비디오 셧으로부터 영역, 모션 및 퍼지 이론을 이용한 계층적 대표 프레임 선택

(Hierarchical Keyframe Selection from Video Shots using Region, Motion and Fuzzy Set Theory)

강 행 봉 †

(Hang-Bong Kang)

요 약 내용 기반의 비디오 인덱싱 및 검색을 위해서는 비디오 데이타를 셧(shot)으로 분할하고, 또 각 셧을 나타내는 대표 프레임을 선택하는 것이 필요하다. 하지만, 대표 프레임을 선택하는 것은 주관적이어서 일관되게 자동적으로 대표 프레임을 선택하는 것은 쉬운 문제가 아니다.

본 논문에서는 각 프레임에서의 영역을 바탕으로한 컨텐트 정보 및 시간 축 상의 변화를 이용하여 계층적으로 대표 프레임을 선택하는 방법을 제안한다. 먼저, 비디오 셧에서 카메라 모션을 검출하여 이에 따라 비디오 셧을 분류한다. 다음, 분류된 비디오 셧에 컨텐트의 중요도를 계산하기 위한 퍼지 규칙을 적용하여 대표 프레임을 선택한다. 끝으로, 선택되는 대표 프레임의 수는 브라우징 상세도(detailness)에 따라 계층적으로 선택되게끔 한다.

Abstract For content-based video indexing and retrieval, it is necessary to segment video data into video shots and then select key frames or representative frames for each shot. However, it is very difficult to select key frames automatically because the task of selecting meaningful frames is quite subjective. In this paper, we propose a new approach in selecting key frames based on visual contents such as region information and their temporal variations in the shot. First of all, we classify video shots into panning shots, zooming shots, tilting shots or no camera motion shots by detecting camera motion information in video shots. Then, in each category, we apply appropriate fuzzy rules to select key frames based on meaningful content in frame. Finally, we control the number of key frames in the selection process by adjusting the degree of detail in representing video shots.

1. 서 론

인터넷상의 도처에 널려있는 다양한 형태의 멀티미디어 데이타들을 사용자가 원하는 데로 효과적으로 검색하는 것은 정보화 사회에 있어 중요한 일임에 틀림없다. 이러한 멀티미디어 데이타 중에서 비디오 데이타는 대용량이고 비구조화(unstructured) 되어 있어서 정보 검색에 어려움이 많이 있다. 효율적인 정보 검색을 위해서는, 비디오 데이타의 내용(content)을 기반으로 인덱싱(indexing)하고 검색하는 방법이 필요하다.

내용 기반의 비디오 인덱싱 및 검색을 위해서는 대용량의 비디오 데이타를 셧(shot)으로 분할하고, 또 각 셧에 대해 대표 프레임(key frame or representative frame)을 선택하는 것이 필요하다[1,2]. 대표 프레임이란 그 셧을 대표하는 프레임을 뜻한다. 하지만, 대표 프레임을 선택하는 것은 주관적이어서 일관되게 자동적으로 대표 프레임을 선택하는 것은 쉬운 문제가 아니다. 이러한 문제를 풀기 위한 한 방법은 셧에 존재하는 컨텐트(content)의 공간적 및 시간적인 변화를 고려하여 대표 프레임을 선택하는 것이다.

비디오 셧에 존재하는 컨텐트는 공간적으로는 오브젝트와 이에 관련된 칼라, 밝기 값, 크기, 모양, 텍스츄어 및 모션등의 중요한 특징으로 표현할 수 있다. 여기서 오브젝트는 한 개 또는 여러 개의 영역으로 구성되어

· 본 논문은 1998년도 정보통신부 대학 기초 연구 지원사업에 의한 것임.

† 종신회원 : 가톨릭대학교 컴퓨터전자공학부 교수

hbkang@www.cuk.ac.kr

논문접수 : 1999년 7월 28일

심사완료 : 2000년 2월 17일

있는 의미적인 단위이고, 또 영역 정보는 영상을 분할(segmentation)함으로서 쉽게 구할 수 있기 때문에, 오브젝트보다는 영역에 관련된 정보를 이용하는 것이 더 편리하다. 또, 이러한 영역에 관한 여러 가지 속성들의 시간 축 상의 변화에 따라 셧이 나타내고 있는 컨텐트가 변화하는 것이다. 따라서, 이러한 공간적 및 시간적인 특성을 바탕으로 프레임의 중요도를 계산할 수 있다면, 대표 프레임을 일관되게 선택할 수 있다.

대표 프레임의 선택에 있어서 또 하나의 중요한 이슈는 셧을 대표할 수 있는 프레임의 개수를 조정하거나 원하는 개수의 프레임을 효과적으로 선택하는 것이다. 이것은 다양한 비디오 브라우징을 위해서도 필요하고, 또 시스템이 갖고 있는 저장 장치의 한계를 맞추기 위해서도 필요하다. 따라서, 대표 프레임을 선택할 때는 프레임의 개수를 조정할 수 있어야 하고, 또 정해진 개수의 프레임을 효과적으로 선택할 수 있어야 한다.

본 논문에서는 비디오 셧을 카메라 모션에 따라 구분하여 각 셧에 컨텐트의 공간적 및 시간적인 중요도를 계산하여 이를 바탕으로 계층적으로 대표 프레임을 선택하는 방법을 제안한다. 2 장에서는 대표 프레임 선택에 관한 기존 연구에 대해 기술하고, 3 장에서는 카메라 움직임 및 퍼지 이론에 근거한 컨텐트 표현에 대해 설명하며, 4 장에서는 제안된 대표 프레임 선택 방법에 대하여 기술한다. 5 장에서는 제안된 알고리즘의 실험 결과를 보여 준다.

2. 관련 연구

수 년 전부터 비디오 셧을 대표할 수 있는 의미있는 프레임을 선택하기 위한 다양한 연구가 진행되어 왔다. 간단하게는 각 셧의 가운데 존재하는 프레임을 선택하거나, 또는 첫 번째 프레임과 마지막 프레임을 대표 프레임으로 선택하였다[3]. 하지만, 효율적으로 대표 프레임을 선택하기 위해서는 비디오 셧에 대해 모션 정보및 특징들의 차이를 계산하여 프레임을 선택하는 방법과 미리 정해진 개수만큼의 대표 프레임을 셧에서 일관되게 찾아내는 방법에 관한 연구가 진행되어 왔다.

