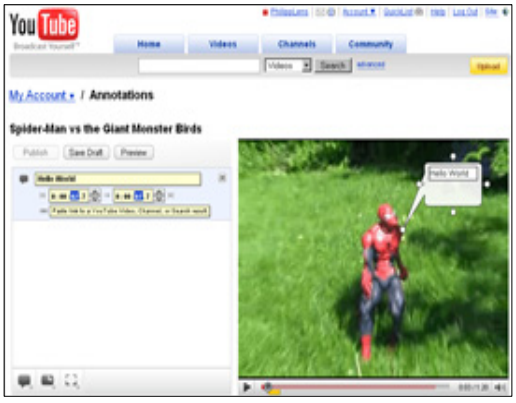
그러나 VideoAnnEx 툴은 사용자가 세그먼트 구간을 직접 명세할 수 있는 기능을 지원하지는 않는다. 앞서 설명한대로, 세그먼트 구간은 미리 정해진 shot segment 알고리즘에 의해 결정될 뿐더러 사용자는 어노테이션 작업을 하기 위해 복잡한 온톨로지 기반의 편집 인터페이스를 사용해야 한다. 이 때문에 VideoAnnEx는 윈도우 기반 시스템에서 실행되고 있으며, 현재 모바일 장치를 통해서는 사용할 수 없다.

2.2 YouTube Annotations

YouTube Annotations는 YouTube가 개발한 영상 어노테이션 툴이다[22]. 사용자는 자신이 업로드한 동영상에 대한 주석이나 말풍선을 남길 수

[그림 1] IBM의 VideoAnnEx의 모바일 인터페이스 화면

있으며 사용자에 의해 어노테이션된 영상은 추후에 영상 공개여부 선택에 따라 다른 사용자들에게 공개될 수 있다. 사용자의 어노테이션 활동을 위해 YouTube Annotations는 사용자 인터페이스를 어노테이션 타임라인, 비디오 플레이어, 그리고 어노테이션 속성 패널의 3부분으로 나누어 어노테이션 활동을 지원한다. 어노테이션 타임라인은 사용자가 어노테이션 하고자 하는 영상구간, 즉 시간대를 설정하도록 지원하며, 비디오 플레이어는 설정된 시간대의 영상이 시각화되는 부분이다. 타임라인 설정을 통해 어노테이션 하고자 하는 영상구간이 비디오 플레이어 화면에 나타나면, 사용자는 어노테이션 속성 패널 창을 이용하여 자신만의 어노테이션 정보를 영상에 기입할 수 있다. YouTube의 Annotations를 통해 기입할 수 있는 어노테이션 타입은 스피치 버블, 노트, 그리고 스포트라이트가 있다. 스피치 버블은 말풍선 기능으로 영상에 기입하는 형태이며 노트는 텍스트입력의 기입을, 그리고 스포트라이트는 사용자와 소통하는 기능으로써 마우스 오비 시 라인과 텍스트가 출력된다. 3가지 타입 모두 어노테이션 시, 다른 YouTube 동영상, 채널 및 검색 결과에 반응할 수 있도록 링크를 붙일 수 있다.



[그림 2] YouTube Annotations 서비스

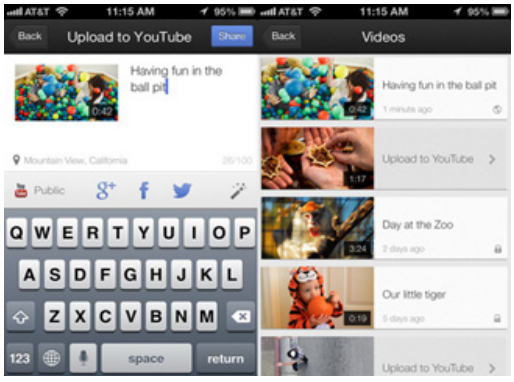
편리한 사용자 인터페이스 덕분에 YouTube의 Annotations는 영상 어노테이션 툴로써 비교적 많은 사용자들에게 알려져 있지만, 자신이 업로드한 영상에 대해서만 어노테이션 작업을 수행할 수 있다는 제한과 웹 기반 인터페이스 디자인으로 인해 모바일 기기 상에서 사용하기에는 불편한 단점이 있다.

2.3 AnViAnno

AnViAnno는 안드로이드 플랫폼에서 실행되는 모바일 어노테이션 서비스이다[15]. AnViAnno는 영상과관련된 MPEG-7 메타데이터 정보저장 방식을 활용하여 영상과 관련된 객체, 특정 이벤트 시간 및 해당 영상이 찍힌 장소 등 다양한 어노테이션 정보를 입력할 수 있도록 지원한다. 어노테이션 작업 후 업로드 된 영상 정보는 SeViAnno 사이트에 접근을 통해 확인할 수 있다.

2.4 YouTube Capture

YouTube Capture는 YouTube 사가 제작한 모바일 기반 영상 어노테이션 프로그램이다[21]. 현재 YouTube은 iOS를 탑재한 모바일 기기에서만 작동되며, 카메라를 통해 캡쳐한 영상 또는 선택한 기존 영상을 SNS에 실시간으로 업로드할 수 있는 기능을 제공한다. 사용자는 YouTube Capture를 통해 영상 길이 편집, 태깅 입력 등의 어노테이션 작업을 수행할 수 있으며, 자동 색보정, 손떨림



[그림 3] YouTube Capture 모바일 서비스

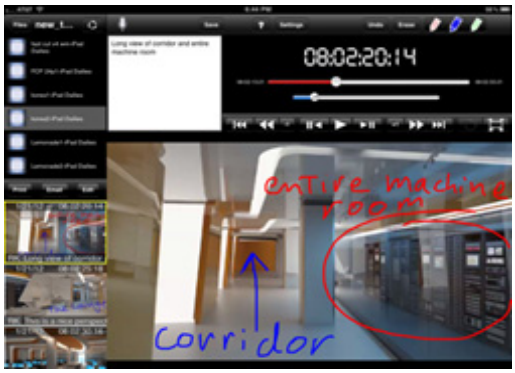
보정, 다듬기 및 음악 트랙 삽입 등의 효과를 주어 영상을 가공할 수 있다. 어노테이션된 영상 정보는 YouTube 서버에 저장되어 추후 다시 활용된다.

2.5 MAMI

MAMI(Multimodal and Mobile Personal Image Retrieval)는 Telefonica 연구소에서 연구 개발한 윈도우 모바일 기반의 음성 어노테이션 프로토타입이다[13]. MAMI는 추후 이미지 검색을 위해 사용자가 입력한 음성 태깅 정보를 저장한다. MAMI는 별도의 서버 없이 사용자가 입력한 음성정보와 그 밖에 메타정보(위치정보, 날짜 및 시간 등)를 로컬 저장소에 보관하여 추후 이미지 검색 시 활용한다.

