사무실 환경 내 다중카메라 영상의 이벤트분석을 통한 자연어 기반 동영상 검색시스템

Natural Language based Video Retrieval System with Event Analysis of Multi-camera Image Sequence in Office Environment

임수정, Soojung Lim*, 홍진혁, Jin-Hyuk Hong**, 조성배, Sung-Bae Cho***

요약 최근 동영상을 저장하거나 효과적으로 검색하는 시스템의 필요성이 증가하고 있다. 기존 동영상 검색 시스템은 메뉴나 텍스트 기반의 키워드로 동작하는데, 검색을 위한 정보가 빈약하여 많은 동영상이 동시에 검색되는 경우가 많고, 사용자가 해당 시스템을 사용하기 위한 지식이 필요하다. 본 논문에서는 키워드 기반 질의에 비해 사용자의 의도나 다양한 정보를 포함 할 수 있고 이벤트나 사람 뿐 아니라 인간의 기본적인 행동까지 검색하는 자연어 질의 기반 동영상 검색 시스템을 제안한다. 먼저, 사무실에서 수집된 동영상에 대한 도메인 분석을 통해 메타데이터로 이벤트 데이터베이스를 구축하고, 해당 영역에서의 사용자 질의에 대한 전처리 과정과 분석 작업을 통해 스크립트 데이터베이스를 구축한다. 이렇게 구축된 시스템을 바탕으로 대화형 질의와 답변 쌍 매칭을 통해 동영상을 검색한다. 10명의 사용자를 대상으로 메뉴 기반 검색 시스템과의 성능 평가와 프로세스 평가 비교 검증을 통해, 성능이나 사용자 만족도면에서 제안하는 시스템이 우수함을 보였다.

Abstract Recently, the necessity of systems which effectively store and retrieve video data has increased. Conventional video retrieval systems retrieve data using menus or text based keywords. Due to the lack of information, many video clips are simultaneously searched, and the user must have a certain level of knowledge to utilize the system. In this paper, we suggest a natural language based conversational video retrieval system that reflects users' intentions and includes more information than keyword based queries. This system can also retrieve from events or people to their movements. First, an event database is constructed based on meta-data which are generated by domain analysis for collected video in an office environment. Then, a script database is also constructed based on the query pre-processing and analysis. From that, a method to retrieve a video through a matching technique between natural language queries and answers is suggested and validated through performance and process evaluation for 10 users. The natural language based retrieval system has shown its better efficiency in performance and user satisfaction than the menu based retrieval system.

해심어: Video retrieval, Natural language based retrieval, Domain analysis, Natural language query

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (IITA-2007-C1090-0701-0011)

*주저자 : 연세대학교 컴퓨터과학과 석사과정

**공동저자 : 연세대학교 컴퓨터과학과 박사과정

***교신저자 : 연세대학교 컴퓨터과학과 교수; e-mail: sbcho@cs.yonsei.ac.kr

1. 서론

최근 동영상에 대한 쉬운 저장 기술과 인코딩 방식으로 인해 동영상 데이터의 양이 크게 늘어났고, 동영상을 효과적으로 검색하는 시스템의 필요성이 증가하고 있다. 많은 연구가 진행 중이지만, 텍스트 문서에 대한 검색 엔진에 비해 동영상 데이터에 대한 검색은 효율적이지 못하다[1].

기존의 동영상 검색 시스템은 주로 메뉴나 텍스트 기반의 키워드로 동작하는데, 검색을 위한 정보가 빈약하여 다수의 동영상이 동시에 검색되는 경우가 많고, 사용자가 시스템 사용을 위한 지식이 필요하기도 한다. 또한 검색의 대상이 되는 데이터가 복잡하고 다양해짐에 따라 검색을 위한 시간 복잡도가 크게 상승하기 때문에 메뉴나 텍스트 기반의 키워드 검색은 한계가 있다. 사용자의 의도를 충분히 반영하지 못하고 시스템에서 제공되는 질의 형태로만 검색 가능한 시스템 환경 역시 기존 동영상 검색 시스템의 단점이다.

