
광범위한 지역 감시시스템에서의 행동기반 비디오 요약

Activity-based video summarization in a wide-area surveillance system

HyeYoung Kwon. 권혜영. Youn-mi Lee. 이윤미. Kyoung-Mi Lee. 이경미.

덕성여자대학교 전산정보통신학과 지능형멀티미디어연구실

요약 본 논문에서는 광범위한 지역을 감시하기 위해 설치된 여러 대의 카메라로부터 획득된 비디오에 대해 행동을 기반으로 한 비디오 요약 시스템을 제안한다. 제안된 시스템은 시야가 겹치지 않은 다수의 CCTV 카메라를 통해서 촬영한 비디오들을 30분 단위로 나누어 비디오 데이터베이스를 구축하여 시간별, 카메라별 비디오 검색이 가능하다. 또한 비디오에서 키프레임을 추출하여 카메라별, 사람별, 행동별로 비디오를 요약할 수 있도록 하였다. 또한 임계치에 따라 키프레임 검색정도를 조절함으로써 비디오 요약정도를 조절할 수 있다. in, out, stay, left, right, forward, backward와 관련된 11가지 행동들을 추출하여 요약된 정보를 가지고 현재 사람의 행동이 어떤 영역에서 어떤 방향으로 움직이고 있는 지에 대한 정보를 보여줌으로써 더 자세히 행동추적을 할 수 있다. 또한 카메라 3대에 대한 전체적인 키프레임에 대한 행동별 통계를 통해서 감시지역의 행동기반 이벤트를 간단히 확인해 볼 수 있다.

핵심어: *surveillance system, summarization, wide-area, activity*

1. 서론

현대사회에서는 국가적, 사회적 안전뿐만 아니라, 개인과 재산에 대한 안전과 보호의 관심이 높아짐에 따라, 주요 시설물 및 관공서, 학교, 기업, 가정에 이르기까지 보안의 중요성과 적용범위가 넓혀져 가고 있다. 특히, CCTV나 PC카메라 등을 이용한 영상 감시시스템은 영상을 통해 능동적으로 감시 대상을 구분하고 추적할 수 있는 효과적인 감시수단으로 이용되고 있다. 그러나 지금까지 개발된 대부분의 영상감시시스템은 감시요원이 카메라로 들어오는 영상을 육안으로 직접 지켜봐야 하기 때문에 비용과 인력의 소모가 크므로, 특정 감시요원 없이 자동적으로 감시 시간동안의 상황을 요약하고 검색할 수 있는 시스템이 필요하다.

비디오에서의 검색 및 요약 시스템은 이미 활발히 연구되고 있는데, 최근 들어 감시시스템에서의 비디오 요약 및 검색 시스템에 대한 연구가 관심을 받고 있다. Silva et. al은 유비쿼터스 환경에 설치된 다수의 카메라로부터 획득된 비디오를 요약하는 시스템을 제안하였다[1]. 이 시스템은 바닥에 감지센서를 장착하여 사람의 움직임을 추적하였고, 키프레임을 추출하여 시간별, 위치별로 비디오를 검색 요약하였다. Yamazawa et. al은 OVISS라는 시스템을 구축하여 4개의 영역 안에 센서를 바닥에 설치하여 in, out, stay,

move에 대한 사람이 들어오고 나가는 행동들을 검출하여 시간대별로 보여주고 있다[2]. Dmitry O. Gorodnichy는 ACE(Annotated Critical Evidence) surveillance system을 만들어 현대의 카메라에서 밤에 촬영한 영상을 통해서 7가지 행동(in, out, stay, left, right, backward, foreword)을 추출하여 보여주고 있다[3].

본 논문에서는 광범위한 지역에 설치된 다수의 CCTV 카메라를 이용하여 시간 독립적 배경 제거와 움직임 검출을 통해 11가지 행동(in, out, stay, left, right, forward, backward, left_forward, left_backward, right_forward, right_backward)에 기반한 키프레임을 추출하고 비디오를 요약하는 시스템을 제안한다. 비디오 요약 시스템은 감시지역에서 촬영한 비디오들에 대해서 요약정도를 임계치에 따라 대략적 또는 세밀하게 검색할 수 있도록 하여 방대한 양의 감시지역 비디오를 보다 효율적으로 요약 검색할 수 있다[4]. 또한 감시지역 안에서 낮부터 저녁까지 촬영한 영상을 통해 사람들의 행동들이 어떤 방향으로 움직이고 있는지를 추출하여 좀 더 자세히 사람들의 행동추적을 할 수 있다.