2.6 Scribbeo

Scribbeo는 iOS 환경에서 실행되는 모바일 영상 어노테이션 프로그램이다[19]. Scribbeo은 개인 또는 그룹간의 공유되는 영상 또는 이미지에 대해 어노테이션 작업을 지원하며, 네트워크를 통해 사용자 간의 변경된 부분을 저장 및 공유할 수 있는 기능을 제공한다. 사용자는 Scribbeo Server 모듈 설치를 통해 자신의 어노테이션 작업을 지인과 공유할 수 잇으며 PDF 또는 HTML 형식의 파일로 저장하여 이메일로 전송할 수 있다.



[그림 4] Scribbeo 모바일 서비스

2.7 MoviBing

MoviBing은 마이크로소프트 연구소에서 개발한 모바일 기반의 자동 영상 어노테이션 프로그램이다[10]. MoviBing은 사용자가 캡쳐 및 선택한 영상정보를 중앙 서버에 전송하여 처리함으로써 실시간 및 자동 어노테이션 서비스 기능을 제공한다. 서버는 수신한 영상 스트리밍과 관련된 이미지 및 관련 링크 정보를 기존의 웹에서 크롤링한 정보와 필터링 알고리즘을 사용하여 데이터베이스에서 검색한다. 이 과정에 거쳐 추출된 링크정보는 중앙 서버에 요청에 의해 Bing 으로부터 얻어와 사용자에게 제공된다. 사용자에 의해 새롭게 선택된 영상 어노테이션 정보는 서버에 저장되어 추후 다시 활용된다.

2.8 숙명여대 음성 주석달기 시스템

숙명여대의 음성 주석달기 시스템은 독서장애인이 청각과 음성을 이용하여 어노테이션 작업을 수행할 수 있도록 연구 및 개발된 시스템이다[1]. 음성 주석달기 시스템은 사용자로부터 입력된 음성을 분석, 랜더링 등의 작업을 거쳐 가공한 후, 음성합성 기술을 적용하여 어노테이션 출력을 수행한다. 가공된 어노테이션 정보는 어노테이션 형식에 따라 XML 파일 형태로 구조화되어 저장되며, 이 과정에서 어노테이션의 이름, 위치, 링크 정보가 함께 저장된다.

위에서 본 바와 같이 모바일 기반의 영상 어노테이션에 관한 학술적, 상업적 수준의 다양한 연구가 수행되었다. 위에서 제시한 비디오 어노테이션 시스템들 외에도 Marquee[7]나 VAnnotator[9]와 같이 다양한 영상 어노테이션 프로그램들이 특수 효과 지원을 위해 제안되었다.

그러나 이러한 툴 및 프로그램들이 제공하는 어노테이션 인터페이스 방식은 여러 가지 제약사항이 많은 모바일 기기 상에서 수행되기에는 부담이 되고 있다. 주된 문제점들은 다음과 같다.

1. 현재 대부분의 모바일 기반 영상 어노테이션 프로그램들은 사용자에게 다양한 기능을 제공하기 위해 복잡한 화면 구성을 따르고 있다. 이는 제약적인 화면크기의 모바일 기기에서 사용하기에는 다소 번거로운 작업으로써 숙련된 기술을 요구한다.
2. 선행된 모바일 기반의 영상 어노테이션 시스템들은 단순히 터치스크린이나 키패드를 터치하는 방식에 크게 의존하고 있다. 이는 사용자의 다양한 제스처를 고려하지 않은 단순 인터페이스 설계방식으로써 직관성이 떨어지며 프로그램 기능이 다양해질수록 화면 구성이 복잡해져 사용에 불편함을 준다.
3. 셋째, 현재 나와 있는 대부분의 모바일 기반의 영상 어노테이션 시스템은 주로 비디오 세그먼트(영상 구간)가 아닌 특정 비디오 프레임(영상 장면)에 어노테이션 기입을 하는 것에 초점을 맞추고 있다.

위에서 언급한 문제들을 해결하기 위한 솔루션으로써 우리는 단순하고 직관적인 비디오 어노테이션 방식인 워키태깅을 제안한다. 워키태깅의 핵심적인 의의는 사용자의 음성 및 직감적인 제스처 방식을 활용해 어노테이션 작업을 수행함에 있다. 이를 위해 본 연구는 기존의 음성인식 기술 및 사용자의 제스처에 대한 연구들을 분석하였으며[3, 4, 11], 이를 워키태깅 인터페이스 설계에 적용하였다. 워키태깅 시스템의 상세 설명은 제 3장 시스템 디자인 섹션에서 설명하도록 하겠다.

3. 시스템 디자인

워키태깅은 안드로이드 기반 모바일 플랫폼에서 실행되는 영상 어노테이션 서비스이다. 워키태깅은 워키토키(무전기)의 무전 방식에서 착안한 개념으로써 기존의 영상 어노테이션 인터페이스를 개선하여 사용자에게 단순하면서도 직감적인 인터페이스를 제공하는 것을 목표로 설계되었다. 특히, 기존의 모바일 영상 어노테이션 방식이 갖고 있는 주요 문제점인 복잡한 화면 구성, 터치 및 문자입력 방식의 지나친 의존성, 그리고 영상 세그먼트 어노테이션 기능의 부재를 개선하기 위해 워키태깅은 다음의 세 가지 사항을 인터페이스 설계에 반영하도록 하였다.

- 단순한 화면 구성
- 직감적 제스처를 활용한 어노테이션 방식
- 영상 세그먼트에 대한 어노테이션 기능

본 연구는 단순한 인터페이스 화면 설계의 첫단계로써 어노테이션 작업에 필요한 필수 정보를 분석하였다. 영상 프레임(영상 한 컷)과 세그먼트(영상 구간)에 대해 어노테이션 작업 시 필요한 필수 정보는 다음과 같다.

[영상 세그먼트]
1. 어노테이션 단어 입력
2. 어노테이션 시작시간 정보
3. 어노테이션 끝 시간 정보

[영상 프레임]
1. 어노테이션 단어 입력
2. 어노테이션 시점 정보

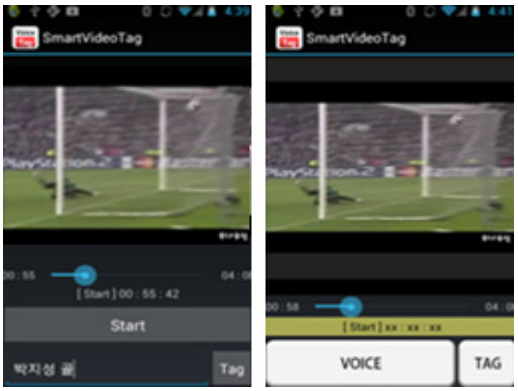
워키태깅은 부차적인 기능을 줄이고 어노테이션에 필요한 기능 위주로 인터페이스를 구성할 수 있도록 노력하였다. 대신, 사용자의 음성과 직감적인 제스처를 활용하여 어노테이션 작업을 수행할 수 있도록 인터페이스를 설계함으로써 사용자의 편리성을 도모하도록 하였다. 음성과 사용자 제스처의 활용으로 워키태깅 인터페이스는 기존의 터치 및 문자 입력 의존성에서 탈피하였으며, 영상 세그먼트에 대한 어노테이션 작업까지 가능하도록 설계될 수 있었다.