본 논문에서는 다양한 동영상의 효과적 검색을 위해 자연어 기반 동영상 검색 시스템을 제안한다. 제안하는 시스템은 사무실 내의 이벤트 분석 과정을 통해 사용자의 의도나 다양한 정보, 발생 가능한 모든 상황에 대해 정의를 하고, 이를 토대로 이벤트나 사람, 나아가 기본 행동까지 검색이 가능하다. 기존의 메뉴 기반 검색 인터페이스와의 성능 평가와 프로세스 평가 비교를 통해 제안하는 방법의 유용성을 검증하였다.

2. 관련 연구

2.1 다중카메라 시스템 및 뷰 선택

다중 카메라 시스템은 한 명의 피험자 또는 하나의 이벤트에 대한 다양한 시각을 가진 영상을 얻을 수 있다. 이러한 시스템에서는 카메라가 커버할 수 있는 영역을 벗어나는 사각지대나, 사람과 물체의 방향에 따라 사용자가 원하는 영상이 가려져서 잘 보이지 않는 경우 등이 발생하지 않아, 대부분의 경우 원하는 영상정보를 얻는다[2].

뷰 선택은 다중 카메라 이벤트 시퀀스로부터 하나의 이벤트 시퀀스를 만들기 위해, 동일한 시점에서 이벤트를 가장 잘 보여주는 최적의 카메라를 선택하는 것으로, 본 논문에서는 동영상 검색의 전처리 과정으로 다중카메라로부터 수집된 동영상 시퀀스들 최적의 뷰를 갖는 하나의 시퀀스로 만들기 위해 각 이벤트별 뷰 선택을 위한 규칙을 사용하였다. 규칙은 사용자로부터 사전에 조사된 이벤트, 사람, 물체 등의 관심도에 대해 도메인 온톨로지의 도메인 지식을 구성한 것을 바탕으로 만들어졌다[3].

2.2 동영상 검색 시스템

동영상을 검색하는 방법은 해당 동영상을 어노테이션 등의 작업을 통해 텍스트를 이용하여 검색하거나 이미지를 기반으로 검색하는 것으로 나뉜다. 텍스트 기반 검색의 경우, 각 키 프레임마다 해당하는 정보와 특징을 통해 생성된 메타데이터를 이용하는 방법을 가장 많이 사용하며[4], 시나리오나 대본, 청각 장애인용 자막을 이용하여 영상을 검색하기도 한다. 이미지 기반 검색은 동영상으로부터 키 프레임을 추출하여 색깔이나 질감, 움직임이나 색깔 히스토그램을 사용하여 이를 비교하여 유사한 영상을 검색한다. 이미지 기반 검색은 보통 시스템 자동화를 위해 사용된다.

Volkmer 등은 대화 스크립트를 사용하여 영상과 텍스트를 매핑하여 주제별로 영상을 분류한 후, 해당 키워드를 기반으로 검색하는 시스템을 제안하였다[5]. 이를 통해 사용자로부터 해당 카테고리를 선택하게 한 후, 텍스트를 입력 받아 검색 엔진으로부터 얻어진 스크린샷을 검색 결과로 제공한다. B. Allen은 검색 기준을 토대로 키워드기반 검색 및 이미지 기반 검색을 제공하는 시스템으로, 이미지를 업로드 하여 유사성과 차이점을 비교 검색하는 시스템을 제안하였다[6]. 하지만 키워드와 영상의 1:1 매칭검색이라는 한계가 있고, 이미지 기반 검색의 경우 오차범위가 크다.

제안하는 사무실 환경 내에서의 자연어 기반 동영상 검색 시스템의 경우, 동영상 자체에 포함 되어 있는 컨텍스트 정보에 대해 사전 이벤트 분석을 통해 정의를 하였기 때문에 보다 자세한 검색이 가능하고, 따라서 단순 이벤트나 사람에 대한 검색뿐 아니라 행동에 대한 검색도 가능하다.

