

실내 멀티카메라 환경에서의 퍼지 규칙 기반 이벤트 시퀀스 요약

박한샘⁰, 조성배
연세대학교 컴퓨터과학과

sammy@sclab.yonsei.ac.kr, sbcho@cs.yonsei.ac.kr

Fuzzy Rule-based Summarization of Event Sequences in an Indoor Multi-camera Environment

Han-Saem Park⁰ and Sung-Bae Cho
Department of Computer Science, Yonsei University

요약

최근 동영상 데이터의 폭발적인 증가와 함께 이의 요약을 위한 연구가 활발히 이루어지고 있다. 동영상 데이터 수집 과정에서 하나의 카메라가 아닌 멀티 카메라를 활용하는 경우도 늘고 있는데 이들 대부분은 실내에서 넓은 영역을 커버하거나 물체를 추적하기 위한 목적으로 멀티 카메라 시스템을 사용하였다. 본 논문에서는 하나의 이벤트를 여러 방향으로부터 입력하여 하나의 대상에 대한 다양한 시각과 정보에 초점을 맞추며, 이를 바탕으로 수집된 이벤트 시퀀스에 대한 요약 문제를 다룬다. 요약과정은 여러 개의 카메라 뷰 가운데 최적의 뷰를 선택하는 카메라 뷰 선택과정과, 그렇게 만들어진 하나의 전체 시퀀스를 요약하는 요약과정으로 나누어진다. 본 논문에서는 사용자 조사 및 분석을 통해 얻은 사용자 선호도 통계 정보로부터 카메라 뷰 선택을 위한 규칙을 획득하였고, 사람의 의사결정과정을 모방하고자 퍼지 규칙기반 시스템을 사용하여 이벤트를 평가한 후 그 점수에 근거한 요약을 수행하였다.

1. 서론

최근 디지털 카메라가 보급되고, 압축, 저장 기술 등 멀티미디어 처리기술과 대용량 데이터 전송 기술이 발달하며, 누구나 쉽게 동영상 데이터를 수집하고 이용할 수 있게 되었다[1]. 이는 공공건물에서의 CCTV에서부터 개인용 휴대 단말기에 이르기까지 다양한 방식으로 동영상 데이터를 수집 가능하게 하였는데, 이러한 동영상 데이터들은 텍스트문서나 음성, 그림파일 등의 다른 데이터 보다 훨씬 구체적이고 사실적이며 정확한 정보를 포함하고 있어 활용가치가 높다. 이처럼 동영상 데이터가 일반화되어 대량의 동영상 데이터 가운데 원하는 정보를 얻는 것이 중요해 졌으며[2], 이를 위해서 동영상을 분석하고 요약, 검색하기 위한 연구가 활발히 이루어지고 있다 [1-3].

요약의 대상이 되는 동영상은 영화, 뉴스, 스포츠 등 전문가에 의해 촬영된 동영상에서 미리 셋업 된 환경에 카메라를 설치하여 연구자가 직접 수집한 동영상까지 그 범위가 다양하다. 영화, 뉴스와 같이 장면(scene)과 샷(shot)의 구분이 명확하고 배경의 변화가 동적인 동영상의 요약 및 검색을 위한 절차는 크게 내용 분석(content analysis), 구조 분석(structure parsing), 요약(summarization) 및 검색(indexing & retrieval)의 네 단계로 나누어진다[2]. 내용 분석은 하위수준의 특징으로부터 의미 수준의 내용을 매핑하는 단계로서, 영상 처리 기법과 인식 기술 등을 이용하여 자동으로 수행되기도 하고[4], 사람에 의해 수동으로 수행되기도 한다[5]. 구조 분석은 내용 분석 결과를 바탕으로 동영상을 개별 장면 혹은 샷으로 나누는 단계이다. 요약은 이렇게 분석된 정보를 이용하여 동영상에서 중요한 내용을 간결하게 보여주는 단계이다.