이처럼 워키태깅 인터페이스는 설계과정에서부터 사용자의 편리성을 높이기 위해 고려되었다.

이는 기존에 모바일 영상 어노테이션 프로그램들이 터치 및 문자 입력을 통해 어노테이션 서비스를 제공한 것과는 차별되는 점이다. 결과적으로 워키태깅은 기존의 영상 어노테이션 프로그램들이 갖고 있던 복잡한 제어 및 입력 방식에서 벗어나 단순하고 직감적인 방식을 취함으로써 모바일 환경에서 사용이 용이하도록 설계되었다.

3.1 사용자 인터페이스 디자인

워키태깅 사용자 인터페이스는 모바일 환경에서 손쉽게 사용할 수 있도록 최대한 단순화시켜 설계되었다. 그 결과, 워키태깅 사용자 인터페이스는 영상 재생을 위한 영상 플레이어 화면과 어노테이션 시점 설정을 위해 필요한 'Voice' 버튼으로 구성되어 있다. 본 연구는 사용자 실험을 위해 별도로 'Tag' 버튼을 두어 리액션 타임을 측정할 수 있도록 하였다.



(a) (b)

[그림 5] 기존 어노테이션 인터페이스(a) 및 워키태깅 인터페이스(b)

기능적인 면에서 워키태깅은 사용자가 직접 어노테이션 하고자 하는 영상 시점을 자유롭게 설정할 수 있으며, 설정한 구간범위 전체(부분 영상)에 대한 어노테이션 및 설정한 시점 순간의 장면(영상 한 컷)에 대한 어노테이션 모두 가능하도록 인터페이스를 설계하였다. 이 과정에서 구간 영상 및 순간 영상 어노테이션 작업을 위해 필요한 메타데이터 정보를 각 각 분석하였고, 사용자 입장에서 각 모드 선택 후에는 동일한 인터페이스를 사용할 수 있도록 설계하였다. 결과적으로 워키태깅 시스템의 사용자 인터페이스는 구간영상 및 순간영상에 대한 어노테이션 기능모두 지원하면서도 사용자 입장에서 동일한 인터페이스를 통해 제어할 수 있도록 설계함으로써 어노테이션 작업의 부하를 줄일 수 있도록 하였다[그림 5](b). 다음은 각 모드별 동작을 위해 분석한 메타데이터 정보 및 사용자 인터렉션에 대한 설명이다.

3.1.1 순간 영상(영상 한 컷) 어노테이션

워키태깅 시스템 상에서 사용자가 순간 영상에 대한 어노테이션을 하기 위해서는 'voice' 버튼을 누른 채로 원하는 어노테이션 음성 정보를 입력한 후, 버튼을 떼면 된다. 사용자가 지정한 어노테이션 시점시간은 'voice' 버튼을 클릭하는 시점으로 정해지며, 마이크로폰을 통해 입력된 음성 어노테이션 데이터와 누른 시점의 영상 시간정보는 추후에 워키태깅 원격서버에 전송된다. 워키태깅 시스템은 사용자의 음성 정보를 정확히 포착하기 위해 어노테이션을 남기는 시점의 영상에 대한 음소거 작업을 수행한다.

3.1.2 구간 영상(부분 영상) 어노테이션

구간 영상에 대한 어노테이션은 기본적으로 순간 영상에 대한 어노테이션과 같은 원리로 동작한다. 차이점은 순간 어노테이션 모드는 영상 한 컷에 대한 어노테이션이기 때문에 'voice' 버튼을 누른 시점의 시간 정보만 필요했지만, 구간 영상 어노테이션은 사용자가 지정한 구간정보를 알기 위해 구간의'끝 시간'에 대한 정보가 추가적으로 필요하다. 워키태깅 시스템은 '끝 시간'에 대한 정보 수집을 위해 두 가지 방식의 인터페이스를 제안한다.

첫째는 'voice' 버튼을 누른 시점부터 손을 떼는 시점까지의 시간을 측정하는 것이다. 즉, 사용자가 버튼을 누르고 있는 지속 기간을 측정하는 원리이

다. 워키태깅 시스템은 사용자가 'voice' 버튼을 누른 시점과 지속 기간을 측정함으로써 사용자가 어노테이션 하고자 하는 구간의 끝 시간을 파악할 수 있다. 그러나 어노테이션 하려는 구간이 길어지게 되면(예를 들어, 10초 이상) 사용자에게 불편함을 줄 수 있어 짧은 구간의 어노테이션에 적합한 방식이다.

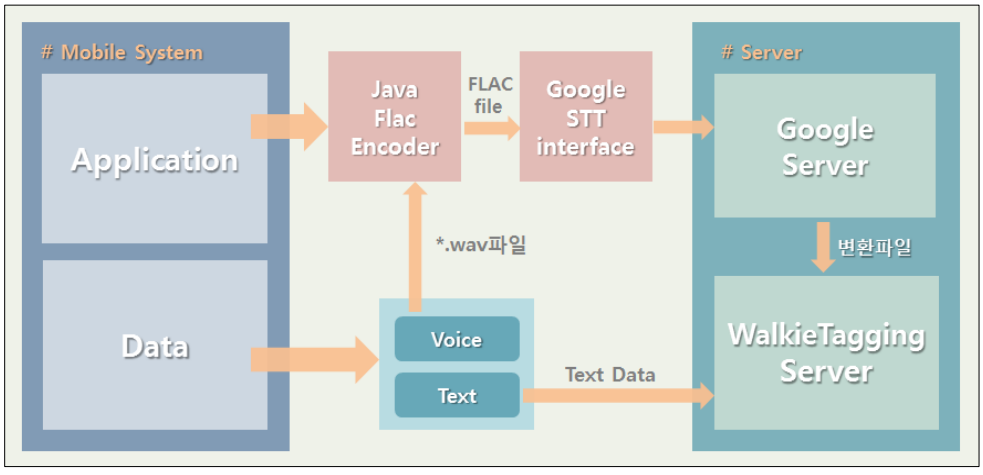
둘째는, 손가락을 누른 채 드래깅하는 방식이다. 이 방식은 부분적으로 첫 번째 방식과 같지만, 어노테이션 하고자 하는 구간을 손가락을 누른 상태로 드래깅하여 빠르게 검색 및 지정할 수 있다는 점이 다르다. 이 모드는 기존의 모바일 인터페이스에서 손가락을 이용하여 화면 확대 또는 축소를 하는 사용자 경험에 착안하여 고안된 방식으로써, 사용자에게 친숙한 인터페이스를 제공한다. 본 연구는 이 모드에 대해 한 손가락을 사용하는 방식과 두 손가락을 사용하는 방식으로 세분화하여 설계하였다. 한 손가락 사용방식은 한 손가락을 터치한 상태에서 위 또는 아래로 드래깅하여 사용하는 방법이고, 손가락 두 개를 사용하는 방식은 두 손가락을 터치한 상태에서 두 손가락의 간격을 넓혔다 좁혔다 하면서 사용하는 방식이다. 두 손가락을 사용하여 드래깅을 하는 경우, 워키태깅 시스템은 영상 구간 길이(시간)를 측정하기 위해 사용자가 손가락 드래깅을 지속한 시간과 두 손가락 사이의 거리 간의 수학적 모델링을 통해 계산하였다. 구체적인 공식은 다음과 같다.

$$\text{지속 시간} = (\text{드래깅 시간}) \times (\text{손가락 사이의 거리})/C$$

위의 공식에서 손가락 사이의 거리는 픽셀단위로 계산되며, C는 손가락 동작에 대한 민감도 상수이다. 본 연구에서는 C의 값을 600으로 하여 지속시간을 측정하였다.