2.3 스크립트 데이터베이스 설계

이벤트의 정의 및 분류를 포함한 도메인 분석은, 주어진 환경을 정의하고 보다 다양한 항목들을 제공하기 위해 선행되어져야 한다. 특히 가장 중점이 되는 이벤트에 대한 정의 및 분류는 기존 연구에서 물체 중심의 활동 분류와 사람 또는 이벤트 중심의 분류로 크게 나누어진다.

J. Fogarty 등은 센서로부터 얻어진 데이터를 통해 시스템 사용자를 방해 가능한 시점에 대한 예측을 위해 물체를 중심으로 각 활동들을 분류하였고[7], D. Moore 등은 인간의 행동 인식을 위해 물체를 중심으로 각 활동을 분류하여 해당 상황을 인지하였다[8].

N. Oliver 등은 다양한 센서로부터 얻어진 사무실 환경 내의 activity에 대해 센서로부터 바로 수집가능한 데이터와 수동 어노테이션이 필요한 의미적 데이터 간의 계층을 설정하여 이벤트를 분류하였고[9], C. Wojek 등은 사람 수에 따라 연구실 환경 내에서의 이벤트를 분류하였다[10].

본 논문에서는 사무실 환경을 빛, 소리 등의 환경 정보, 사무실 구성원, 사무실 내의 물체, 발생하는 이벤트, 이벤트 보다 작은 단위인 기본 행동 등으로 구성하였다. 기존 연구에서 물체나 이벤트에 대해 개별적으로 접근하는 것의 한계를 넘어, 이를 동시에 고려하여 분석하는 과정을 통해 사람 수에 따라 이벤트를 분류 하고 그를 세분화하였고, 이를 통해 보다 자세한 항목

에 대해 검색 가능하도록 하였다.

3. 제안하는 대화형 동영상 검색 시스템

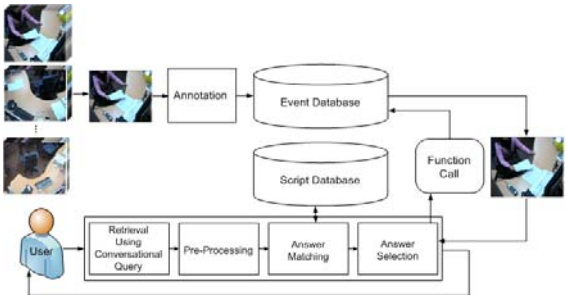

그림 1. 제안하는 방법의 개요

본 논문에서 제안하는 이벤트 검색 방법은 그림 1과 같이 동작한다. 먼저 사무실 환경 내에 설치된 다중카메라로부터 일과에 관한 동영상 데이터를 수집하고, 뷰 선택 과정을 통해 하나의 시퀀스로 만들어진 동영상의 각 이벤트 별 키 프레임 마다 사전에 정의된 컨텍스트 항목으로 어노테이션을 수행한다. 동영상 검색을 위한 자연어 질의는 형태소와 패턴을 분석하여 사전에 정의된 스크립트 데이터베이스로부터 해당 질의와 답변을 매칭한다. 가장 적절한 답변이 선택되면 이벤트 데이터베이스에서 실제 데이터 값을 추출하여 사용자에게 제공한다. 이때, 실제 데이터 값과 같은 투플에 저장되어 있는 해당 이벤트의 시작 프레임 정보와 끝 프레임 정보를 이용하여 검색 결과에 해당하는 동영상을 제공한다.

3.1 이벤트 데이터베이스의 구조 및 생성과정

이벤트 검색을 위한 도메인 분석 및 이벤트 추출을 통해 구성된 사무실 환경은 환경정보(카메라 정보, 시간 정보, 빛 등), 이벤트(entry, leaving, calling 등), 이벤트와 직접적(phone, screen, vacuum cleaner 등)으로 또는 간접적으로(stapler, binder, heater 등) 관련 있는 물체, 사람(사람 수에 따른 구분) 간의 관계로 정의 가능하다. 본 논문에서는 기존 연구들의 확장 모델로, 사람 수에 따른 이벤트의 분류 후 이벤트를 구성하는 primitive motions(standing, sitting, moving 등)과 basic activities(walking, holding, using 등)로 세분화 하였다. 그 후 이벤트를 구성하는 물체와 환경 정보 중 소리, 장소 정보를 매칭 하였고, 마지막으로 사람 간 관계 정보를 매칭 하였다.