2. 광범위한 지역 감시시스템

광범위한 지역 감시시스템은 다수의 사람 추적이 가능한 감시시스템을 대상으로 한다. 다수의 비접침 카메라는 서버와의 연결을 통하여, 넓은 지역에 걸쳐 설치된 다른 카메라와 추적된 사람에 대한 정보를 주고받을 수 있다. 사람은 사람의 위치, 색상, 움직이는 방향, 중심점, 인체를 구성하는 경계영역 등의 정보를 포함한 사람모델을 이용하여 추적 대상의 정보를 전달할 수 있다. 생성된 사람모델의 정보는 각 카메라들과 서버와의 정보전달을 위하여 서버의 사람리스트에 등록, 저장된다. 사람리스트에 저장된 사람모델의 정보는 매 프레임마다 사람의 움직임에 따라 사람모델을 갱신시켜 추적대상의 정보변화의 손실을 최소화시킴으로써 정확한 추적이 가능토록 한다[5].

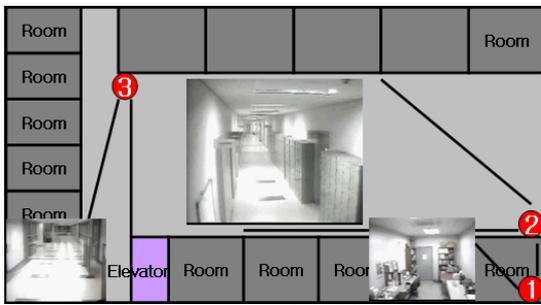


그림 1. CCTV 카메라 설치 영역

본 논문에서는 그림 1에서 보는 바와 같이 3대의 비접침 CCTV 카메라를 연결한 감시시스템에서 얻은 비디오에서 행동기반 키프레임을 검출하고 요약하여 검색할 수 있는 시스템을 제안한다.

3. 행동기반 키프레임 검출

키프레임은 비디오에서 중요한 이벤트나 내용을 담고 있는 프레임으로써, 좋은 비디오 요약 시스템을 구축하기 위해선 좋은 키프레임을 검출하는 것이 매우 중요하다. 본 논문에서는 키프레임을 이용하여 사람의 행동을 추적하기 위해서 행동기반 키프레임을 검출하였다.

사람의 행동을 CCTV로 촬영한 동영상으로 관찰한 결과, 카메라 3대를 통해서 현 영상에서 나타났다가 없어지는 사람들과 사람들이 복도에서 걸어들 때 복도의 길이가 한정되어 있어 대체로 상하좌우로만 움직이는 것을 보게 되었다. 그리하여 영상처리를 통해 얻을 수 있는 11가지 행동들로 정의하여 구축하였다.

3.1 키프레임 검출

본 논문에서는 행동의 움직임이 큰 프레임을 키프레임으로 정의하였고, 올바른 물체 검출을 위해 시간 독립적 배경 제거를 통한 물체를 검출하였다. 다수의 비디오카메라에서 획득된 비디오는 시간의 경과에 따라 빛과 조명에 의해 끊임없이 변화된다. 이러한 입력영상의 조명 잡음은 물체의 정보를 왜곡, 분실시키기 쉬우므로 올바르게 물체를 검출하기 매우 어렵게 된다. 본 논문에서는 시간에 따라 적응적 조명을 모델링하여 다음 프레임에서 나타날 조명과 음영을 미리 예측하고 제거하는 시간 독립적 배경 제거법을 이용한다[3].

조명 보정된 영상은 본래영상을 이용하여 배경을 분리한다. 배경영상을 분리하기 위해서 평균 배경영상을 계산한다. 새로운 프레임이 들어올 때마다, 평균 배경영상과 비교하여 차이가 보다 적은 픽셀을 배경으로 제거하고 평균 배경영상을 갱신시킨다.