직접 환경을 구축하고 카메라를 설치하여 수집된 데이터의 경우 주로 실내에서 여러 대의 카메라를 함께 사용

하는 멀티 카메라 환경을 대상으로 한다[6-8]. 이러한 데이터를 대상으로 하는 경우에는 집이나 사무실과 같은 실내를 대상으로 하기 때문에 배경 변화가 정적이고, 따라서 장면보다는 사용자의 위치나[6], 해당 도메인에서 발생하는 행동(activity)이나 이벤트(event) 단위로 동영상을 나누는 방법을 주로 사용한다[7]. 이러한 분석을 통해서 실내에서 발생하는 행동에 대한 연구를 수행하는데 도움을 줄 수 있으며, 또한 요약이나 검색과 같은 서비스를 통해 지난 일에 대한 사람의 기억을 돕기 위한 목적으로도 사용될 수 있다[6].

본 논문에서는 사무실 환경에 설치된 여러 대의 카메라를 이용하여 수집된 동영상을 대상으로 하며, 이를 이벤트 단위로 나누고, 사용자가 원하는 이벤트에 대해 해당 이벤트를 가장 잘 보여주는 카메라를 선택하고, 전체 이벤트 시퀀스 가운데 사용자의 요청에 맞추어 요약된 이벤트 시퀀스를 제공해 주는 시스템을 제안한다. 제안하는 시스템은 사용자로부터 이벤트나 사람에 대한 관심도(Degree of interest)를 입력받아 이벤트를 평가하는데 활용하며, 중요한 이벤트를 선별하기 위한 이벤트 평가 과정에는 사람의 의사결정 과정을 모방하기 위하여 퍼지 규칙 기반 추론이 사용되었다[9]. 또한 이벤트를 잘 보여주는 카메라 뷰 및 화면 전환을 선택하고, 요약하는 과정에 도움을 받기 위해서 사용자 조사, 도메인 지식, 관련연구 조사를 통해 구성한 도메인 온톨로지를 활용하였다.

2. 멀티카메라 시스템

기존의 멀티카메라 시스템은 주로 넓은 영역을 커버하거나, 이동중인 사람, 물체를 추적(tracking)하기 위해 사용되었다. J. Black 등은 실외환경에서의 움직이는 물체를 추적하고 검출하기 위한 멀티카메라 시스템을 개발하였고[10], 최근에는 실내 환경에서의 멀티카메라 시스

템에 대한 연구를 진행 중이다. 멀티카메라를 활용한 또 다른 주요 목적은 보안을 위해서이다. F. Porikli는 건물 내에 여러 대의 카메라를 설치하여 움직이는 물체를 추적하고 그 정보를 요약하였다[11]. Y. Sumi 등은 학회장에 설치되어진 센서와 멀티카메라를 이용해 피험자들의 행동을 분석한 뒤 간단한 요약을 수행하였고[7], G. C. Silva 등은 유비쿼터스 환경 내에서 가정 내 구성원들의 정보를 요약하거나 검색하기 위해 멀티카메라를 사용하였다[6].

하지만 멀티카메라 시스템은 넓은 영역을 커버하고 물체를 추적하기 위한 목적 이외에, 하나의 대상에 대한 다양한 시각을 가진 영상을 제공해 준다는 장점을 갖는다. 하나의 대상을 다양한 시각에서 바라봄으로써 주어진 상황에 대한 정확한 정보를, 그리고 필요에 따른 다양한 정보를 얻을 수 있는 것이다. C. Zhang 등은 IR 태그와 네 대의 카메라를 이용하여 사무실 환경에서 사람의 위치정보로부터 이벤트를 얻고, 사람, 이벤트, 위치에 대한 간단한 쿼리를 통해 원하는 이벤트를 검색할 수 있는 시스템을 구축하였다[8]. 또한 S. Lim 등은 사무실 환경에 다수의 카메라를 가운데를 향하도록 설치하고, 사무실 이벤트를 모델링하였다[12]. 이는 대상을 다양한 시각에서 바라볼 수 있는 장점에 초점을 맞춘 것이다. 본 논문에서 또한 멀티카메라 시스템의 이러한 가능성에 초점을 맞추어 연구를 진행하며, 사용자가 원하는 동영상 요약을 위해 사용자의 이벤트, 사람, 물체에 대한 사용자의 관심도를 입력받아 활용한다.