표 1은 이벤트 정의 결과 생성된 총 30개의 이벤트, 4개의 primitive motions, 15개의 basic activities, 이벤트와 관련된 14개의 물체, 2개의 사람 간 관계, 2개의 소리 정보, 4개의 장소정보를 보여준다.

정의된 이벤트들을 토대로, 동영상의 키프레임 별 어노테이션을 수행하여 데이터베이스를 구축하였다. 어노테이션된 메타데이터는 앞서 정의된 각 항목에 카메라 정보, 시작과 끝 프레임 정보, 그리고 사람 ID 정보를 포함하는 13개의 항목으로 구성된다.

3.2 검색 시스템의 동작 과정

사용자로부터 질의가 입력되면, 시스템은 자연어 질의 분석을 통해 질의 유형을 정의한다. 질의 유형은 화행 분류를 통해 수행되어 지는데, 화행을 파악함으로서 질의에 대한 답변 매칭 시 사용자 의도에 적합한 답변을 제시할 수 있게 한다[11]. 질의 유형은 시간항목에 대한 질의 유형인 When, 피험자 항목에 대한 질의 유형인 Who, 이벤트 항목에 대한 질의 유형인 What, 사용 물체 항목에 대한 질의 유형인 Which things, 피험자의 위치 항목에 대한 질의 유형인 Where로 구성된다. 각 질의 유형은 표 1에서와 같이 발생가능한 모든 질의 값과 검색을 위한 sql문을 정의하고 있다. 질의 유형이 결정되면, 시스템은 사전 생성된 스크립트로부터 답변 쌍 매칭을 통해 해당 결과를 돌려주고, 이벤트 데이터베이스는 이를 이용하여 sql문을 수행한다. 예를 들어 “수정이가 출근 할 때의 장면” 에 대해서는 When의 질의 유형이 선택되고, When이 정의하고 있는 다음과 같은 sql문을 통해서 검색을 하게 된다.

```
SELECT Start frame, End frame
FROM EventDB
WHERE actor= ‘soojung’  AND event= ‘entry’ ;
```

시스템은 검색 결과 조건을 만족하는 실제 데이터 값과 같은 투플에 저장되어 있는 해당 이벤트의 시작 프레임 정보와 끝 프레임 정보를 이용하여 검색 결과에 해당하는 동영상을 제공한다.

3.3 스크립트 데이터베이스 설계

질의 결과는 해당 질의의 답변으로 발생가능한 질의 값에 대해 정의한다. 스크립트는 스크립트 언어 및 템플릿 변수를 통해 정의 되는데, 스크립트 언어는 대화의 범주를 나타내거나, 스크립트가 속한 주제나 사용자 입력의 패턴을 명시하고, 답변을 출력하는 항목들을 포함한다. 템플릿 변수는 개체의 범주에 대해 명시를 하거나 해당 변수가 가질 수 있는 값 등을 명시한다[12].

스크립트 언어 내에서 〈script〉는 대화의 범주를 나타내고, 〈topic〉은 스크립트가 속한 주제를 명시한다. 〈necessary_var〉는 스크립트 행동 수행에 있어서 필수적으로 필요한 정보 기술을 명시하며, 〈candidate_var〉는 행동 수행에 있어서 있으면 도움이 될 만한 변수에 대해 명시한다. 사용자의 입력은 〈query〉를 통해, 스크립트의 행동은 〈action〉을 통해 정의한다. 답변은 [answer_string]의 형태로 〈answer〉를 통해 출력되며, 〈function〉은 해당 도메인에 관련된 기능을 수행할 수 있도록 한다. 또 템플릿 변수 내에서 〈object〉는 개체의 범주에 대해 명시하고, 〈variable〉은 추출할 변수에 대해, 〈value〉는