Wolf[4]는 대표 프레임을 선택하기 위해서 모션 정보를 이용하였다. 셧에 존재하는 모션 정보를 계산하여 지역적 최소값(local minima)을 갖는 프레임을 택하여서 이를 대표 프레임으로 선정하였다. 왜냐하면, 모션이 일시 정지된 프레임이 일반적으로 보는 사람에게 강조를 뜻하는 프레임이기 때문이다. Zhang et al.[5]은 대표 프레임을 선택하기 위해 칼라 정보와 모션 정보를 이용하였다. 첫 번째 프레임을 대표 프레임으로 선택하고 연속된 프레임에서 기준 프레임과의 특징의 차이를 칼라 히스토그램을 이용하여 계산하였다. 이러한 차이가 임계값보다 크면, 이 프레임을 새로운 대표 프레임으로 선택하고, 이를 기준 프레임으로하여 같은 작업을 반복하였다. 또, 셧에 존재하는 모션에 따라 대표 프레임 선택의 기준이 되는 프레임을 다르게 정하였다.

미리 정해진 개수의 대표 프레임을 선택하는 방법으로서, Lagendijk et al.[6]은 셧이 지속되는 시간과 모션에 근거하여 일련의 대표 프레임을 할당하는 방법을 제안하였다. 이 대표 프레임들은 비디오 셧에 최적으로 분포되게 고안되어졌다. Sun et al.[7]은 미리 정해진 수의 대표 프레임을 선택하기 위해 적응적 클러스터링 알고리즘(adaptive clustering algorithm)을 사용하였다. 셧 경계를 검출하지 않고, 프레임 간의 차이를 콘텐트의 변화로 사용했으며, 이에 근거하여 비디오 셧을 두 개로 클러스터링 하였다. 한 클러스터는 컨텐트 변화가 적은 그룹으로서 대표 프레임 선택에서 제외되는 그룹이고, 다른 클러스터는 컨텐트 변화가 심한 것으로서 대표 프레임으로 선택되는 그룹이다. 중복되는 프레임을 제외되는 그룹에서 반복적으로 제거함으로서 원하는 개수의 대표 프레임을 추출하는 방법을 제안하였다.

기존의 대표 프레임 선택은 모션이나 칼라의 변화에 의한 것으로서, 셧에서 컨텐트의 공간적인 중요성이나 시간 축 상의 변화량에 의한 중요도를 고려하지 않았다. 더욱이, 정해진 개수의 프레임 선택뿐만 아니라, 컨텐트의 중요도 및 변화량이나, 저장 장치의 한계 또는 사용자의 의도에 따라 대표 프레임의 개수를 정하기가 어려운 단점이 있다.

3. 대표 프레임 선택 방법

본 논문에서는 각 프레임에서의 영역을 바탕으로한 컨텐트 정보 및 시간 축 상의 변화를 이용하여 계층적으로 대표 프레임을 선택하는 방법을 제안한다. 먼저, 비디오 셧에서 카메라 모션을 검출하여 이에 따라 비디오 셧을 분류한다. 카메라 모션 정보는 카메라 감독의 의도가 표현되어 있는 중요한 정보이므로, 이 카메라 모션 정보에 의해 분류된 비디오 셧으로부터 컨텐트의 중요도 및 변화량을 퍼지 규칙을 적용하여 계산하고, 이를 바탕으로 대표 프레임을 선택한다. 끝으로, 선택되는 대표 프레임의 수는 브라우징 상세도(detaliness)에 따라 다양하게 선택되게끔 한다.

3.1 카메라 모션 검출

비디오 셧은 일반적으로 카메라 모션에 의해 구별되므로, 비디오 셧을 특징 지우기 위해서는 카메라 모션을

효과적으로 검출하는 것이 필요하다. 비디오 데이타에서 모션 정보는 카메라 모션과 오브젝트 모션의 두 종류로 분류할 수 있다. 카메라 모션은 일반적으로 프레임 전체에 존재하는 움직임이고 오브젝트 모션은 프레임의 특정 영역에 국한되어서 발견되는 것이다. 따라서, 비디오 셧의 특징을 찾아내기 위해서는 카메라 모션을 이용하는 것이 바람직하다. 본 논문에서는 "팬"(panning), "줌"(zooming) 및 "틸트"(tilting) 등의 카메라 모션을 검출하기 위해서 Lucas-Kanade의 경사도를 이용한 옵티칼 플로우 계산 방식을 사용한다[8].

3.1.1 프레임에서의 모션 검출

비디오 셧에 존재하는 카메라 모션을 계산하기 위해 먼저 각각의 프레임에서 모션을 계산하고, 이를 바탕으로 셧 전체에 존재하는 모션을 검출한다. 프레임에서 모션을 구하기 위해 우리는 영역으로 분할 된 프레임 데이타를 사용한다. 왜냐하면, 각각의 프레임은 중요한 영역들로 특징 지울 수 있고, 또 이런 영역들에는 보통 같은 방향의 옵티칼 플로우가 존재하고 있으므로, 영역 안에 존재하는 모션을 계산하면 프레임 전체 영역을 계산하는 것 보다 간단하고 정확하게 움직임을 계산할 수 있다. 프레임을 영역으로 분할하기 위해서 모폴로지 필터(morphology filter)를 사용하여 영상을 단순화한다[9]. 단순화된 영상으로부터 편평한 영역(flat region)을 구하고, 이런 편평한 영역으로부터 watershed 알고리즘을 이용하여 영역의 경계를 검출한다[10]. 각각의 프레임에서 분할된 영역 중에서 5개의 커다란 영역을 주 영역(dominant region)으로 정하고, 각각의 주 영역에서 대표 모션을 계산한다. 대표 모션을 구하기 위해 옵티칼 플로우 벡터를 8개의 방향으로 양자화하여, 각 영역에서 8개 방향에 대한 옵티칼 플로우 벡터의 개수를 계산한다. 만약 방향θ를 갖는 옵티칼 플로우 벡터(m(x,y))의 개수(N_k)가 전체 옵티칼 플로우 벡터의 개수의 합(N_{total})의 반 이상이 될 경우, 이 방향(θ)을 영역의 대표 모션 벡터의 방향(Motion Phase)으로 간주한다. 또, 모션 벡터의 크기(Motion Intensity)는 대표 모션 벡터와 같은 방향의 옵티칼 플로우 벡터의 크기를 평균하여 계산한다. 즉,