3. 멀티카메라 사무실 환경

3.1. 실험환경 설정

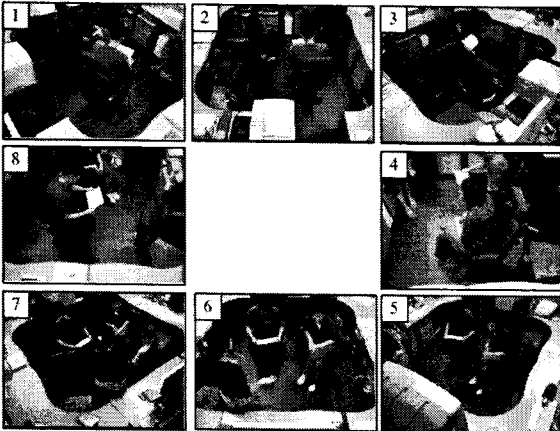


그림 1. 멀티카메라를 통해 수집한 영상

사무실 내에서의 동영상 데이터 수집을 위해, 연구실 내에 4m×4m 영역을 목적영역으로 설정하고 8대의 카메라는 공통적으로 가운데를 향하도록 설치하여 시스템이 동일한 이벤트에 대한 다양한 시점을 제공할 수 있도록 하였다. 그림 1은 8대의 카메라를 통해 실제로 수집한 영상을 보여준다. 실험에 사용되어진 카메라는 소니 네트워크 카메라 SNC-P5이며, 수집 영상은 320×240의 해상도, 15 fps의 MPEG 파일로 저장하였다.

3.2. 이벤트, 사람, 물체 정보의 어노테이션

본 논문은 사무실에서 수집한 동영상상을 이벤트의 시점으로 보고 그 상위에서 요약 서비스를 제공한다. 이벤트와 사람, 물체, 사람의 위치와 같은 기본 정보는 사람이 수동으로 어노테이션 작업을 통해 얻도록 하였으며, 이 작업은 아래와 같은 이벤트 정의를 바탕으로 하였다.

- Entry (A),
if stand (A, entrance-area) and face (A, in)
- Leaving (A),
if stand (A, entrance-area) and face (A, out)
- Calling (A),
if hold (A, phone) and speak (A)
- Cleaning (A),
if hold (A, vacuum cleaner) and stand (A, center-area)
- Eating (A),
if hold (A, food)
- Nap (A),
if rest (A, corner-area)
- Work (A),
if sit (A, corner-area) and
{use (A, computer) or hold (A, document)}
- Printing (A),
if exist (A, printer-area) and hold (A, printout)
- Conversation (A, B),
if {exist (A, x) and exist (B, y) and close (x, y)} and
{speak (A) or speak (B)}
- Meeting (A, B, C),
if {exist (A, x) and exist (B, y) and exist (C, z) and close (x, y, z)} and
{speak (A) or speak (B) or speak (C)} and
{hold (A, document) or hold (B, document) or hold (C, document)}
- Seminar (A, B),
if {stand (A, screen-area) and sit (B, center-area)} and speak (A)

사무실 환경 내에서 발생할 수 있는 일반적인 이벤트를 사전에 정의하였으며, 이벤트의 종류는 Entry, Leaving, Calling, Cleaning, Eating, Nap, Work, Printing, Conversation, Meeting, Seminar의 11 가지이다.

영상처리 및 패턴인식 기법을 통한 이벤트 인식, 물체 인식 등의 연구 또한 중요한 이슈로써, 기존의 비디오 요약 연구에서는 대부분 자동 혹은 반자동으로 물체나 이벤트의 어노테이션을 수행하였다[1, 4]. 본 논문에서는 사람의 어노테이션으로 대체하였으며, 이 부분은 향후 자동화 할 예정이다.