표 1. 이벤트 정의 결과

	event	primitive motions	basic activities	object	person	sound	where
0 person	no event	–	–	–	–	–	–
1 person	entry	standing & moving	walking				into the office
	leaving	standing & moving	walking				out from the office
	calling	standing | sitting	holding | telling	phone		machine	inside the office
	cleaning – vacuuming – mopping	(standing | bowing) & moving	cleaning | holding	vacuum cleaner			inside the office
	eating	standing | sitting	holding	food			inside the office
	drinking	standing | sitting	holding	cup | can | bottle | glass			inside the office
	computer work – web searching – game – working – typing – sending e-mail – searching – retrieving – drawing	sitting	watching | using	monitor keyboard mouse			at own desk | inside the office
	printing	sitting | standing	holding | using	document		machine	in front of printer | inside the office
	reading	sitting	holding | watching	book | document | newspaper			at own desk | inside the office
	studying	sitting	watching|writing down	book | document | newspaper			at own desk | inside the office
	resting – napping – stretching	sitting	sitting relaxed lying down | stand				at own desk | inside the office
	copying	standing	holding	document | book		machine	inside the office
	scanning	sitting | standing	using | holding	document | book		machine	inside the office
	working	sitting	sitting | holding	document | book			inside the office
2~3 persons	conversation	standing | sitting	talking (to) looking (at)		each other	voice	inside the office
	meeting	sitting	talking (to) holding	document	each other one another	voice	inside the office
	seminar	sitting	watching speaking	screen	one another	voice	inside the office
	greeting	standing	talking (to) hand-up looking (at)		one another	voice voice voice	inside the office inside the office inside the office

해당 변수가 가질 수 있는 값을, 마지막으로 〈default〉는 스크립트의 〈answer〉 태그에서 템플릿 변수 값이 존재하지 않을 때 사용될 값에 대해 명시한다.

N. Ashish 등은 사용자의 질의를 도메인에 대한 클래스 분류와 클래스 별 구축되어진 온톨로지에 따른 질의 분류의 두 단계로 나누어 분류하였다[13]. 본 논문에서도 이를 토대로 질의 유형별로 도메인을 분류하고, 각각 사람, 이벤트, 물체 등의 범주를 나눈 후 이를 토대로 동의어 및 발생 가능한 질의와 답변 쌍을 생성하였다. 그림 2는 생성된 스크립트의 예를 보여준다.

```
<script>
<topic> where 질문형</topic>
<necessary_var>
   <variable>$where$who$</variable>
   <variable>$where$when$</variable>
   <variable>$where$event$</variable>
   <variable>$where$object$</variable>
</necessary_var>
<query>$where$when$ $where$event$ 할때 $where$who$는 어디 있었지?</query>
<action>
   <return value = true>
      <answer>$where$when$ $where$event$ 할때 $where$who$는
              Where(who[],when[],event[])에 있었습니다.</answer>
   </return>
...
<query>$where$when$ $where$object$는 어디에 있었지?</query>
<action>
   <return value = true>
      <answer>$where$when$ $where$object$는
              Where(when[],object[])에 있었습니다.</answer>
   </return>
</action>
</script>
```

그림 2. 생성된 스크립트의 예

4. 실험 결과

4.1 실험환경 설정

사무실 내에서의 일상생활에 대한 실제 영상 데이터 수집을 위해, 사무실 내에 4m×4m 영역을 목적영역으로 설정하고 8대의 카메라를 설치하였다. 실험에 사용되어진 카메라는 소니 네트워크 카메라(SNC-P5)로서, 영상은 320×240의 해상도에 30fps의 MPEG 동영상 포맷으로 저장하였다.

4.2 데이터 수집

동영상 데이터는 사무실 내의 일상생활에서 발생하는 모든 이벤트들에 대해 미리 정의해 둔 컨텍스트 정의에 기반 하여 3명의 피험자에 대한 시나리오를 설계한 후, 그에 따라 촬영 되었다. 시나리오는 오전 9시부터 6시까지의 업무시간을 기준으로 하루 일과를 가정하였다. 데이터 샘플은 Entry 30개/Leaving 30개/Calling 14개/Cleaning 74개/Conversation 4개/Eating 7개/Nap 4개/Meeting 82개/Printout 9개/Seminar 7개/Work 20개로 수집되었다. 그림 3은 실제 데이터 수집에 사용된 시나리오의 예를 보여준다.