3.2 시스템 아키텍처 디자인

구조도 [그림 3]은 워키태깅 서비스의 전체 시스템 아키텍처이다. 워키태깅 시스템은 크게 모바일 플랫폼과 원격서버 두 부분으로 나뉜다. 모바일 플랫폼은 사용자가 실질적으로 워키태깅 서비스를 실행하는 환경으로써 워키태깅 애플리케이션 구동에 필요한 모듈 및 서비스 그리고 사용자가 수행한 어노테이션 작업들을 원격 서버로 전송하기 위한 모듈들을 포함하고 있다. 원격서버는 모바일 플랫폼 상에서 워키태깅 서비스를 통해 발생한 사용자의 음성 어노테이션 데이터를 가공 및 처리하는 모듈과 이들 데이터 및 관련 메타 데이터들을



[그림 6] 워키태깅 시스템 아키텍처 구조도

저장하고 관리하는 데이터베이스로 구성되어 있다.

본 연구를 통해 개발한 워키태깅 서비스는 사용자에게 음성기반의 어노테이션 방식을 제공하기 위해 구글에서 배포한 음성인식 프로그래밍 인터페이스 라이브러리를 활용하였다. 또한 사용자들의 음성이 저마다 다르고, 워키태깅을 사용하는 환경이 대부분 일상적인 공간이라는 점을 감안하여 오디오 데이터의 무손실 압축을 지원하는 FLAC 파일형식으로 음성을 저장하도록 설정하였다. 영상 어노테이션에 필요한 시스템 동작 과정은 다음과 같다.

1. 일단 워키태깅 애플리케이션이 실행되면 구글 음성인식 라이브러리와 자바 FALC 인코딩 라이브러리가 프로그램과 함께 메모리에 로드된다.
2. 이 후, 프로그램이 포그라운드에서 수행되는 동안 사용자가 발생시킨 음성 데이터는 자바 FLAC 인코딩 라이브러리에 의해 FLAC 파일로 변환되고 구글의 스피치-텍스트 변환 라이브러리에 의해 텍스트 형태로 변환된다.

이러한 과정을 거쳐 변형된 사용자의 음성 어노테이션 정보는 구글 서버를 거쳐 최종 변형된 텍스트 파일로 워키태깅 원격 서버에 저장된다. 요약하면, 본 연구에서 개발한 워키태깅은 모바일 환경에서 실행되는 음성기반의 어노테이션 서비스로써 구글 음성 인식 라이브러리와 자바 FLAC 인코딩 라이브러리를 통해 사용자의 음성 어노테이션 데이터를 텍스트로 변환하여 원격 서버에 저장한다.

4. 실험설계

본 연구는 워키태깅이 제안하는 인터페이스 방식의 사용성 평가를 위해 YouTube Capture를 비교평가 실험 하였다. YouTube Capture는 현존하는 모바일 영상관련 서비스 프로그램 중 가장 많은 사용자를 보유하고 있으며, 전체 앱 사용 순위에서도 높은 순위를 차지하고 있다[16, 17]. 실험은 YouTube Capture와 워키태깅의 어노테이션 입력 방식에 대한 비교평가가 이루어 지도록 설계하였다. 또한 추가적으로 워키태깅이 새롭게 제안하는 영상 세그먼트(구간영상) 어노테이션 기능의 사용성 평가를 위해 세그먼트 어노테이션 인터페이스를 별도로 설계하여 실험하였다.[1] 워키태깅은 영상 세그먼트의 인터페이스로써 두 가지 방식의 손가락 제스처 활용을 제시하였고, 두 방식에 대한 비교 평가도 실험하였다. 본 연구에서 수행한 실험 설계를 다음의 표와 그림을 통해 요약하였다.

〈표 1〉 영상 프레임의 어노테이션 방식

Mode	어노테이션 방식
YouTube	텍스트 입력
워키태깅	음성 입력

〈표 2〉 영상 세그먼트의 어노테이션 방식

Mode	어노테이션 방식
YouTube	텍스트 입력
워키태깅	음성 및 한 손가락 제스처 사용
워키태깅	음성 및 두 손가락 제스처 사용

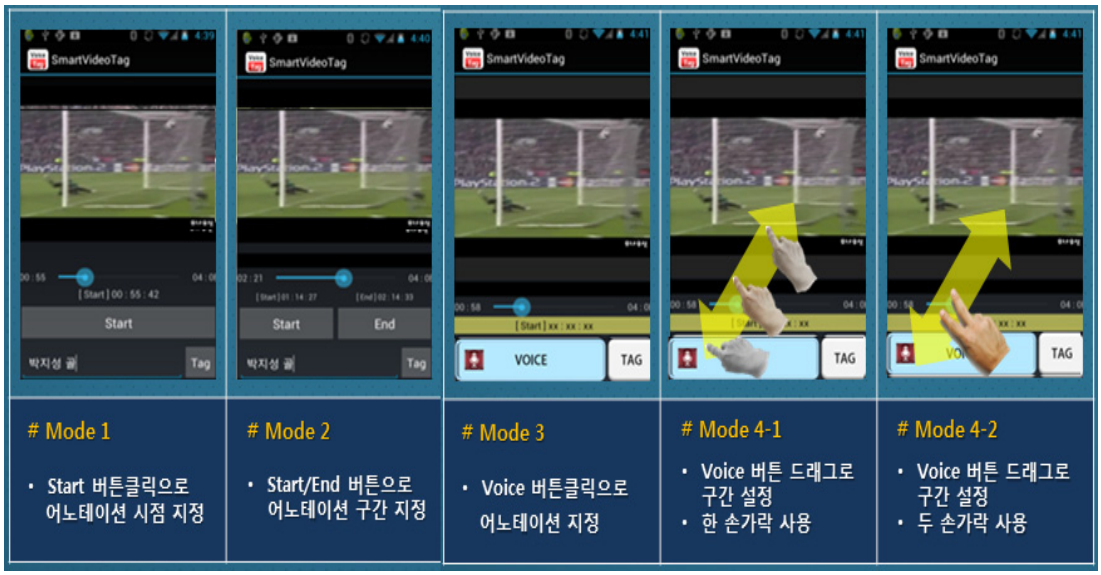
• Mode 1

Mode 1은 YouTube Capture에서 사용하는 영상 프레임에 대한 문자입력 기반의 어노테이션 방식이다. Mode 1은 어노테이션 하고자 하는 시점에서 Start 버튼을 누른 후, 터치패드를 이용한 텍스트 입력으로 어노테이션 작업을 수행한다.