4. 제안하는 방법

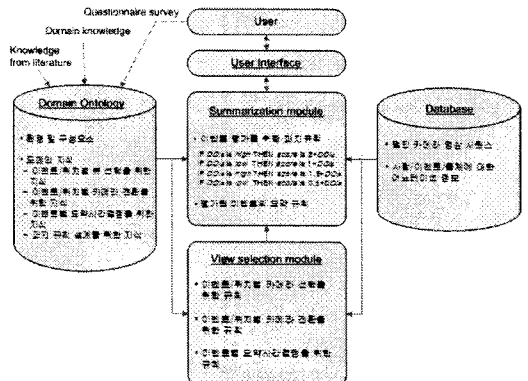


그림 2. 멀티 카메라 요약 시스템의 동작 구조

그림 2는 이벤트 시퀀스의 요약을 위해 본 논문에서 제안하는 시스템의 동작구조를 보여준다. 앞에서 설명했듯이 요약 과정은 카메라 뷰 선택(view selection)과 요약(summarization)으로 나뉜다. 한 시점의 이벤트를 가장 잘 보여주는 최적 뷰 선택을 위한 규칙은 사용자 조사, 도메인 지식, 관련연구 조사를 통해 얻은 도메인 온톨로지로 부터 만들어지며, 이렇게 얻어진 규칙에 따라 데이터베이스에서 데이터를 가져와 뷰 선택을 수행한다. 이벤트 시퀀스의 요약 또한 도메인 온톨로지로 부터 구성된 퍼지 규칙을 기반으로 하여 이벤트의 평가 및 평가 결과에 따른 요약을 수행하며, 이 때 사용자로부터 입력 받은 값을 함께 활용한다.

4.1. 사용자 입력

사용자가 원하는 뷰 선택 및 요약을 수행하기 위하여 사용자의 이벤트, 사람, 물체에 대한 관심도(Degree of interest)를 입력받아 활용하고자 하였다. 사용자 관심도는 인터페이스 디자인이나 사람과의 인터랙션 등을 위해 중요한 요소로 사용되고 있다[12]. 본 논문에서는 0 (관심 없음) ~ 3 (관심 많음) 사이의 정수를 입력받아 사용하였으며, 이 값은 뷰 선택과 요약을 위해 사용 될 수 있다.

본 논문에서는 샷(shot)을 기본 단위로 하여 전체 이벤트를 구분하였으며, 샷은 동일한 이벤트, 사람, 물체의 정보를 갖는 연속된 프레임(frame)으로 정의하였다. 사용자의 관심도는 샷별로 계산되는데, 하나의 샷 S_i 를 위한 관심도 값은 다음과 같이 정의된다.

$$DOI_{E_j, S_i}^2 = \sum_{frame f \in S_i} DOI_{E_j}(f) \quad (1)$$

$$DOI_{P_k, S_i}^2 = \sum_{frame f \in S_i} DOI_{P_k}(f) \quad (2)$$

$$DOI_{O_l, S_i}^2 = \sum_{frame f \in S_i} DOI_{O_l}(f) \quad (3)$$

식 (1)~(3)은 각각 하나의 샷(S_i)에서 이벤트(E_j), 사람(P_k), 물체(O_l)에 대한 관심도를 계산하는 식을 보여준다. 샷에 속한 모든 프레임에 대해 관심도 값을 더한 후 이벤트의 지속시간의 영향이 너무 크지 않도록 제곱근을 취한 값을 샷에 대한 최종 관심도 값으로 사용하였다.

4.2. 도메인 온톨로지

도메인 온톨로지는 환경, 구성요소에 대한 기술과 뷰 선택 및 요약을 위해 활용되는 도메인 지식 기술로 구성된다. 환경 및 구성요소는 앞 섹션에서 설명된 오피스 환경의 각종 구성요소 및 구성요소간의 관계를 기술한다. 도메인 지식은 어떤 이벤트를 가장 잘 보여주는 뷰의 선택을 위한 정보와 이벤트 내에서 사용자의 위치가 변화에 따른 뷰 전환을 위한 정보, 그리고 이벤트 평가를 위한 퍼지규칙의 설계를 위한 정보로, 사전 사용자 조사, 도메인 지식, 관련연구 조사 등을 통해 구성되었다.

도메인 지식을 구성하기 위해 실시한 사용자 조사는 20대 10명의 대학원생을 대상으로 수행되었으며, 과정을 예를 들어 설명하면 다음과 같다. 그림 2는 Seminar