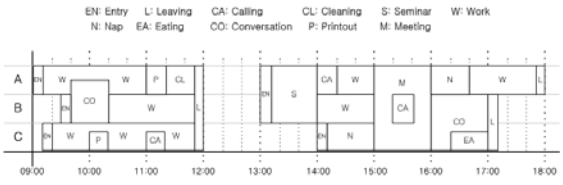

그림 3. 설계된 시나리오의 예

이렇게 수집된 동영상 데이터는 어노테이션 툴을 사용하여 어노테이션 한 후, 생성된 메타데이터를 데이터베이스에 저장하여 검색에 사용할 수 있게 하였다. 검색을 위한 데이터베이스는 Microsoft Access 2003을 이용하여 구축하였다.

4.3 시스템 동작 예

사용자의 질의에 대해 시스템은 자연어 질의 분석을 통해 사전 생성된 스크립트로부터 답변 쌍 매칭을 통해 동영상 검색의 결과를 사용자에게 제공한다. 예를 들어 동작 시나리오를 도식화한 그림 4에서와 같이 "누가 청소기로 청소했는지 보여줘"라는 질의에 대해 시스템은 키워드나 단어의 형태소 분석과 문형 분석을 통해 질의를 분석하고, 대화형 질의와 답변 쌍 매칭을 통해 "질의 유형 Who"에 대해 "event = 'vacuuming', object = 'vacuum cleaner'"을 조건문으로 하는 sql문을 수행한다. 그

후 시스템은 결과로 추출된 항목 "Soojung"과 조건을 만족하는 투플에서 시작 프레임과 끝 프레임 "Start frame: 1900, End frame: 2300"을 추출하여 동영상과 연동하게 된다. 그림 5는 해당 질의에 대해 실제 어플리케이션에서 검색된 결과를 보여준다.

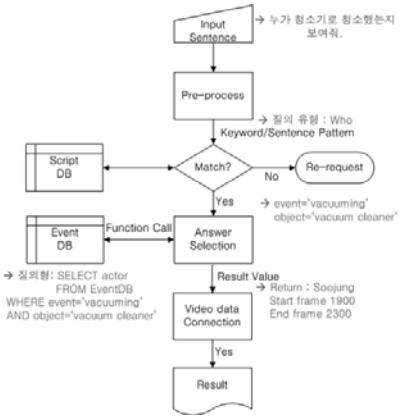

그림 4. 동작 시나리오

그림 5. 실제 어플리케이션에서의 검색 결과

4.4 사용성 평가

제안하는 방법의 유용성을 보이기 위해 20~29세의 남·녀 대학생과 대학원생 10명을 대상으로 성능평가와 프로세스평가를 수행하였다. 성능평가는 목적에 따라 사용자가 얼마나 빠른 시간에 검색을 수행하였는지 와 검색 성공률로 평가하는데, 본 논문에서는 주어진 문제에 대해 검색 결과를 얻기까지 걸리는 시간으로 기존의 메뉴 선택 기반 검색 시스템과 제안하는 자연어 기반 검색 시스템을 비교한다. 메뉴 선택 기반 검색의 경우에는 사람, 이벤트, 물체, primitive motions, basic activities의 도메인에 대해 검색 가능하며, 각 도메인의 개수를 15(s)개와 30(L)개 두 가지로 구분하여 각각의 성능 평가를 비교한다. 프로세스평가는 시스템 사용 후 느끼는 만족도 평가로, SUS(System Usability Scale)의 항목을 사용한다. 각 문항의 응답은 Likert 척도를 사용하여 강한 부정, 부정, 보통, 긍정, 강한 긍정의 1에서 5까지의 점수를 부여하도록 하였다. 홀수 문항은 평가 값에서 1을 빼고, 짝수 문항은 5에서 평가 값을 뺀 후 이를 모두 더하고, 여기에 2.5를 곱하여 0부터 100사이의 값을 갖는 점수로 환산한다.