• Mode 2

Mode 2는 YouTube Capture에서 사용하는 문

1) 현재 YouTube는 영상 세그먼트에 대한 어노테이션 기능 지원을 하지 않고 있다. 이 때문에 본 연구는 터치 및 문자입력을 통해 어노테이션하는 기존의 인터페이스 방식에 기반한 영상 세그먼트 어노테이션 인터페이스를 별도로 설계하였다. 그리고 이를 워키태깅의 영상 세그먼트 어노테이션 인터페이스와 비교 평가하는 실험을 진행하였다.



[그림 7] 각 어노테이션 동작 방식에 대한 설명 및 사용자 인터페이스 화면. 워키태깅의 구간 영상 어노테이션 방식은 손가락 제스처 모드에 따라 2가지 버전으로 세분화되었다

자입력 방식에 기반하여 새롭게 설계된 영상 세그먼트 어노테이션 인터페이스이다. Mode 2는 어노테이션 하고자 하는 구간의 시작점에서 START 버튼을, 구간의 끝점에서 End 버튼을 누른 후, 터치패드를 이용한 텍스트 입력으로 어노테이션 작업을 수행한다.

• Mode 3

Mode 3은 어노테이션 하고자 하는 시점에서 Voice 버튼을 누른 채로 음성을 남겨 어노테이션 작업을 수행한다.

• Mode 4-1

Mode 4-1는 어노테이션 하고자 하는 시점에서 손가락 하나로 Voice 버튼을 눌러 드래깅을 하면서 영상 구간을 설정한 다음, 음성을 남겨 어노테이션 작업을 수행한다.

• Mode 4-2

Mode 4-2는 어노테이션 하고자 하는 시점에서 두 손가락으로 Voice 버튼을 누른 후, 두 손가락 간격을 넓혔다 좁혔다 하여 영상 구간을 설정한다. 영상 구간이 설정되면 Mode 4-1과 같이 음성을 남겨 어노테이션 작업을 수행한다.

본 연구는 실험과정에서 YouTube Capture의 인터페이스와 동일한 방식의 실험용 앱을 만들어 이를 사용하였다. 실제 실험 설계 과정에서는 YouTube Capture와 워키태깅 인터페이스 방식에 대해 순차적으로 번호를 부여하였다. 이는 실험의 공정성을 위한 실험설계로써 실험 참가자들이 실험 과정에서 배포용 YouTube Capture 인터페이스를 사용했을 때 발생할 수 있는 편견을 배제하도록 하였다. 또한 워키태깅 인터페이스에 대해서도 실험 전에 미리 해당 인터페이스가 워키태깅 인터페이스임을 밝히지 않은 채로 실험에 참여하도록 함으로써 객관성을 확보하도록 노력하였다.

워키태깅 사용자 실험은 안드로이드 4.1 플랫폼 [Ice Cream Sandwich]에서 수행되었으며, 각 실험 참가자들 별로 4가지 모드를 모두 사용한 후에 설문 및 인터뷰에 응하도록 하였다. 이는 설문과 실험에 대한 사전 학습효과를 방지하여 사용자의 편향을 줄이기 위한 실험 설계였다. 설문지 실험은 4

가지 모드 사용 실험을 완료한 실험 참가자에 대해 수행되었으며, 정량적 평가를 위해 객관문항의 설문을 제시하여 답변하도록 요구하였다. 또한 정성적 차원의 실험 분석을 위해 설문 실험을 완료한 실험 참가자에 대해 개별 인터뷰를 진행함으로써 심층적인 사용성 평가가 이루어지도록 하였다. 실험 참가자 별로 인터뷰한 내용은 현장에서 녹음되어 기록되었다. 사용자 실험을 통해 측정하고자 한 평가 항목은 크게 2가지이며, 각 인터페이스 상에서 어노테이션 작업에 걸리는 작업 시간과 어노테이션 방식의 편리성 평가가 이에 해당한다. 각 항목에 대한 구체적인 실험 설계 내용은 다음 절에서 언급한다.

4.1 사용 작업 시간 평가

사용자들의 수월한 어노테이션 작업을 위해서는 인터페이스 반응 및 작업 시간이 짧아야한다. 어노테이션을 위한 인터페이스 사용 시간이 길다는 것은 그만큼 수행해야 할 작업량이 많다는 것을 의미하기 때문이다. 이에 따라 본 연구는 각 영상 어노테이션 방식에 대한 성능평가 지표로써 반응 속도에 대해 측정하였다.

작업 시간 측정을 위해 본 연구는 각 모드 별로 사용자가 어노테이션 작업을 시작해서 끝 마치는데 걸리는 시간을 모바일 장치를 통해 측정할 수 있도록 프로그래밍 하였다. 안드로이드에서 기본적으로 제공하는 타이머 함수를 응용하여 코드 상에서 사용자의 어노테이션 작업이 수행되는 시점의 시간과 어노테이션 작업이 끝나는 시점의 시간을 각 각 측정하여 두 변수 간의 차이를 구하였다. 얻어진 결과 값은 사용자 인터페이스를 통해 시각화시켜 실험참가자들이 자신의 작업 시간을 확인할 수 있도록 설계하였다.

4.2 어노테이션 방식의 유용성 평가

본 연구에서는 어노테이션 방식의 유용성 평가를 위해 Likert-Scale 방식의 객관식 설문문항을 사용하였다. 실험참가자는 모든 어노테이션 방식에 대해 사용이 얼마나 편리한 지를 평가하도록 요구받았다. 본 실험에서 Likert-Scale 설문기법을 사용한 이유는 사용자가 답변으로 선택한 객관문항 번호를 평점으로 간주하여 누적합산에 대한 평균값으로 사용하기 위함이다. 또한 설문문항을 마친 사용자에 대해서는 답변에 대한 추가적인 이유를 물어 기록하였다. 실험에서 사용한 구체적인 설문문항은 다음과 같다.

[유용성 평가]
- 사용하신 방식이 편리하십니까?

1. 매우 편리하지 않다.
2. 편리하지 않다.
3. 그저 그렇다.
4. 편리하다.
5. 매우 편리하다.

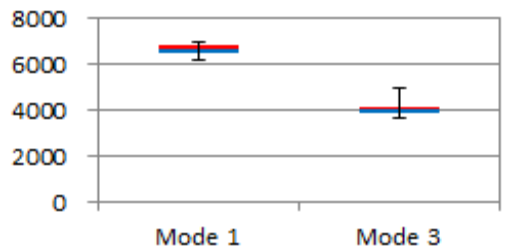
- 위 문항에 대한 답변이유는 무엇입니까?

5. 실험 결과

실험은 총 46명의 참가자들로 진행되었다. 참가자 구성은 학생 26명, 직장인 14명, 기타 4명이었으며, 성비 구성은 남성 38명, 여성 8명이었다.