이벤트를 8대의 카메라로 캡춰한 화면이며, 사용자 조사를 통해 이 중 Seminar를 가장 잘 설명하는 뷰는 6번 또는 7번이라는 정보를 얻었다. 모든 이벤트에 대해 이와 같은 조사 결과를 저장하였으며, 이 정보는 뷰 선택 및 이벤트 시퀀스의 요약에 활용된다. 예를 들면 Seminar이벤트를 요약할 때는 6번과 7번 뷰를 사용하게 된다. 위치변화가 큰 이벤트에 대해서는 뷰의 전환이 필요한데, 이동할 때마다 가장 가까운 뷰를 선택하여 따라가는 방법, 2가지 주요한 뷰를 잡아 적은수의 전환을 수행하는 방법, 가장 잘 보이는 하나의 뷰만을 사용하는 방법 중 이벤트를 가장 잘 보여주는 전환방법을 물어 그 정보를 저장하였다. 그 외에 하나의 이벤트에 대한 동영상 얼마동안 보여주는 것이 적절한지, 요약을 위한 이벤트 평가시 어떤 기준에 의해 평가를 해야 할지 등의 정보 또한 도메인 지식에 포함된다.

4.3. 뷰 선택, 이벤트 평가 및 요약

뷰 선택은 멀티 카메라 이벤트 시퀀스로부터 하나의 이벤트 시퀀스를 만들어 내는 과정으로, 이는 동일한 시점의 여러 이벤트 가운데 최적의 이벤트를 선택하는 과정과 하나의 이벤트를 가장 잘 보여주는 최적의 뷰를 선택하기 위한 과정을 포함한다. 기존의 논문에서는 여러 센서로부터 필요한 하나의 센서를 선택하기 위해 간단한 규칙을 주로 사용하였다. Y. Sumi 등은 이벤트 우선순위를 정하여 우선순위가 높은 이벤트에 해당하는 센서를 선택하였으며[7], G. C. Silva 등은 시간 주기를 정해놓고 규칙적으로 카메라를 전환하거나, 사람의 위치가 바뀌는 것을 추적하여 카메라를 전환하는 방법을 사용하였다[6]. 본 논문에서는 사용자가 입력하는 이벤트, 사람, 물체에 대한 관심도를 비롯하여 다양한 입력에 대해 최적의 카메라 뷰를 선택하기 위하여 사전에 수행된 사용자 조사를 거쳐 도메인 온톨로지의 도메인 지식을 구성하였으며, 이를 바탕으로 이벤트 별 뷰 선택을 위한 규칙을 생성하였다. 그림 3의 수도코드는 특정 시점에서 사용자가 가장 관심 있는 하나의 이벤트를 선택하고, 그 이벤트를 가장 잘 설명하는 최적의 뷰를 선택하는 뷰 선택 과정을 기술한다.

```

procedure
var
    N: the number of events in given time point
    M: the number of views for a given event
function
    SelectEvent(Ei): a function that select an event
        with the highest DOIE
        SelectView(Ei, Vi): a function that select a view
            for an event Ei based on rules in domain
            ontology
begin
    for i=1 to N
        SelectEvent(Ei)
    for j=1 to M
        SelectView(Ei, Vi)
end
    
```

그림 3. 최적 뷰 선택과정

뷰 선택을 마치고 생성된 하나의 이벤트 시퀀스를 요약하기 위해 요약에 앞서 각 이벤트를 평가하게 되는데, 이 과정에서 사람의 평가과정을 모방한 TSK 퍼지 시스템을 이용하였다. 도메인 온톨로지의 지식을 바탕으로 설계된 퍼지 규칙은 아래와 같다. 이벤트, 사람, 물체에 대한 사용자의 관심도를 입력으로 하여 2:2:1의 비율로

요약을 위한 최종 점수에 반영되도록 하였다.

- $IF DOI_E$ is high and DOI_P is high and DOI_O is high
THEN score = $2 * DOI_E + 2 * DOI_P + 1 * DOI_O$
- $IF DOI_E$ is high and DOI_P is high and DOI_O is low
THEN score = $2 * DOI_E + 2 * DOI_P + 0.25 * DOI_O$
- $IF DOI_E$ is high and DOI_P is low and DOI_O is high
THEN score = $2 * DOI_E + 0.5 * DOI_P + 1 * DOI_O$
- $IF DOI_E$ is high and DOI_P is low and DOI_O is low
THEN score = $2 * DOI_E + 0.5 * DOI_P + 0.25 * DOI_O$
- $IF DOI_E$ is low and DOI_P is high and DOI_O is high
THEN score = $0.5 * DOI_E + 2 * DOI_P + 0.25 * DOI_O$
- $IF DOI_E$ is low and DOI_P is high and DOI_O is low
THEN score = $0.5 * DOI_E + 2 * DOI_P + 0.25 * DOI_O$
- $IF DOI_E$ is low and DOI_P is low and DOI_O is high
THEN score = $0.5 * DOI_E + 0.5 * DOI_P + 1 * DOI_O$
- $IF DOI_E$ is low and DOI_P is low and DOI_O is low
THEN score = $0.5 * DOI_E + 0.5 * DOI_P + 0.25 * DOI_O$