표 2. SUS 항목

SUS	문항
SUS-1	나는 이 시스템을 자주 사용할 것 같다
SUS-2	시스템에 불필요하게 복잡한 부분이 있다
SUS-3	시스템이 사용하기 쉽다고 생각한다
SUS-4	이 시스템을 사용하기 위해서 전문가가 필요할 것 같다
SUS-5	이 시스템은 다양한 기능이 조직적으로 잘 결합되어 있다
SUS-6	이 시스템은 너무 불안정한 것 같다
SUS-7	대부분의 사람이 이 시스템의 사용방법을 빨리 익힐 것이다
SUS-8	이 시스템은 사용하기 귀찮은 부분이 있다
SUS-9	나는 이 시스템을 사용했다는데 자부심이 생긴다
SUS-10	이 시스템을 계속 사용하기 위해서는 많은 것을 배워야 할 것 같다

표 2는 SUS 항목의 구성을 나타내고, 그림 6은 제안하는 시스템과의 비교를 위해 제작된 메뉴 기반 검색 시스템의 어플리케이션을 보여준다.

그림 6. 메뉴 기반 검색 시스템

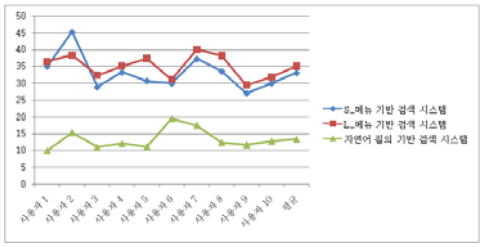

그림 7. 성능 평가 결과

그림 7은 메뉴 기반 검색 시스템과 제안하는 자연어 질의 기반 시스템의 성능평가 결과로서, 각각을 사용하여 "한샘이가 앉아서 전화기를 잡고 전화를 할 때", "수정이가 서서 청소를 하며 청소기를 잡고 있던 순간", "종욱이가 자리에 앉아서 키보드를 써서 컴퓨터를 할 때", "종욱이가 자리에 앉아서 음식을 들고 먹을 때", "한샘이가 앉아서 종이를 들고 미팅을 할 때"의 주어진 5개 질문에 대해 검색을 수행하는데 까지 걸린 시간의 평균을 나타낸다.

메뉴 기반 검색 시스템의 경우 초기 시스템에 대해 지식을 습득하는 시간을 제외하고는 검색 도메인의 부피가 증가함에 따라 검색 시간이 증가하였다. 이는 도메인 개수를 n, 메뉴 선택 한 개 당 드는 수행 시간을 T_n이라 했을 때, 메뉴 기반의 시간 복잡도가 5개의 메뉴 선택을 통해 최대 $5^* T_n$이

되므로, 자연어 기반의 검색 시스템의 복잡도가 C인 것에 비해 평균 수행 시간이 길어지기 때문이다. 제안하는 시스템의 경우에는 자연어 질의를 이용하여 사용자가 검색하고자 하는 문장을 직접 입력하기 때문에 시간 복잡도가 비교적 일정하게 유지되므로, 복잡한 도메인 환경에서 메뉴기반 시스템에 비해 유용성을 확인할 수 있다. 그림 8은 도메인 개수의 증가에 따른 시간 복잡도와의 관계를 나타낸다.

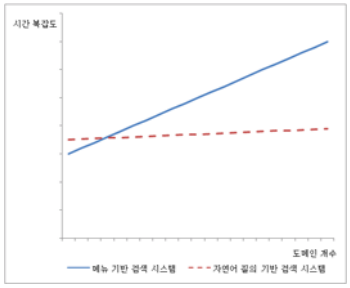

그림 8. 도메인 개수와 시간복잡도의 상관관계

그림 9는 두 시스템을 사용 후 느끼는 사용자의 만족도 여부에 대한 평가로, SUS의 10개 항목을 사용하여 평가한 결과를 100점 만점으로 환산하여 보여준다. 제안하는 시스템의 만족도는 평균 72.5점으로, 메뉴 기반의 검색 시스템(평균 61.25) 보다 우수한 것을 확인하였다.