$$\textit{Motion Phase} = \begin{cases} \theta & \text{, if } N_k > N_{total}/2 \\ don't\ \ care & \text{, otherwise} \end{cases} \quad (1)$$

$$\textit{Motion Intensity} = \sum_{i=1}^{N_k} m(x,y)/N_k \ , \quad \text{where motion phase of } m(x,y) = \theta \quad (2)$$

이렇게 하여 프레임의 각각의 주 영역에 존재하는 대표 모션 벡터의 방향과 모션 크기를 계산한다.

3.1.2 비디오 셧에서의 모션 검출

각 영역의 대표 모션 벡터의 방향과 크기를 이용하여 프레임 전체에서의 모션을 결정한다. 먼저, 프레임의 각 영역에서 각 방향에 대한 대표 모션의 개수를 구하고, 특정 방향의 모션 개수가 다른 방향의 모션 개수보다 적어도 두 배 이상 클 때, 그 프레임의 모션 방향으로 정한다. 그림 1은 카메라 모션이 "팬"인 경우를 보여주고 있고 방향은 180도이다.

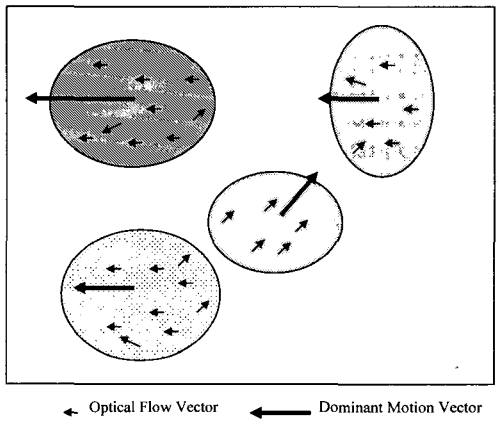


그림 1 영역에서의 모션 벡터

만일 "팬" 혹은 "틸트"등의 모션 방향을 정할 수 없을 경우에는 프레임이 "줌 인/아웃"(zoom in/out) 인 경우인가를 테스트한다. 이렇게 하기 위해서는 먼저 초점(Focus of Expansion or Focus of Contraction)을 계산한다. 주 영역에 존재하는 각 화소에서 그 화소가 갖고 있는 옵티칼플로우 벡터의 방향으로 선을 연장한다. 각 화소에서는 선이 지나가는 횟수를 계산하고, 가장 많은 선이 지나가는 화소가 가장 많은 옵티칼플로우 벡터가 향하는 방향이므로 초점(FOE or FOC)이 될 가능성이 높다(그림 2 참조). 만약 임계값 이상의 횟수를 기록하는 화소가 존재하지 않는 경우에는 "줌 인/아웃" 모션이 존재하지 않는다. 이 초점은 주 영역 안에 존재할 수도 있고, 밖에 존재할 수도 있다. 계산된 초점이 정확한지를 체크하기 위해 이 초점을 기준 점으로 프레임을 8개의 방향으로 분할한다. 이때, 각 영역들은 부분 영역으로 나누어 질 수도 있다(그림 3 참조). 분할된 부분 영역으로부터 다시 대표 모션의 방향과 크기를 구하고,

초점을 향한 대표 모션 방향의 개수가 임계값 이상이면 이 프레임을 "줌 인/아웃" 경우로 간주한다. 이 때, 작은 크기의 모션을 가진 부분 영역은 무시한다. 만일 "팬", "틸트" 또는 "줌" 경우를 만족하지 않을 경우 "카메라 모션 없음"이라고 간주한다.