퍼지 규칙을 통한 평가를 수행하기 전에 각 관심도 값을 퍼지화(fuzzification)하기 위해 15, 45를 기준으로 꺾이는 사다리꼴 모양의 퍼지 멤버쉽 함수를 사용하였다.

요약과정에서는 앞서 퍼지 규칙에 의해 평가된 점수를 기준으로, 점수가 높은 이벤트부터 차례대로 요약을 위해 선택된다. 이 때, 동일한 이벤트는 여러 번 선택되지 않으며 -동일한 이벤트가 중간에 발생한 이벤트로 인해 나눠지기도 함- 평가 결과 점수가 낮은 이벤트들은 요약대상에서 제외된다. 각 이벤트의 요약시간은 도메인 온톨로지의 사용자 조사결과로부터 얻어지며, 해당 샷의 가운데 부분을 요약을 위해 선택하였다.

5. 실험 결과

5.1. 시나리오 및 데이터 수집

데이터 수집은 앞서 정의된 사무실 내의 이벤트에 대해 3명의 사용자를 가정하고, 하루 일과에 대한 시나리오를 설계한 후, 그에 따라 촬영, 수집되었다. 시나리오는 오전 9시부터 오후 6시까지의 업무시간을 기준으로 설계하였다. 그림 4는 데이터 수집을 위해 설계된 시나리오를 보여준다. 여기에서 약자로 쓰여진 EN은 entry, PR은 printing, CONVERS는 conversation, CA는 calling, LE는 leaving을 각각 의미하며, 위쪽의 화살표는 시간의 흐름을, A, B, C는 이벤트에 참여하는 세 명의 사람을 의미한다.

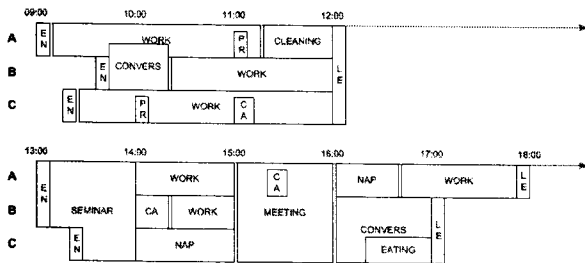


그림 4. 사무실 이벤트 시퀀스 수집을 위한 시나리오

5.2. 뷰 선택 결과

그림 5의 시나리오를 바탕으로 이벤트에 대한 관심도를 표 1과 같이 가정했을 때, 멀티 카메라 이벤트 시퀀스로부터 카메라 선택을 통해 얻은 하나의 이벤트 시퀀스는 표 3과 같이 정리된다. 물체에 대한 관심도는 모두 0으로의 입력을, 사람에 대한 관심도는 A-1, B-1, C-3을 가정하였다.

표 1. 사용자 입력(이벤트에 대한 관심도)

이벤트	DOI_E
Entry	2
Leaving	1
Calling	1
Cleaning	3
Eating	0
Nap	0
Work	1
Printing	3
Conversation	2
Meeting	3
Seminar	3

표 2에서 중간에 비어있는 시간은 아무 이벤트도 발생하지 않은 경우에 해당하며, C#은 카메라 번호를, F#은 프레임 번호를, 이벤트는 이벤트를, 사람은 시나리오 상에서 이벤트에 참여하는 사람을 각각 의미한다. 사용자 입력을 고려하면 자연스러운 결과이지만, 사람 C와 관련된 이벤트가 상대적으로 많이 선택되었으며, 사용자의 관심도가 높은 Printing, Cleaning, Seminar, Meeting 등의 이벤트에 해당하는 카메라는 모두 선택이 되었다.