키프레임은 다음 식1과 같이 배경이 제거된 전경 영상에서의 크기가 임계치(th1)보다 큰 프레임으로 간단하게 정의될 수 있다. (영상: p, 배경영상: b, x: 전경영상)

$$p_i - b_i = x_i (x_i > th1) \quad (식1)$$

그러나 이런 경우 움직임 없이 한 자리에 오래 서 있는 물체를 포함한 모든 프레임들이 키프레임으로 검출되어 키프레임의 개수가 너무 크며 중복되는 프레임을 포함하고 있어 좋은 키프레임들이 검출되었다고 보기 어렵다. 따라서 배경과의 차이뿐만 아니라, 사람의 움직임도 함께 고려한다면 움직임이 적은 중복 프레임을 제거함으로써 보다 적은 수의 키프레임을 검출할 수 있게 된다. 식2와 같이 현재 프레임은 이전 프레임과의 차이를 통해 움직임을 계산하고, 계산된 움직임 값이 임계치(th2)보다 큰 경우만 키프레임으로 검출 된다. (현재프레임 : f, 이전프레임: b)

$$f_i - b_i = x_i (x_i > th2) \quad (식2)$$

3.2 행동 검출

검출된 키프레임에 대해, 움직임이 있는 픽셀들을 연결 요소로 그룹화시켰다. 그룹화를 마치면, 주어진 영상에 움직임이 있는 연결 영역들의 리스트가 만들어진다. 연결 영역들은 영역 면적, 평균 색상, 최대, 최소의 bounding box의 꼭지점 위치, 물체의 중심점 등과 같은 정보를 포함한다.

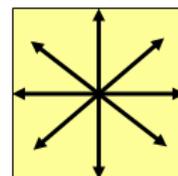


그림 2. 행동 방향

행동 검출은 전 키프레임과 현재 키프레임 상에서의 차이 통하여 구할 수 있다. 본 논문에서는 위에서 구한 연결 영역의 정보들 중 물체의 중심점을 이용하여 행동을 검출을 하였다. 사람의 행동을 파악하기 위해서 CCTV로 찍은 영상을 관찰한 결과, 그림2와 같이 8가지 방향의 움직임을 가지고 11가지의 행동을 판단할 수 있었다. 아래 표 1은 11가지 행동들을 추출할 수 있는 상태들을 설명하고 있으며, 행동기반 키프레임은 배경이 제거된 전경 영상에서의 전 후 키프레임의 차이를 통해서 표 1에서 설명했던 바와 같이 행동들을 추출하게 된다.

표 1. 11가지 행동들

행동	상태	설명	
1	IN	전 키프레임에서는 없었는데 현재 키프레임에 들어온 상태	
2	OUT	전 키프레임에서는 보였는데 현재 키프레임에서는 보이지 않는 상태	
3	MOVE	LEFT	전 키프레임과 현재 키프레임과의 차이에서 왼쪽으로 이동한 상태
4		RIGHT	전 키프레임과 현재 키프레임과의 차이를 이용하여 오른쪽으로 이동한 상태
5		FORWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 전방으로 이동한 상태
6		BACKWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 후방으로 이동한 상태
7		LEFT_FORWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 왼쪽으로 이동하면서 전방으로 움직인 상태
8		LEFT_BACKWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 왼쪽으로 이동하면서 후방으로 움직인 상태
9	RIGHT_FORWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 오른쪽으로 이동하면서 전방으로 움직인 상태	
10	RIGHT_BACKWARD	전 키프레임과 현재 키프레임과의 차이를 이용하여 오른쪽으로 이동하면서 후방으로 움직인 상태	
11	STAY	전 키프레임에서의 움직임과 현재 키프레임에서의 움직임이 유사한 상태	

행동기반 키프레임을 처음 추출할 때에 영역 면적값을 가지고 11가지 행동을 추출하려 하였으나, 그림 3의 카메라 2에서 보고 있는 것처럼 맨 처음 들어온 영상부터 사람이 크게 들어와 in보다는 forward나 backward로 인식되어 행동을 추출할 수 없었다. 그리하여 바로 앞의 키프레임과 현재 키프레임을 비교하여 현재 움직이는 사람의 중앙값의 x,y좌표를 가지고 추출한 결과, 그림 3처럼 추출하고자 했던 11가지 사람들의 행동들을 추출할 수 있었다. 그림 3는 적절한 임계치 값을 이용하여 추출한 영상들이다.