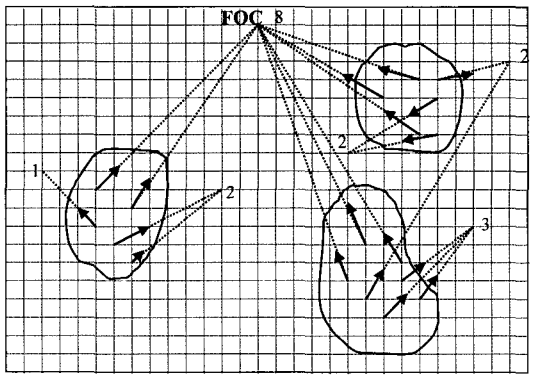


그림 2 "줌"을 검출하기 위한 초점 계산 방법

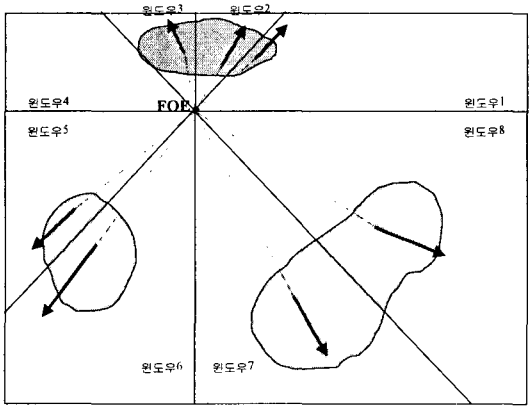


그림 3 "줌" 테스트를 위한 부분영역 분할

프레임에 존재하는 모션을 계산한 후, 이를 이용하여 연속된 프레임으로 구성된 비디오 셧에 존재하는 카메라 모션을 계산한다. 각 프레임에서 계산된 모션의 오류를 제거하기 위해, 우리는 슬라이딩 윈도우 개념을 사용하여 비디오 셧의 카메라 모션을 측정한다. 셧이 시작되는 시점에 초기 윈도우를 설정한다. 이 윈도우 안에 들어있는 프레임들로부터 대표되는 모션을 계산한다. 다음에, 이 윈도우를 시간 축 상으로 한 프레임씩 움직이면서 각 윈도우의 대표 모션을 구하고 초기의 대표 모션과 동일한 가를 체크한다. 이 때, 대표 모션은 윈도우 안 에서 가장 많은 개수를 갖는 모션이다. 만일, 계산한 대표 모션이 초기의 대표 모션과 다르면, 우리는 초기의 카메라 모션이 종료했다고 간주한다. 이것은 그림 4에 잘 나타나 있다. 그림 4에서 보면 사이즈 5의 윈도우를 설정하고 초기 윈도우 안에서 "팬(P)"을 가진 프레임수가 제일 많으므로, 초기 대표 모션은 "팬(P)"으로 정하였다. 윈도우를 시간 축 상으로 하나씩 증가하면서 윈도우가 21번 프레임까지 갔을 때 대표 모션이 초기 모션과 같지 않은 것을 검출하고, 이 윈도우 안에서 초기 대표 모션과 다른 모션을 갖는 프레임을 찾아서 모션의 경계를 검출한다. 여기서는 "팬"이 1번부터 22번까지이고 23번부터는 "카메라 모션 없음(N)"의 프레임이 시작된다. 즉, 이 셧은 "팬"과 "카메라 모션 없음"의 모션으로 구성되어 있음을 알 수 있다. 이러한 방식으로 비디오 셧을 카메라 모션에 따라 분류한다. 그림 5는 제안된 방식의 카메라 모션 검출 알고리즘을 보여주고 있다.

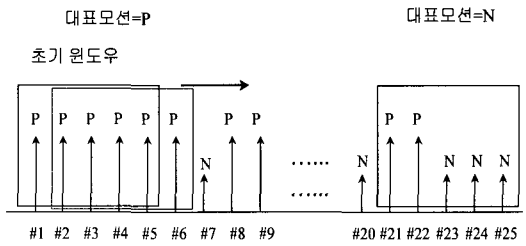


그림 4 슬라이딩 윈도우를 이용한 셧에서의 모션 검출 방법

3.2 퍼지 이론에 근거한 비디오 컨텐트 표현

모션에 의해 분류된 셧에 대하여 그 셧의 특성을 잘 나타내도록 대표 프레임을 선택하는 것이 바람직하다. 대표 프레임을 선택하는 데 있어서 가장 중요한 것은 비디오 셧에 존재하고 있는 컨텐트를 효과적으로 표현하는 것이다. 본 논문에서는, 이러한 컨텐트를 의미있는 특징 및 이 특징들의 시간적인 변화로 나타낸다. 우선 의미있는 특징들로서 영역의 사이즈, 평균 밝기 및 중심 좌표 등을 선택한다. 다음에 공간 영역 및 시간 영역에서 이런 특징들의 변화를 계산한다. 왜냐하면, 카메라 모션이 없는 경우에는 공간 영역에서의 중요도가 대표 프레임을 선택하는 기준이 되고, 카메라 모션이 있는 경우에는 특징들의 변화가 심한 프레임이 대표 프레임이 될 가능성이 많다. 그림 6은 두 가지 경우를 보여 주고 있는데 그림 6(a)와 같이 카메라 모션이 없는 경우 프

레임 #3처럼 프레임 자체가 의미가 있는 것이 중요하고, 그림 6(b)처럼 카메라 모션이 있는 경우, 전 프레임에 비해 컨텐트가 많이 변화하는 프레임 #19를 대표적인 프레임으로 간주할 수 있다.

3.2.1 공간 영역에서의 컨텐트의 중요도

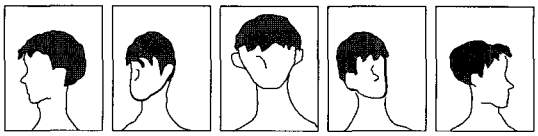
공간 영역에서 프레임이 갖고 있는 중요도를 계산하는 것은 매우 주관적인 문제이므로, 우리는 커다랗고 밝은 영역이 프레임 가운데 위치하였을 때 그 프레임이 셋 중에서 가장 중요하다고 가정한다. 인간이 느끼는 데로 프레임의 중요성을 계산하는 것은 매우 어려운 문제이다. 본 논문에서는 컨텐트가 갖고 있는 의미의 정도를 표현하기 위해 퍼지 이론을 사용한다[11]. 앞서 기술하였듯이 컨텐트로는 의미 있는 특징중의 하나인 영역을 사용한다. 이 영역들의 속성인 영역 크기(S), 평균 밝기(I) 및 중심 위치(L)에 대하여 퍼지 멤버쉽 함수를 구현하여 각각의 퍼지 멤버쉽 값(μ_S, μ_I, μ_L)들을 프레임의 중요도를 계산하는데 이용한다.

```
==========================================
Choose five dominant regions from each frame
Compute dominant motion phase and motion intensity from
five dominant regions
Compute quantized HSV color histogram
Detect Shot boundaries
For each shot:
      for(i= strat_frame_no; i< end_frame_no; i++)
{
/* panning or tilting test */
    if (number of regions  whose dominant motion phase =
      Θ ) Threshold)
  frame_motion[i] = panning(or tilting) with Θ
    else
/* zooming test */
  compute FOE or FOC
        if (FOE || FOC)
    frame_motion[i] = zooming
  else
    frame_motion[i] = no_camera_motion
}
shot_motion = sliding_window(window_size, start_frame,
end_frame, frame_motion)
==========================================
```