5.1 작업 시간 실험 결과

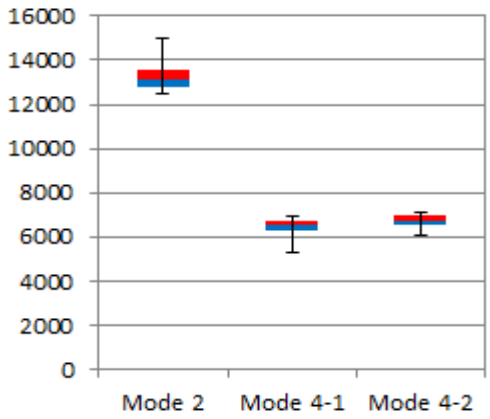
순간 영상(영상 한 장면) 어노테이션의 작업 시간 측정 실험 결과, 기존의 어노테이션 방식(mode



[그림 8] 순간 영상 어노테이션 작업 시간 평균값(단위 : ms)

1)에 비해 워키태깅(mode 3)방식의 작업 시간이 평균적으로 훨씬 적게 걸린 것을 관찰하였다. 이는 터치패드로 텍스트를 입력하는 시간보다 음성으로 어노테이션 하는 시간이 더 적기 때문인 것으로 나타났다.

본 실험에서는 두 어노테이션 방식의 정확한 비교 평가를 위해 실험 참가자들에게 동일한 영상 컨텐츠 및 영상 시점을 제공한 후, 어노테이션 작업을 수행하도록 요청하였다. 구체적인 순간 영상 어노테이션 작업 시간의 평균값은 mode 1이 6,659ms, mode 3이 4,044ms를 기록하였다.



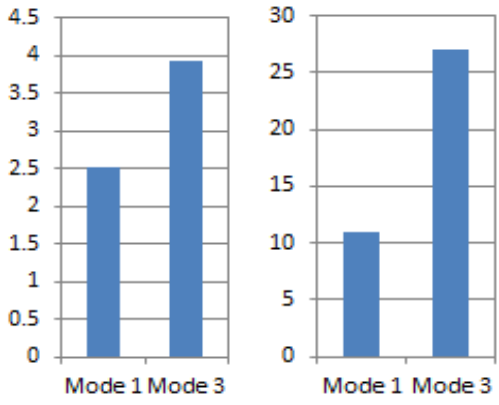
[그림 9] 구간 영상 어노테이션 작업 시간 평균값(단위 : ms)

구간 영상(부분 영상) 어노테이션의 작업 시간 실험에서도 역시 워키태깅 방식의 작업 시간이 더 짧게 나타났다. 기존 방식인 mode 2의 경우 어노테이션 구간의 끝 장면이 나올 때 까지 기다려야 했던 반면에, 워키태깅 방식은 'Pinch in and out' 방식을 통해 구간의 끝을 빠르게 검색할 수 있어 어노테이션 작업 시간이 단축된 것으로 드러났다.

구간 영상 실험 역시 정확한 비교 평가를 위해 실험 참가자들에게 동일한 영상 콘텐츠 및 영상 구간을 제공한 후, 어노테이션 작업을 수행하도록 요청하였다. 평균 시간은은 mode 2가 1,3240ms, mode 4-1이 6,340ms, 그리고 mode 4-2가 6,705ms를 기록하였다.

5.2 사용자 편리성 평가 실험 결과

순간 및 구간 영상에 대한 사용자 편리성 실험에서 기존 어노테이션 방식에 비해 워키태깅 방식의 평균 점수가 모두 더 높게 나타났다. 순간 영상에 대한 실험에서 두 방식에 대한 평균 평점은 각각 2.52점, 3.93점으로 나타났다. 구간 영상에 대해서는 mode 2가 평균 2.81점, mode 4-1이 3.952점, mode 4-2가 3.67점을 기록하였다. 또한 각 모드별로 사용자로부터 편리성 평점 4점('유용하다'이상을 답변 한 경우)이상을 획득한 사용자 수를 집계한 결과에서도 워키태깅이 기존 방식보다 더 좋은 결과를 보였다. 구체적으로 순간 영상 실험에서 기존 모드가 11명의 사용자로부터 4점 이상을 받은 반면, 워키태깅은 실험 참가자의 절반 이상인 27명으로부터 같은 답변을 얻었다. 구간 영상 실험에서는 기존 방식이 13명, 워키태깅 방식이 각 각 28명, 23명의 사용자로부터 이와 같은 평가를 받았다.



[그림 10] 순간 영상 어노테이션에 대한 사용자의 편리성 평균 점수(a) 및 평점 4점 이상을 획득한 사용자 수(b)

사용자의 편리성에 대한 설문 후 진행된 인터뷰에서 대부분의 실험 참가자들은 워키태깅 서비스가 편리한 이유로 음성의 사용 및 단순한 인터페이스를 언급하였으며, 빠른 어노테이션이 가능하

기 때문이라는 의견도 있었다. 다음은 몇 몇 실험 참가자들의 발췌된 인터뷰 내용이다.

모드 3은 어노테이션을 할 때 타이핑해야 하는 부분이 없어서 편했다. 평소에 문자를 보낼 때 오타 때문에 짜증이 많이 나는데 어노테이션 작업도 문자입력을 통해서 해야 한다면 별로 사용하고 싶지 않았을 것 같다." [실험참가자 1]

"*말로 태깅을 한다는 개념이 신선하고 편했다. 손으로 치는 작업이 없으니까 마치 자동으로 주석을 다는 느낌이 든다. 내가 말한 태깅을 가지고 나중에 영상을 검색할 때 활용할 수 있다면 더 좋을 것 같다.*" [실험참가자 2]

모드 1이나 모드 3이나 둘 다 나에겐 익숙한 방식이다. 그러나 모드 3은 말로 하는 방식이라 확실히 빠른 것 같다. 그래서 편하다고 느꼈다." [실험참가자 3]

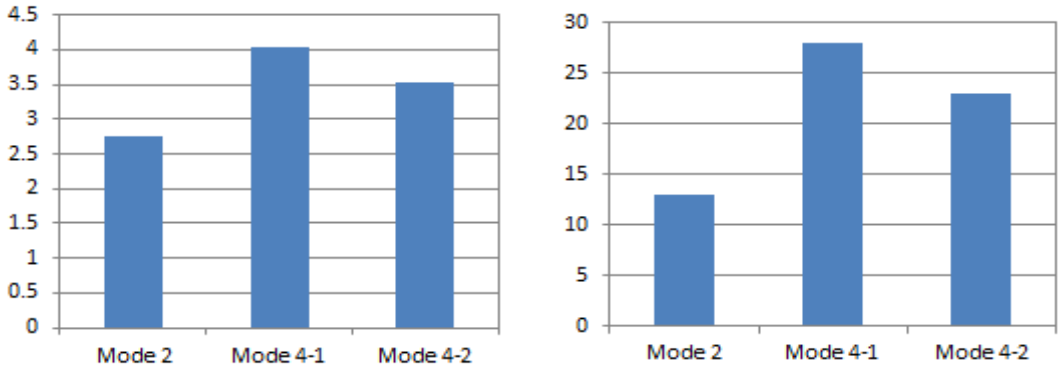
구간 어노테이션에 대한 사용자 편리성 평가 역시 워키태깅 방식이 사용자에게 더 편리함을 주고 있음을 관찰하였다. 또한, 손가락 사용 개수에 따른 방식 차이에서는 손가라 하나를 사용한 어노테이션 방식이 두 손가락을 사용한 방식보다 사용자에게 더 편리함을 주는 것으로 나타났다. 각 방식에 대한 평균평점은 각 각 2.76점, 4.04점, 3.52점으로 손가락 하나를 사용한 어노테이션 방식이 가장 좋은 평가를 얻었다.

mode 4-1이 편하다고 느낀 이유로는 직감적이라는 것과 구간설정 제어가 가능하다는 이유가 많았다. 다음은 실험참가자들의 발췌된 인터뷰 내용 일부이다.