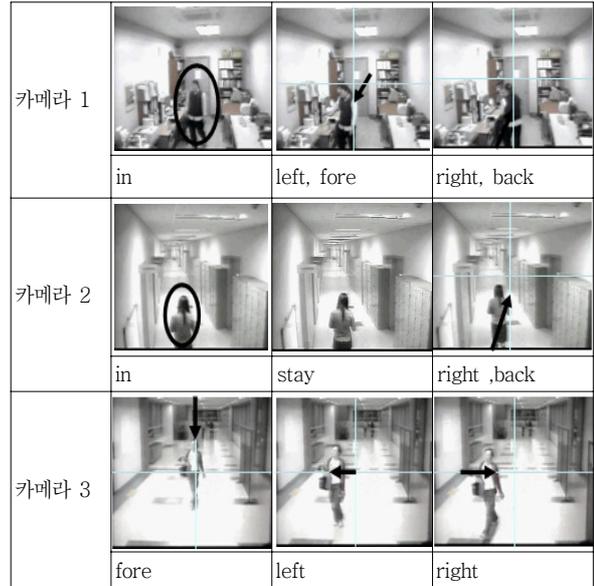


그림 3. 행동기반으로 추출된 영상들

4. 키프레임 기반 비디오 요약 시스템

본 논문에서 제안된 행동기반 비디오 요약 시스템은 pentium(R) D CPU 3.0GHz와 1.00GB RAM 메모리 사양의 windows XP상에서 UNIMO CCN-541 보안 카메라 세대에서 획득한 영상을 처리하기 위하여 JAVA(JMF)를 이용하였다.

4.1 데이터베이스

현재 본 시스템에서는 3대의 CCTV카메라에서 6시간동안 촬영한 비디오를 30분 단위로 나누어 비디오 데이터베이스에 날짜시간별, 카메라별, 사람별, 행동별로 저장하였다. cameraDB는 사용되고 있는 3대의 카메라에 대한 정보를 저장하고 있으며, keyframeDB는 비디오에서 추출된 행동기반 키프레임들의 전체적인 정보를 저장해 놓은 데이터베이스로 카메라 번호, 연월일, 시분초, 키프레임 번호, bgcount, prevcount, 검출된 물체 수, 영역, 라벨, minX, minY, maxX, maxY, center_x, center_y, position_x, position_y로 구성되어 있다. bgcount는 배경 영상이 제거된 전경 영상의 크기로, 3.1절의 전경 영상에 사용된 임계치(th1)보다 커야만 한다. prevcount는 이전 영상과 차이로 구해진 움직임의 크기로, 3.1절의 움직임에 사용된 임계치(th2)보다 커야만 한다. 또한 position_x는 left, right에 대한 정보만 저장하고 있으며, position_y는 in, out, stay, forward, backward에 대한 정보들을 저장하고 있다.

4.2 시스템 소개

본 논문에서 제안하는 비디오 검색 및 요약 시스템은 그림 3과 같이 전체적인 GUI를 보여주고 있다. 비디오 검색은 연월일과 시분을 입력하고 카메라 번호를 하나 또는 다수로 지정하여 검색되어지며, 비디오는 30분 단위로 나누어 검색되어진다. 또한 카메라가 설치된 감시 영역에 대한 정보도 확인할 수 있다. 그림 1과 같이 카메라들의 설치 위치와 시야를 볼 수 있다.

또한 제안하는 시스템은 검색된 감시 비디오에 대한 키프레임을 행동 순으로 나열함으로써 그림 5와 같은 비디오 요약을 제공하고 있다. 제안하는 시스템에서 비디오 요약은 카메라별 요약과 사람별 요약, 더불어 행동별로 요약된 결과 화면을 볼 수 있다. 하나 또는 다수의 카메라가 선택된 경우, 카메라별 요약은 각각의 카메라에서 촬영한 비디오의 키프레임 요약 결과를 차례대로 보여준다. 예를 들어, 카메라 1,2,3 모두 선택된 경우 카메라 1의 비디오 요약, 카메라 2의 비디오 요약, 카메라 3의 비디오 요약 순으로 결과가 보인다. 또한 행동별 검색은 11가지(in, out, stay, left, right, forward, backward, left_forward, left_backward, right_forward, right_backward)행동에 대한 키프레임 요약 결과를 보여준다. in, out, stay, left, right, forward, backward를 하나만 선택했을 경우에는 카메라와 상관없이 7가지 행동에 대한 내용들이 각각 보여지며, left를 선택하고 forward나 backward를 선택하면, left된 행동이면서 forward나 backward인 행동들을 보여주고, right를 선택하고 forward나 backward를 선택하면, right된 행동과 함께 forward나 backward로 행동한 요약 영상들을 2개씩 선택하여 볼 수 있다.