표 2. 뷰 선택을 거쳐 선택된 이벤트 시퀀스

C#	F# (시작)	F# (끝)	물체	사람	이벤트
7	229	330		A	Entry
7	361	410		A	Work
5	411	530		C	Entry
5	571	730	Computer	C	Work
7	731	980		A, B	Conversation
2	981	1200	Printer	C	Printing
7	1201	1330		A, B	Conversation
5	1331	1810	Computer	C	Work
7	1811	2030		A	Printing
5	2031	2180	Phone	C	Calling
5	2201	2350	Computer	C	Work
2	2351	2760	Vacuum	A	Cleaning
5	2761	2890		C	Leaving
2	2891	2950	Vacuum	A	Cleaning
7	2996	3030		A	Leaving
7	3591	3630		A	Entry
8	3631	3660		B	Entry
6	3721	3960		A, B	Seminar
5	3961	4055		C	Entry
6	4061	4520		A, B, C	Seminar
6	4561	4800	Phone	B	Calling
5	4801	5420		C	Nap
6	5471	5760	Document	A, B, C	Meeting
6	5761	6030	Document	B, C	Meeting
6	6031	6330	Document	A, B, C	Meeting
6	6331	7210	Document	B, C	Meeting
5	7241	7340		C	Leaving
7	7346	7605		A	Work
7	7611	7660		A	Leaving

5.3. 요약 결과

요약을 위한 실험은 표 3의 이벤트 시퀀스를 바탕으로 수행되었다. 이 동영상은 29개의 샷을 포함하며 섹션 5.2와 동일한 사용자 입력을 가정하고 이벤트 평가를 수행하였다. 그림 5는 퍼지 추론을 거쳐 평가된 샷별 최종 점수의 변화를 보여준다. 가장 최종 점수가 높은 샷은 Seminar 이벤트에 대한 샷 #20과 Meeting 이벤트를 보여주는 샷 #23 ~ 샷 #26 인데, 이 샷들은 공통적으로 사용자의 관심도가 높은 이벤트를 보여주며, 사람 C를 포함하고 오랜 시간 지속되었다. Printing 이벤트에 대한 샷 #6 또한 높은 점수로 평가되었다. 샷 #6은 지속시간은 짧지만 사용자의 관심도가 높은 Printing 이벤트에 대한 것이며, 역시 사람 C를 포함한다.

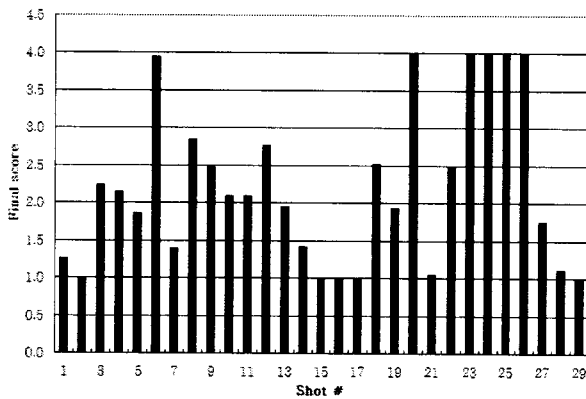


그림 5. 샷별 최종 점수

요약된 이벤트 시퀀스는 표 4에 정리된 것과 같다. S#는 샷 번호를 의미하며 C#에서는 이벤트에 따라 이상의 뷰가 선택되기도 하였다. 점수는 퍼지 추론을 거쳐 최종 평가된 샷의 점수를 보여준다. 카메라 전환 과정에서 뷰의 선택은 섹션 4.2에 기술된 것과 같이 도메인 온톨로지의 사용자 조사결과를 바탕으로 설계된 규칙에 의해 수행되었다.

표 4. 요약된 이벤트

S#	C#	F# (시작)	F# (끝)	사람	이벤트	점수
3	6⇒1	400	570	C	Entry	2.2
4	1	616	690	C	Work	2.1
6	1⇒6⇒1	950	1230	C	Printing	3.9
9	3⇒6⇒3	1810	2050	A	Printing	2.5
10	1	2065	2140	C	Calling	2.1
12	6⇒3⇒1⇒6	2395	2750	A	Cleaning	2.8
13	1⇒6	2750	2900	C	Leaving	2
20	2⇒3	4216	4365	A, B, C	Seminar	4
22	1	5081	5155	C	Nap	2.5
25	2	6441	6515	A, B, C	Meeting	4
26	1	6731	6805	B, C	Meeting	4
27	6	7241	7335	C	Leaving	1.8