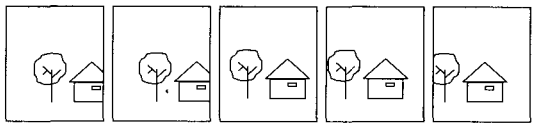
그림 5 카메라 모션 검출 알고리즘

영역으로 분할 된 각 프레임의 중요도를 계산하기 위해 우선 각 영역들로부터 의미있는 정도를 계산한다. 다음에 이 영역들로 구성되어 있는 프레임 전체에 대한 중요도를 계산한다. 이러한 의미있는 정도를 계산하기 위해 여러 가지 실험을 하였다. 영역은 세 개의 속성을 가지고 있으므로, 이 속성들이 각기 다른 값을 갖는 하나의 영역으로 이루어진 30 장의 그림에서 가장 크고 밝고 중심에 위치해 있는 프레임을 기준으로하여 이에 가까운 순서데로 프레임을 정렬시키는 실험을 10명의 학생에게 행하였다. 그림 7은 실험 데이타의 일부분이다. 여기서 얻은 결론은 한 영역의 의미있는 정도는 그 영역의 속성들의 퍼지 멤버쉽 값들의 최소값(conjunction)으로 표현된다. 예를들어, 한 개의 커다란 영역이 프레임 중심에 위치하고 있지만 그 영역의 밝기는 밝지 않을 때, 그 영역은 다른 두 개의 높은 퍼지 멤버쉽 값(μ_S, μ_L) 에도 불구하고, 밝기에 대한 퍼지 멤버쉽 값(μ_I)에 의해 중요도가 결정된다. 따라서, 우리는 다음과 같은 규칙을 만들었다. 즉, 프레임 f_i의 영역 j에 대한 중요도의 퍼지 멤버쉽 값은 다음과 같이 계산할 수 있다.

Rule 1: $\mu_{region\_spatial}(f_i^{\ j}) = min\ (\mu_I(f_i^{\ j}),\ \mu_S(f_i^{\ j}),\ \mu_L(f_i^{\ j}))$



(a)



(b)

그림 6 (a) 카메라 모션이 없는 경우, (b) 카메라 모션이 있는 경우

영역에 대한 중요도를 바탕으로, 여러 영역으로 구성된 프레임에 대해 똑 같은 실험을 행하였다. 각 프레임의 중요도는 각 영역의 중요도에 대한 퍼지 멤버쉽 값들의 최대값(disjunction)으로 표현된다. 예를들어, 한 프레임이 중심 부분에 한 개의 커다란 밝은 영역을 가지고 있고, 다른 부분에 작은 크기의 어두운 영역들을 가지고 있어도, 그 프레임은 다른 영역에 상관없이 오로지 커다란 영역에 의해 인식되어지기 때문이다. 따라서, 영역 j, k 및 l 로 구성된 프레임의 중요도는 다음과 같이 계산할 수 있다.

Rue 2: $\mu_{frame\_spatial}(f_i) = max\ (\mu_{region\_spatial}(f_i^{\ j}), \mu_{region\_spatial}(f_i^{\ k}), \mu_{region\_spatial}(f_i^{\ l}))$

3.2.2 시간 영역에서의 컨텐트의 중요도

카메라 모션이 있는 비디오 셧에서의 프레임의 중요도를 계산하기 위해 시간 영역에서의 의미있는 특징들의 변화량을 계산한다. 특징들의 변화량은 기준 프레임과 해당되는 프레임 사이의 컨텐트의 변화량이므로 본 논문에서는 프레임에 있어서 대응되는 주 영역들의 크기의 변화, 영역 중심의 이동 및 영역 평균 밝기의 변화 등의 세가지 변화량을 사용한다. 이런 변화들은 선형적인 관계에 있지 않기 때문에, 이들 변화량을 측정하기 위해 앞에서와 유사한 실험을 행하였다.

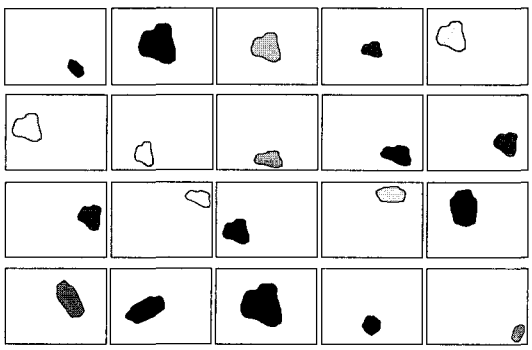


그림 7 실험 데이타

먼저, 한 개의 영역이 시간 축 상의 변화에 대한 변화량을 계산하기 위해 영역의 속성들의 변화량인 크기 변화, 중심 변화 및 밝기 변화를 퍼지 멤버쉽 값(μ_{SC}, μ_{LC}, μ_{IC})으로 계산하고, 시간적으로 떨어진 두 프레임 간의 영역의 변화량은 각 속성들의 퍼지 멤버쉽 값의 최대값(disjunction)으로 계산한다. 예를들어, 작지만 밝고 중심에 위치한 영역이 밝기와 중심이 조금 변화하고, 위치가 많이 이동되었을 때, 인간은 많이 변화한 위치 속성을 강하게 느끼기 때문이다. 각 영역에서의 두 프레임(f_{ref} ,f_{tar})간의 컨텐트의 변화량은 다음과 같은 규칙으로 계산한다.

Rule 3: $\mu_{region\_temporal}(f_{ref}^{\ j}, f_{tar}^{\ j}) = max\ (\mu_{IC}(f_{ref}^{\ j}, f_{tar}^{\ j}), \mu_{SC}(f_{ref}^{\ j}, f_{tar}^{\ j}), \mu_{LC}(f_{ref}^{\ j}, f_{tar}^{\ j}))$

여러 개의 영역으로 이루어진 프레임 간의 컨텐트 변화량을 계산하기 위해, 우리는 실험에의해 각 영역들의 컨텐트 변화량중 최대로 변화한 컨텐트 변화량을 프레임 전체의 컨텐트의 변화량으로 계산한다. 왜냐하면, 프레임에서의 특징은 그 프레임에서의 가장 두드러진 것에 의해 특징 지워지기 때문에 전체 변화량은 가장 두드러진 특징의 변화량만큼 감지되기 때문이다. 3개의 영역을 갖고 있는 두 프레임(f_{ref}, f_{tar})간의 시간축 상의 컨텐트 변화량은 다음과 같은 규칙으로 계산한다.