다수의 카메라가 연결된 감시시스템에서 촬영된 비디오에 대한 요약은 여전히 많은 분량을 차지하고 있다. 따라서 본 시스템에서는 사용자가 시스템에서 제공해 주고 있는 요약 내용을 자율적으로 조절할 수 있도록 하였다. 그림 4 하단의 스크롤바를 조정함으로써 bgcount에 대한 임계치(th1)와 prevcount에 대한 임계치(th2)를 정하게 된다.

또한 30분 비디오의 내용을 한 눈에 보기 위하여 본 논문에서는 카메라에 나타난 사람들에 대한 전체적인 통계 정보를 제공하고 있다. 그림 6은 3대의 카메라의 전체적인 통계 정보로 카메라 1은 다른 2대의 카메라보다 행동량이 적으며, 카메라 2는 3대의 카메라 중 가장 많은 행동량을 보여주고 있다. 카메라 3은 20분 이 후로 활동량이 가장 많은 것을 볼 수 있다. 이렇게 3대의 카메라에 대한 30분 동안, 행동이 검출된 키프레임들을 보여주고 있다.

5. 결 론

본 논문에서 제안한 광범위한 지역 감시시스템에서의 행동기반 비디오 요약은 여러 대의 카메라를 동시에 감시해야 하는 어려움을 해결하기 위해 키프레임 요약 및 사람별 키프레임 검색을 통해 효과적으로 비디오를 요약할 수 있었다. 또한 임계치를 이용한 카메라별, 시간별, 행동별 키프레임 검색을 통해 시스템이 제공해 주고 있는 요약 내용을 자율적으로 요약할 수 있도록 하여 좀 더 검색의 효율을 높였다.

본 연구에 대한 향후 연구 과제는 현재 영상은 1인을 대상으로 하여 촬영한 영상을 가지고 11가지 행동들을 추출하고 있다. 앞으로는 다수의 사람들을 대상으로 촬영한 영상을 가지고 11가지 영상을 추출하여 더욱 강건한 시스템을 만들고자 한다.

참고문헌

- [1] G.C. de Silva, T. Yamasaki and K. Aizawa, "Evaluation of video summarization for a large number of camera in ubiquitous home", In proc. ACM multimedia, 820-828, 2005.
- [2] N. Babaguchi, Y. Fujimoto, K. Yamazawa and N. Yokoya, "A system for visualization and summarization of omnidirectional surveillance video," proceedings of 8th international workshop on multimedia information system 2002 18-27.
- [3] O. Gorodnichy, "ACE surveillance : the next generation surveillance for long-term monitoring and activity summarization", The First International Workshop on Video Processing for Security. Quebec City, Quebec, Canada. June 7-9, 2006.
- [4] 권혜영, 이경미, "광범위한 지역 감시시스템에서의 물체 기반 비디오 요약", 한국정보과학회 가을 학술발표논문집(B): pp.544~548, 2006,
- [5] 이윤미, 이경미, "모델기반 다중 사람추적과 다수의 비접촉 카메라를 결합한 감시시스템", 정보과학회논문지(B):컴퓨팅의 실제, 제12권 제4호 pp.241~253, 2006,