"*손가락의 제스처를 활용하여 내가 어노테이션 할 구간을 선택하는 방법이 직감적이다. 맵을 확대하거나 축소할 때 사용했던 방식과 비슷하다고 느껴서 그런 것 같다.*" [실험참가자 1]

"*어노테이션 하는 방식의 정형성을 깬 것 같다. 기계적으로 학습해서 사용한다는 기분이 전혀 들지 않았다. 그래서 편하게 느꼈다.*" [실험참가자 2]

"*버튼 사용보다 손가락을 왔다 갔다 하는 게 더 편하다. 원하는 구간을 빠르게 검색할 수 있기 때문이다. 물론 정확성은 버튼이 훨씬 좋다고 생각한다. 손가락 두 개를 사용하는 것이 하나를 사용하는 것보단 불편했다. 두 손가락을 같이 움직여야 했기 때문에 그런 것 같다.*"
[실험참가자 3]



(a) (b)

[그림 11] 구간 영상 어노테이션에 대한 사용자의 편리성 평균 점수(a) 및 평점 4점 이상을 준 사용자 수

〈표 3〉 사용자 작업시간 실험 결과에 대한 T-test

(a)

	Mode 1	Mode 3
평균	6659.456522	4043.956522
분산	56762.34251	44456.22029
관측수	46	46
가설 평균차	0	
자유도	89	
t 통계량	55.75753388	
P(T <= t) 단축 검정	2.58428E-71	
t 기각치 단축 검정	1.662155326	
P(T <= t) 양축 검정	5.16857E-71	
t 기각치 양축 검정	1.9869787	

(b)

	Mode 2	Mode 4
평균	13240.95652	6339.543478
분산	329987.5981	114987.8092
관측수	46	46
가설 평균차	0	
자유도	73	
t 통계량	70.16958881	
P(T <= t) 단축 검정	4.78649E-69	
t 기각치 단축 검정	1.665996224	
P(T <= t) 양축 검정	9.57239E-69	
t 기각치 양축 검정	1.992997126	

〈표 4〉 사용자 편리성 실험 결과에 대한 T-test

(a)

	Mode 1	Mode 3
평균	2.52173913	3.934782609
분산	0.343961353	0.462318841
관측수	46	46
가설 평균차	0	
자유도	88	
t 통계량	-10.6731213	
P(T <= t) 단축 검정	7.54825E-18	
t 기각치 단축 검정	1.662354029	
P(T <= t) 양축 검정	1.50965E-17	
t 기각치 양축 검정	1.987289865	

(b)

	Mode 2	Mode 4
평균	2.760869565	4.043478261
분산	0.319323671	0.353623188
관측수	46	46
가설 평균차	0	
자유도	90	
t 통계량	-10.60731721	
P(T <= t) 단축 검정	7.99877E-18	
t 기각치 단축 검정	1.661961084	
P(T <= t) 양축 검정	1.59975E-17	
t 기각치 양축 검정	1.986674541	

실험결과 워키태깅 어노테이션 인터페이스가 작업 시간 및 유용성 평가 측면에서 평균적으로 모두 좋은 점수를 얻었다. 또한 T-Test를 실시하여 두 집단 간의 유의한 평균 차이가 있음을 검증하였다. 본 연구는 실험 설계과정에서 어노테이션 입력 인터페이스가 개선됨에 따라 사용자들의 만족도 및 인터페이스 이용의 효율성이 높아질 것이라는 가설을 세웠다. 그리고 이를 검증하기 위해 문자입력 방식을 사용하는 기존의 모바일 영상 어노테이션 프로그램과 새롭게 제안하는 음성 및 사용자 제스처 기반의 어노테이션 방식을 사용하는 워키태깅을 비교 평가하였다. 두 서비스의 어노테이션 입력 방식 외에 다른 변수 요인을 통제하여 실험함으로써 본 연구는 실험결과에 대한 객관적 타당성을 확보할 수 있었다. 결과적으로 본 연구는 워키태깅 방식이 기존의 방식보다 더 개선된 것임을 증명할 수 있었다. 이는 말하는 음성이 터치패드를 이용한 문자입력보다 작업 시간이 훨씬 빠르고, 익숙한 제스처 방식을 활용하영 어노테이션 하는 작업이 사용자 입장에서 편리하게 느껴졌기 때문이다.

위의 사실들을 통해 본 연구는 모바일 기기에 탑재되는 영상 어노테이션 프로그램은 이용이 간편하면서도 작업 부담이 적도록 설계되어야 한다는 시사점을 얻을 수 있었다.

6. 결 론

본 연구는 모바일 컴퓨팅 환경에서 사용자들이 영상 어노테이션 작업을 수행할 수 있는 효과적인 방식의 워키태깅 서비스를 제안하였다. 워키태깅은 워키토키(무전기)의 무전 방식에서 착안한 개

넘으로써 단순하면서도 직감적인 방식으로 사용할 수 있도록 설계 및 구현되었다. 이 과정에서 본 연구는 기존의 영상 어노테이션 인터페이스 방식이 갖고 있는 한계점을 분석하였으며, 이를 보완하기 위해 사용자 경험에 기반한 다양한 제스처 분석을 수행하였다. 이처럼 워키태깅 실험은 영상 어노테이션 입력 방식이 개선되면 사용자의 만족도 및 작업능률이 향상될 것이라는 전제하에서 출발하였고, 사용자 실험을 통해 이를 증명하였다. 실험결과는 실제로 워키태깅 서비스가 모바일 환경에서 이용하기에 작업부담이 적으며 사용자의 편리성을 높일 수 있다는 점을 보여주었다.

요약하면, 본 연구에서 제안한 워키태깅 서비스의 의의는 다음과 같다. 첫째, 워키태깅은 모바일 환경에서 단순하고 직감적인 인터페이스 제공을 통해 어노테이션 작업의 편리성 및 효율성을 높인다. 둘째, 영상 세그먼트에 대한 태깅 기능은 기존의 토픽-클라우드 추천 및 협업적 추천 시스템 연구 학제 분야에서 활용될 수 있는 이론적 응용을 제시하였다. 셋째, 워키태깅의 손쉬운 작동방식은 영상 어노테이션 작업의 전문가적 이미지를 탈피하여 범용적 사용 및 참여를 유도한다.

본 연구는 보다 정밀한 실험 평가를 위해 동일한 방식기반의 시스템과 비교평가를 했어야 했으나 현 존하는 음성기반 어노테이션 프로그램의 부재로 수행을 하지 못하였다. 또한 충분한 실험 참가자 수를 확보하지 못해 실험 결과에 대한 통계적 산출의 의미가 상대적으로 부족하였다.