Rule 4: $\mu_{frame\_temporal}(f_{ref}, f_{tar}) = max(\mu_{region\_temporal}(f_{ref}^{\ j}, f_{tar}^{\ j}), \mu_{region\_temporal}(f_{ref}^{\ k}, f_{tar}^{\ k}), \mu_{region\_temporal}(f_{ref}^{\ l}, f_{tar}^{\ l}))$

4. 계층적인 대표 프레임 선택

프레임에 대한 공간상의 중요도와 시간 축 상의 변화량을 바탕으로 이 장에서는 대표 프레임 선택 방법에 대해 기술한다. 비디오 셧은 그 자체가 갖고 있는 카메라 모션에 따라 특이한 성질을 나타내므로, "팬", "틸트", "줌" 및 "카메라 모션 없음" 경우로 분류된 셧에 대표 프레임 선택을 위한 적절한 규칙을 정한다. 대표 프레임은 먼저 기준이 되는 대표 프레임을 정하고 이를 바탕으로 계층적으로 다음 단계의 대표 프레임을 정한다. 이때 대표 프레임이 선택되는 기준 척도를 조절함으로서 대표 프레임 수를 정하는 방법에 관해서도 기술한다.

4.1 대표 프레임 선정 기준

셧을 대표하는 대표 프레임을 선정하는데 있어서 그 기준은 셧의 카메라 모션에 따라 다르게 정한다. "팬"이나 "틸트"인 경우에, 첫 번째 프레임과 마지막 프레임을 대표 프레임의 후보로 선택한다. 왜냐하면, 일반적으로 "팬"이나 "틸트" 셧을 찍는 카메라 감독의 의도가 이 두 프레임에 있기 때문이다[12]. 이러한 두 개의 프레임 중에서 첫 번째 프레임을 기준 프레임으로 선택하여 다음 규칙에 따라 대표 프레임을 선택한다.

Rule 5: 기준 프레임과 현재 프레임의 컨텐트 변화량을 Rule 4에의해 계산하고 그 값이 임계값을 넘으면 그 프레임을 대표 프레임으로 선택하고, 그 프레임을 기준 프레임으로 하여 같은 작업을 반복한다.

그림 8은 이런 과정을 보여주고 있다.

"줌"인 경우에는 하나의 장면이 확대 또는 축소되는 경우로 마지막 프레임이 카메라 감독의 의도가 반영되

어 있는 중요한 프레임이므로, 마지막 프레임을 대표 프레임으로 선택한다[12]. "카메라 모션 없음"의 경우는 다음 규칙에 의해 대표 프레임을 선택한다.

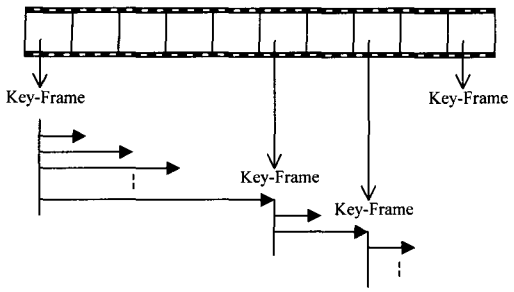


그림 8 "팬" 이나 "틸트" 모션을 가진 셧에서 대표 프레임 선택 방법

Rule 6: 첫 번째 대표 프레임은 셧 중에서 Rule2에의해 가장 중요하고 의미있는 프레임을 선택한다. 다음에 이 프레임을 기준 프레임으로 하여 양방향으로 컨텐트 변화량을 Rule 4에의해 계산하여 임계값 이상이면 그 프레임을 대표 프레임으로 선정하고 다시 이 프레임을 기준 프레임으로 하여 같은 과정을 반복한다.

즉, 그림 9에서처럼 셧 중에서 Rule 2에 따라 가장 의미있는 프레임을 먼저 선정한다. 다음에 이를 기준으로 양방향으로 컨텐트의 변화량을 Rule 4를 사용하여 계산하여 대표 프레임들을 선택한다. 이렇게 함으로서 카메라 모션이 없는 경우 오직 중요한 특징을 갖는 프레임들이 셧을 대표한다.

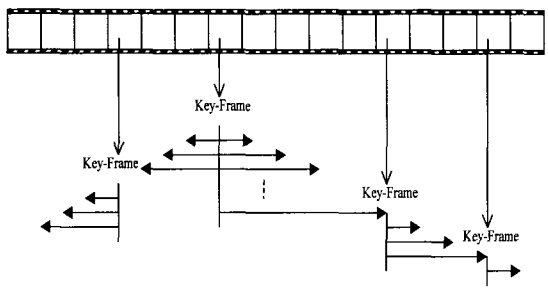


그림 9 "카메라 모션 없음"에서의 대표 프레임 선택 방법

더욱이, 셧이 한 개 이상의 모션으로 이루어 진 경우에는 그 셧을 한 가지의 모션을 갖는 부분 셧으로 분할하고, 각 부분 셧의 카메라 모션에 따라 적합한 규칙을 적용하여 대표 프레임을 선택한다. 예를 들어, "줌"과 "카메라 모션 없음"의 두 개의 부분 셧으로 구성된 셧에서는 먼저 "줌"을 갖는 셧에 대해 줌을 종료하는 프레임을 대표 프레임으로 선택하고, "카메라 모션 없음"의 부분 셧에서는 Rule 6에 따라 대표 프레임을 선택한다.

4.2 대표 프레임 개수 조정

비디오 셧을 대표하는 프레임은 그 응용 분야에 따라 한 개 또는 여러 개를 선택할 필요가 있다. 예를들어, 비디오 데이타를 중요한 장면만을 요약해서 보고 싶으면, 셧당 대표 프레임 개수를 줄이고, 또 조금 자세히 비디오 데이타를 브라우징하고 싶으면 대표 프레임 개수를 증가시키면 된다. 다양한 방식의 브라우징을 위해서, 또 시스템의 저장 한계를 만족하기 위해서는 각 비디오 셧에서의 대표 프레임이 선택되는 개수를 조정할 수 있는 방법이 필요하다. 앞 절에서 기술하였듯이 대표 프레임 선택 방법은 컨텐트의 중요도와 변화량을 계산하여 사용하였으므로, 브라우징하려는 비디오 데이타의 요약 정도에 따라 이러한 변화량을 조정함으로써 가능해진다. 본 논문에서는 브라우징의 상세도(detailness)에 따라 **COARSE, MEDIUM, FINE** 세 가지의 등급을 사용한다. 이러한 상세도는 구해진 컨텐트 변화량에 관한 퍼지 멤버쉽 값에 α-cut 동작을 수행함으로서 얻을 수 있다. 즉,

$$KEY\_FRAMES = \{f_{ref}, f_{tar} \in S, \mu_{frame\_temporal}(f_{ref}, f_{tar}) \geq \alpha\} \quad (3)$$