6. 결론 및 향후 연구

본 논문에서는 사무실 환경에 설치된 멀티 카메라 시스템으로부터 동영상을 수집하여 이를 이벤트 시퀀스로 어노테이션 하고 요약하는 시스템을 제안한다. 이 과정에서 사용자에게 관심도 값을 입력받아 활용하였으며,

최적의 카메라 뷰 선택 및 요약 과정을 위해 도메인 온톨로지에서 얻어진 규칙을 활용하였다. 또한 이벤트 요약 전에 평가를 위해 퍼지 규칙 기반 시스템을 사용하였다. 뷰 선택과 요약 실험에서는 그 결과가 사용자 입력에 따라 자연스러운 결과를 보임을 확인하였다.

향후에는 실험 결과를 평가할 수 있는 사용자 평가 등의 방법이 필요하며, 사람이 직접 수행한 어노테이션 과정을 자동화 할 필요가 있다.

감사의 글

본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음. IITA-2006-(C1090-0603-0046)

참고문헌

- [1] X. Zhu, J. Fan, A. K. Elmagarmid, and X. Wu, "Hierarchical video content description and summarization using unified semantic and visual similarity," *MULTIMEDIA SYST.*, vol. 9, pp. 31-53, 2003.
- [2] N. Sebe, M. S. Lew, and A. W. M. Smeulders, "Editorial introduction: Video retrieval and summarization," *COMPUT VIS IMAGE UND.*, vol. 92, pp. 141-146, 2003.
- [3] Y. Luo, T.-D. Wu, and J.-N. Hwang, "Object-based analysis and interpretation of human motion in sports video sequences by dynamic Bayesian networks," *COMPUT VIS IMAGE UND.*, vol. 92, pp. 196-216, 2003.
- [4] B. L. Tseng, C.-Y. Lin, and J. R. Smith, "Using MPEG-7 and MPEG-21 for personalizing video," *IEEE MULTIMEDIA*, vol. 11, no. 1, pp. 42-53, 2004.
- [5] M. Petkovic and W. Jonker, "An overview of data models and query languages for content-based video retrieval," *Int. Conf. on Advances in Infrastructure for e-business, science, and education on the internet*, 2000.
- [6] G. C. de Silva, T. Yamasaki and K. Aizawa, "Evaluation of video summarization for a large number of cameras in ubiquitous home," *Proc. of the 13th ACM Int. Conf. on Multimedia*, pp. 820-828, 2005.
- [7] Y. Sumi, S. Ito, T. Matsuguchi, S. Fels and K. Mase, "Collaborative capturing and interpretation of interactions," *Pervasive 2004 Workshop on Memory and Sharing of Experiences*, pp. 1-7, 2004.
- [8] C. C. Zhang, S.-B. Cho, and S. Fels, "MyView: Personalized event retrieval and video compositing from multi-camera video images," *LNCS 4557*, 2007.
- [9] A. Dorado, J. Calic, and E. Izquierdo, "A rule-based video annotation system," *IEEE T CIRC SYST VID.*, vol. 14, no. 5, pp. 622-633, 2004.
- [10] J. Black and T. Ellis, "Multi-camera image tracking," *IMAGE VISION COMPUT.*, vol. 24, pp. 1256-1267, 2006.
- [11] F. M. Porikli and A. Divakaran, "Multi-camera calibration, object tracking, and query generation," *IEEE Int. Conf. on Multimedia and Expo*, vol. 1, pp. 653-656, 2003.
- [12] S. Lim, J. Min, H.-S. Park, and S.-B. Cho, "Bayesian network based event recognition in multi-camera environment," *Proc. of the Korea Computer Congress*, vol. 34, no. 1(C), pp. 248-251, 2007.
- [13] D. Kim and W. Woo, "The VR digital storytelling design for users," *Proc. of the KHCI*, vol. 2, pp. 530-535, 2006.
- [14] S. Lertworasirikul, S.-C. Fang, J. A. Joines, and H. L. W. Nuttle, "Fuzzy data envelopment analysis(DEA): A possibility approach," *FUZZY SET SYST.*, vol. 139, no. 2, pp. 3-29, 1998.