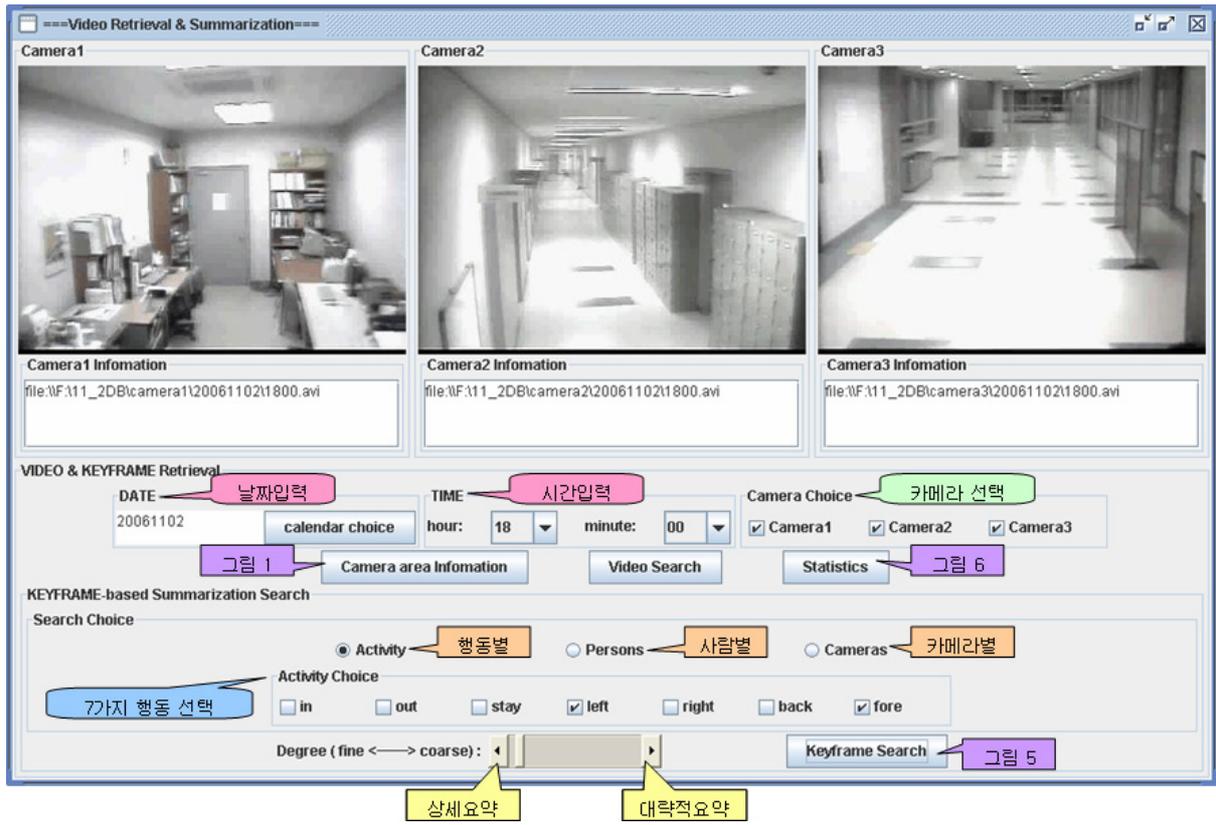


그림 4. 제안된 비디오 요약 시스템의 전체적인 GUI

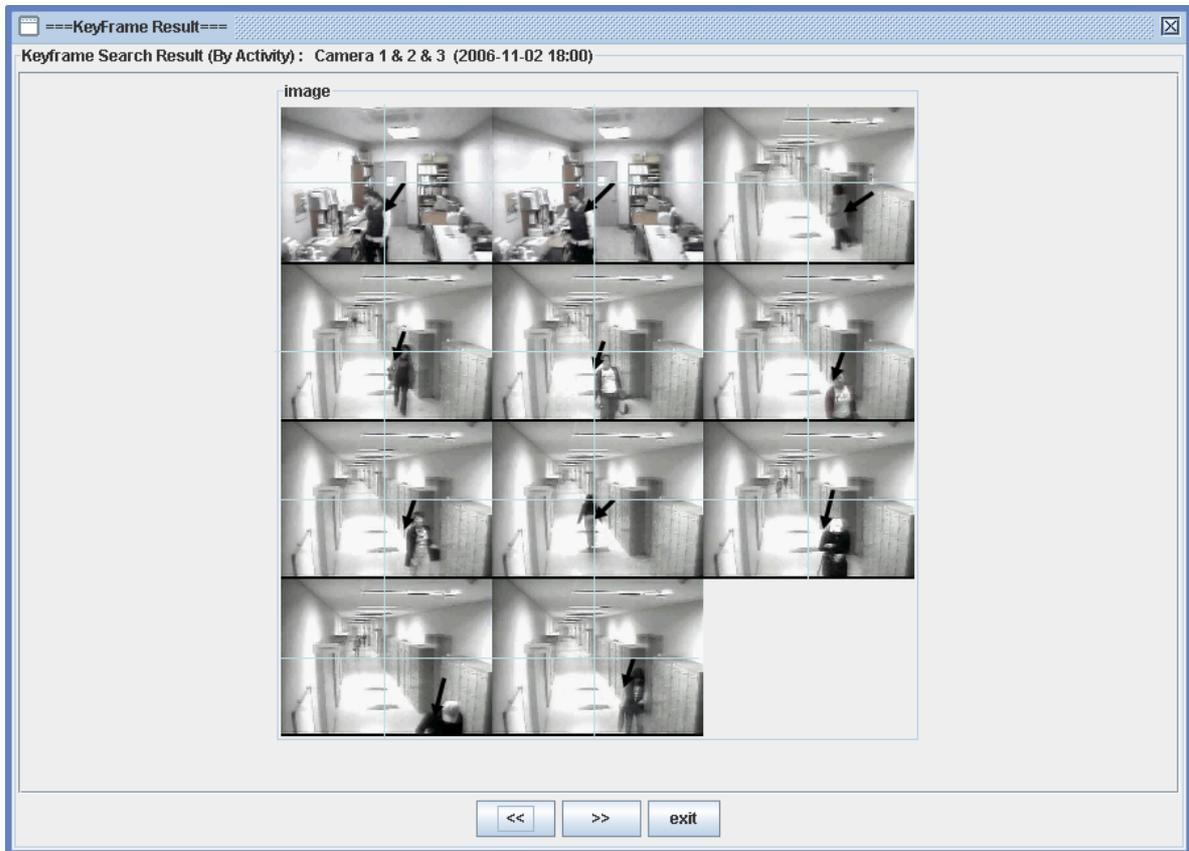