여기서 S는 비디오 셧을 나타내고, f_{ref} 과 f_{tar} 은 각각 기준 프레임과 현재 프레임을 나타낸다. 현재 프레임에서 기준 프레임과의 컨텐트의 변화량을 계산하여, 그 변화량이 α보다 클 때, 대표 프레임으로 선택된다. 예를 들어, 상세도가 **COARSE**인 경우, 매우 중요한 프레임만을 브라우징 하는 경우이므로, 우리는 α-cut 값을 매우 높게 선정하여 대표 프레임을 선택할 수 있다. 이렇게 할 경우 대표 프레임 개수가 작아지므로, 작은 개수의 프레임만을 브라우징하고, 또 저장할 수 있어 저장 공간을 감소시킬 수 있다. 아울러, 정해진 개수의 대표 프레임을 선정하기 위해서는 적절한 α-cut 값을 알아야하는데 이렇게 하기 위해서는 α-cut 값을 1부터 조금씩 감소하면서 얻어지는 대표 프레임 개수를 원하는 프레임 개수와 비교하여 적절한 α-cut 값을 정하여 사

용한다.

5. 실험 결과

시뮬레이션을 위해 "팬"의 카메라 모션을 갖고 있는 flower.mpg, "줌 인/아웃"으로 이루어진 wg_cs_5.mpg, 그리고 카메라 모션이 별로 없는 mjackson1.mpg, news.mpg, 및 NewsClip.mpg 등의 비디오 데이타를 사용하였다. 먼저, MPEG 비디오 데이타로부터 원 영상을 복원하지 않고 DC 영상만을 복원하여, 모폴로지 필터의 opening 작업(*opening by partial reconstruction)*을 통하여 DC 영상을 단순화 시켰다. 단순화된 영상으로부터 수정된 watershed 알고리즘을 이용하여 주 영역들을 계산하였다[10]. 각 프레임 당 5개의 주 영역과 이들의 밝기 값 평균, 사이즈, 그리고 중심 좌표 등을 계산하였다. 이런 공간적인 분할을 통하여 얻어진 영역을 다음 프레임에 투사하여 시간 축 상의 영역들간의 대응 관계를 계산하여 영역으로 분할 된 비디오 데이타를 얻었다.

공간적으로 분할된 영역을 바탕으로 시간 축 상의 영역의 흐름을 찾는 것이 셧 경계 검출에 매우 중요하다. 시간 축 상의 영역의 흐름을 찾기 위해서는 먼저 기준 프레임의 각각의 영역에 대해 다음 프레임의 영역들에 대한 대응 관계(correspondence)를 찾아야 한다. 일반적으로 대응 관계를 찾는 것은 매우 어려운 문제이지만, 비디오 데이타에서는 프레임간에 오브젝트의 움직임이나 사이즈의 변화가 크지 않고, 또 밝기 값의 변화도 크지 않는 특성을 가지고 있으므로, 기준 프레임의 영역을 다음 프레임에 투사하여 해당되는 영역부근에서 대응 영역을 찾는 것이 바람직하다. 다시 말해서, 현재 프레임의 영역이 다음 프레임에서도 비슷한 위치에 있고, 또 영역의 밝기나 사이즈에서 값의 차이가 임계 값 범위 안에 드는 영역을 찾는다면 대응되는 영역으로 간주하여 기준 프레임의 영역과 같은 라벨을 붙인다. 만약, 밝기 정도나 사이즈가 임계값을 벗어나면 새로운 영역이 생긴 것으로 간주하여 새로운 라벨을 붙인다. 그림 10은 이러한 과정을 보여 준다.

분할된 비디오 데이타로부터 시간 축 상의 영역의 불연속 프레임을 찾고, 여기에다 칼라 히스토그램 변화량을 이용하여 셧 경계를 구별하였다[13]. RGB 칼라 스페이스에서는 일반적으로 두 개의 칼라의 근접도(proximity)가 두 칼라의 유사도(similarity)를 나타내지 않기 때문에, 칼라 정보 유사도 측정에 어려운 점이 많다. 더욱이, 24 비트의 RGB 칼라 정보를 사용한다면, 얻는 성과에 비해 계산하는데 많은 시간 및 자원을 소

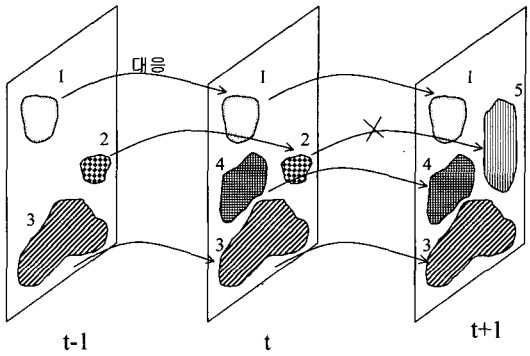
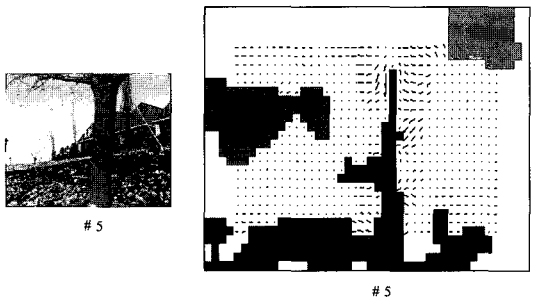
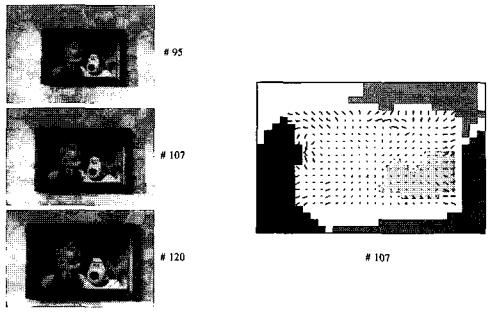