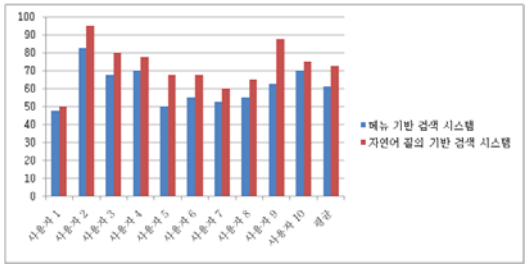

그림 9. 프로세스 평가 결과

5. 결론 및 향후 계획

본 논문에서는 이벤트 분석과 정의를 통해 도메인 지식을 구축하고, 동영상 정보에 대해 어노테이션한 메타데이터로 이벤트 데이터베이스를 구축하였다. 사용자 질의에 대한 전처리 과정과 분석 작업을 통해 스크립트 데이터베이스를 설계하였고, 구축된 시스템은 자연어 질의와 답변 쌍 매칭을 통해 동영상 검색의 결과를 사용자에게 제공하였다. 사용성 평가 결과, 제안하는 시스템은 도메인의 부피가 증가함에 따라 메뉴 기반 검색 시스템에 비해 우수한 성능과 만족도를 보였다.

향후에는 현재 생성되어진 스크립트를 클래스로 분류하고 각 클래스 별 온톨로지를 구축하여 스크립트를 분류, 확장할 예정이다. 또한 발생 가능한 모든 이벤트에 대해 동영상 데이터를 수집하여 시스템의 성능평가에 활용할 예정이다.

참고문헌

[1] S. Chang, W. Chen, H. Meng, H. Sundaram, and D. Zhong, "A fully automated content-based video search engine supporting spatiotemporal queries." *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 8, pp. 602-615, 1998.

[2] S.-J. Lim, J.-K. Min, H.-S. Park and S.-B. Cho, "Bayesian network based event recognition in multi-camera enrironment," *Proc. of The Korea Computer Congress*, vol. 34, pp. 248-251, 2007.

[3] H.-S. Park and S.-B. Cho, "Fuzzy rule-based summarization of event sequences in an indoor multi-camera environment," *Proc. of the KIISE, vol. 34*, pp. 288-292, 2007.

[4] V. Roth, "Content-based retrieval from digital video," *Image and Vision Computing*, 1999.

[5] T. Volkmer, S. Tahaghoghi, and J. Thom, "RMIT University Video Retrieval Experiments"

[6] B. Allen, "Semantic Web Techniques for Searching and Navigating Video Shots in BBC Rushes."

[7] J. Fogarty, S. Hudson, C. Atkeson, D. Avrahami, J. Forlizzi, S. Kiesler, J. Lee and J. Yang, "Predicting human interruptibility with sensors," *ACM Trans. on Computer-Human Interaction*, vol. 12, pp.119-146, 2005.

[8] D. Moore, I. Essa and M. Hayes, "Object spaces: Context management for human activity recognition," *Proc. of the 2nd Int. Conf. on Audio-Vision-based Person Authentication*, pp.1-6, 1998.

[9] N. Oliver, A. Garg and E. Horvitz, "Layered representations for learning and inferring office activity from multiple sensory channels," *Computer Vision and Image Understanding*, vol.96, pp.163-180, 2004.

[10] C. Wojek, K. Nickel and R. Stiefelhagen, "Activity recognition and room-level tracking in an office environment," *IEEE Int. Conf. on Multisensor Fusion and Integration for Intelligent Systems*, pp.25-30, 2006.

[11] S.-I. Lee and S.-B. Cho, "A conversational agent based on structured pattern matching," *Proc. Korea Information Science Society*, 2001.

[12] S.-S. Lim and S.-B. Cho, "Domain independent script language for constructing mixed-initiative conversational agent," *Proc. of The 33st KISS Fall Conference*, vol. 33, pp. 269-273, 2006.

[13] N. Ashish, C. Knoblock, and C. Shahabi, "Selectively materializing data in mediators by analyzing user queries," *Operative Information Systems*, pp. 256-266, 1999.