이러한 부분을 개선하기 위해 본 연구는 더 많은 실험 참가자를 확보하고 지금보다 더 다양한 어노테이션 기법(예 : 실시간 말풍선 삽입)들이 직관적인 방식으로 워키태깅 시스템 기능에 포함될 수 있도록 구현 및 실험할 계획이다. 또한 이론적 활용 측면에서 영상 세그먼트의 태깅 정보가 영상 검색 시스템을 통해 활용될 수 있는 방안을 연구할 것이다.

본 연구에서 제안하는 워키태깅 서비스가 제약조건이 많은 모바일 환경의 영상 어노테이션 서비스 연구에 새로운 활력을 불어 넣어 줄 것으로 기대한다.

참 고 문 헌

[1] 이경희, 이종우, 임순범, “디지털 말하기 책을 위한 음성 주석달기 시스템”, 「한국정보과학회논문지, 컴퓨팅의 실제 및 레터」, 제18권, 제4호(2012), pp.276-281.

[2] 이연호, 오경진, 신위살, 조근식, “링크드 데이터를 이용한 협업적 비디오 어노테이션 및 브라우징 시스템”, 「한국지능정보시스템학회」, 제17권, 제3호(2011), pp.203-219.

[3] Harrison, B. and R. Baecker, “Designing video annotation and analysis systems”, *Proceedings of the conference on Graphics interface*, (1992), pp.157-166.

[4] Pakucs, B., “Butler : A Universal Speech Interface for Mobile Environments”, *MOBILEHCI, Lecture Notes in Computer Science*, Vol3160(2004), pp.399-403.

[5] Yeo, B. and M. Yeung, “Retrieving and visualizing video”, *Communications of the ACM*, Vol.40, No.12(1997), pp.43-52.

[6] Guo-Jun, Q. et al., “Correlative multi-label video annotation”, *In proc. of the 15th international conference on Multimedia (MULTIMEDIA)*, (2007), pp.17-26.

[7] Weher, K. and A. Poon, “Marquee : a tool for real-time video logging”, *In proc. of the SIGCHI conference on Human factors in computing : celebrating interdependence (CHI)*, (1994), pp.58-64.

[8] Ames, M. and M. Naaman, “Why we tag : motivations for annotation in mobile and online media”, *In proc. of the SIGCHI conference on Human factors in computing systems(CHI)*, (2007), pp.971-980.

[9] Costa, M. N. Correia, and N. Guimaraes, “Annotations as multipleperspectives of video content”, *In proc. of the tenth ACM international conference on Multimedia (MULTIMEDIA)*, (2002), pp.283-286.

[10] Motaz, E.-S., X.-J. Wang, N. Hasan, M. Bassiouny, and M. Refaat, “Seamless annotation and enrichment of mobile captured video streams in real-time”, *IEEE Internation Conference on Multimedia and Expo (ICME)*, 2011.

[11] Rosenfeld, R., D. Olsen, and A. Rudnicky, “Universal speech interface”, *Interactions*, Vol.8, No.6(2001), pp.34-44.

[12] Volkmer, T., J. Smith, and A. Natsev, “A web-based system for collaborative annotation of large image and video collections : an evaluation and user study”, *In proc. of the 13th annual ACM international conference on Multimedia(MULTIMEDIA)*, (2005), pp.892-901.

[13] Anguera, X., J. Xu, and N. Oliver, “MAM I : Multimodal Photo Annotation and Retrieval on a Mobile Phone”, *In Proc. ACM International Conference on Multimedia Information Retrieval(MIR)*, Vancouver, Canada, 2008.

[14] Zhiwen, Y. et al., “Supporting Context-Aware Media Recommendations for Smart Phones”, *IEEE Pervasive Computing*, Vol.5, No.3 (2006), pp.68-75, doi:10.1109/MPRV.2006.61.

[15] AnViAnno Video Annotations Tool http://tosini.informatik.rwth-aachen.de:8134/media/anvianno/.

[16] Mobile Applications Ranking Informations http://bgr.com/2013/01/23/mobile-apps-rankings-google-303777/.

[17] Most Popular Video Apps Ranking http://wearesocial.sg/blog/2012/02/social-asia-tuesday-tuneup-14-2/most-popular-video-website-by-country-appappeal/.

[18] Research and Markets : Worldwide Smartphone Markets : 2011 to 2015 http://www.businesswire.com/news/home/20110826005231/en/Research-Markets-Worldwide-Smartphone-Markets-2011-2015.

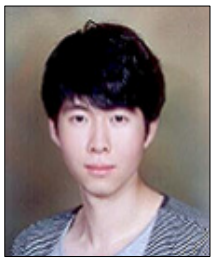
[19] Scribbeo Video Annotations Tool, http://scribbeo.com/.

[20] VideoAnnEx Annotation Tool, http://www.research.ibm.com/VideoAnnEx/.

[21] YouTube Capture Video Annotations Tool, http://www.youtube.com/capture.

[22] YouTube Video Annotations Tool, http://www.youtube.com/watch?v = UxnopxbOdic.

◈ 저 자 소 개 ◈

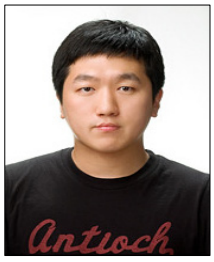


박 준 영 (parkjay@itc.kaist.ac.kr)

현재 KAIST IT융합연구소에서 위촉연구원으로 재직 중이다. 한양대학교 컴퓨터공학과 및 정보기술경영학과 학사학위와 KAIST 지식서비스공학과 석사학위를 취득하였다. 주요 관심분야는 pervasive computing, HCI, 그리고 지식공학 등이다. 구체적으로 위치기반 소셜 컴퓨팅 서비스 및 제스처 기반의 사용자 경험에 관한 연구를 하고 있다.

이 수 빈 (sb.lee@kaist.ac.kr)

현재 KAIST IT융합연구소에서 팀장 및 연구교수로 재직 중이다. KAIST 전기 및 전자공학과 박사학위를 취득하였다. 주요 관심분야는 이동통신시스템, 센서네트워크, 웹, 데이터마이닝, HCI 등이다.



강 동 엽 (dykang@itc.kaist.ac.kr)

현재 KAIST IT융합연구소에서 위촉연구원으로 재직 중이다. KAIST 컴퓨터공학과 학사 및 석사 학위를 취득하였다. 주요 관심분야로는 데이터마이닝과 기계학습 등이다. 구체적으로 probabilistic graphical model, large scale machine learning, matrix factorization 등에 관한 연구를 하고 있다.

석 영 태 (delegacy@gmail.com)

현재 SK텔레콤 솔루션 사업본부 솔루션 개발팀에 재직 중이다. 한국과학기술원 전산학과 학사학위를 취득하였다. 주요 관심분야는 이러닝, 사용자 인터랙션, 데이터마이닝, 소셜네트워크서비스, e-비즈니스 등이다.