그림 5. 키프레임 검출을 이용한 행동별 검색에 따른 비디오 요약

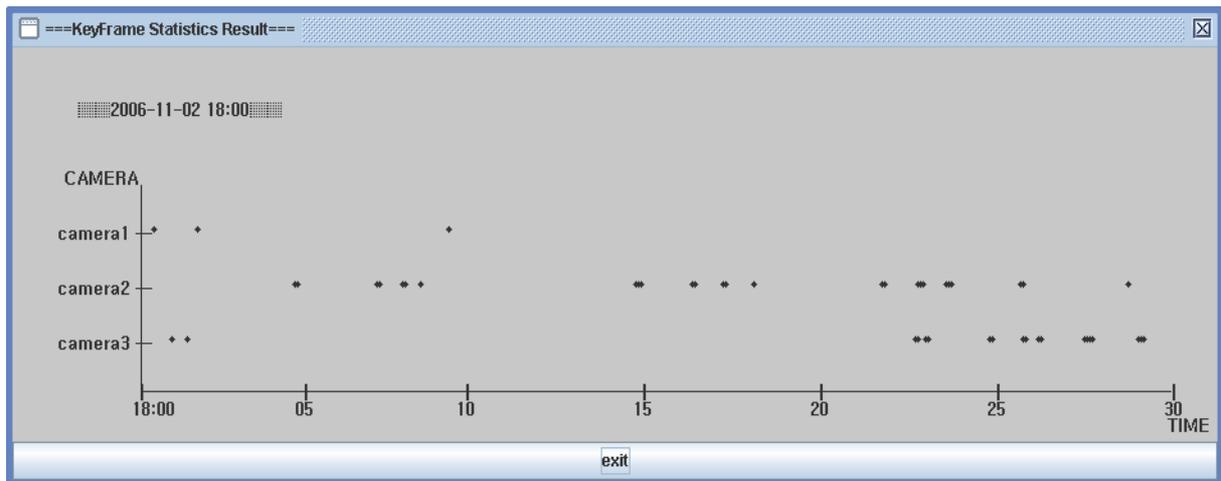


그림 6. 키프레임을 이용한 카메라 3대에 대한 비디오 통계