그림 10 시간 축 상의 영역의 대응 관계



(a) "팬"



(b) "줌"

그림 11 카메라 모션 검출: (a) "팬" 모션, (b) "줌" 모션

모하기 때문이다. 이러한 단점을 보완하기 위해서 본 논문에서는 HSV(Hue, Saturation, Value) 칼라 스페이스를 사용하였다. 이 칼라 스페이스는 자연적이어서, 다른 조명 조건하에서도 색상(Hue)이 변치 않는 장점을 가지고 있다. HSV의 값은 보통 RGB의 비선형 변환으로 구할 수 있다. 이렇게 구해진 HSV 칼라 스페이스를 인

간 시각 시스템에 근거하여 97개로 비 선형적으로 양자화하여, 각각의 프레임 간의 칼라 히스토그램 변화를 계산하였다. 영역 정보와 양자화된 HSV 칼라 히스토그램 변화를 이용하여, 95% 정도의 검출율로 셧 경계를 검출하였다.

각 셧은 영역으로 분할 된 DC 영상들로 구성되어 있으므로, 주 영역에 존재하는 옵티칼 플로우를 이용하여 각 영역의 대표 모션 벡터의 방향과 크기를 계산하였다. 모션 방향은 8개의 방향으로 양자화하였다. 이 때, 작은 모션 크기를 갖고 있는 것은 무시하였다. 3 장에서 기술한 알고리즘에 따라 각 프레임에 존재하는 모션을 검출하였다. 그림 11은 영역으로 분할 된 데이타에서 검출된 카메라 모션을 보여주고 있다. 그림 11(a)는 flower.mpg에서 "팬" 모션이고, 그림 11(b)는 "줌" 모션을 보여주고 있다. 표 1은 모션 검출 결과를 보여준다. 각 프레임에 대한 정확한 모션 검출 율은 비디오 데이타에 따라 다양하지만, 실제로 셧에 대한 모션은 3.1절에서 제안한 슬라이딩 윈도우를 사용하므로, 윈도우 사이즈가 5 인 경우 연속된 3장의 프레임에 대한 카메라 모션이 잘못 검출 되었을때 부 정확한 모션을 검출하게 되지만, 이러한 확률은 표 1에서처럼 매우 작으므로, 비디오 셧에 대한 정확한 모션을 검출할 수 있다.

표 1 모션 검출 결과

비디오데이타	카메라모션	검출 프레임 수	검출율	연속된 3장의 프레임 모션 검출이 fault 인 횟수 (검출율)
mjackson1.mpg	카메라모션없음: 100	97	97%	0 (0%)
wg_cs_5.mpg	줌아웃: 27	27	100%	1 (0.7 %)
	줌 인: 46	32	67%	
	카메라모션 없음: 56	52	93%	
flower.mpg	팬: 150	117	78%	1 (0.6%)
news.mpg	카메라모션 없음: 67	61	91%	0 (0%)
NewsClip.mpg	카메라모션 없음: 60	50	83%	0 (0%)

검출된 카메라 모션에 따라 비디오 셧을 "팬", "틸트", "줌" 및 "카메라 모션 없음" 의 셧으로 구분한다. "팬" 모션 또는 "틸트" 모션을 갖는 셧에는 처음 프레임과 마지막 프레임이 카메라 감독의 의도가 반영된 중요한 프레임이므로[12], 이 두 프레임을 기준으로 Rule 5를 적용하여 찾는다. "줌" 모션을 갖는 셧에는 마지막 프레임이 중요한 프레임이므로, 셧에서의 마지막 프레임을 대표 프레임으로 선정하였다. 그리고, "카메라 모션 없음"의 경우에는 Rule 6을 적용하여 대표 프레임을 선택하였다. 그림 12는 논문에서 제안한 알고리즘에 의해 mjackson1.mpg에서 선택된 대표 프레임들을 보여 주고 있다. 커다란 크기의 영역이 중심에 위치한 프레임들이 선택되었다. 예를들어, 프레임 #34는 프레임 #25부터 프레임 #36까지의 "카메라 모션 없음" 셧에서 처음으로 선택된 대표 프레임이고, 이를 기준으로 컨텐트 변화량이 큰 프레임 #28이 두 번째의 대표 프레임으로 선택되었다. 그림 13은 대표 프레임을 선택하는데 있어서의 상세도 문제를 보여주고 있다. 그림 13(a)는 flower.mpg에서 상세도가 **COARSE**인 경우의 선택된 대표 프레임들을 보여 주고, 13(b)는 상세도가 **FINE**인 경우의 선택된 대표 프레임들을 보여주고 있다. 브라우징의 상세도를 조절함으로서 대표 프레임의 개수를 조절할 수 있어 계층적으로 비디오 데이타를 브라우징할 수 있다.

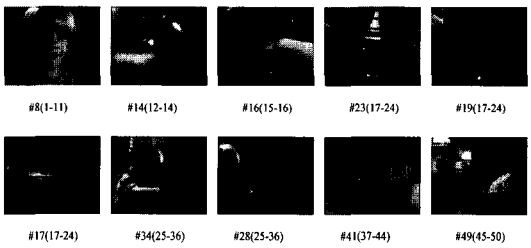


그림 12 mjackson1.mpg에서 선택된 대표 프레임(괄호 안의 번호는 셧에 속한 시작 프레임 및 종료 프레임 번호)

대표 프레임의 선택에 있어서 정확도를 측정하는 것은 매우 어려운 문제이다. 왜냐하면, 인간에 의한 대표 프레임 선택도 실제로는 매우 주관적으로 이루어지기 때문이다. 다른 관점에서 보면, 본 논문에서 제안된 방식은 인간에 의한 대표 프레임 선택 방식보다 더 객관적일 수 있다. 표 2는 사람이 선택한 대표 프레임들과 제안된 알고리즘에 의해 선택된 대표 프레임들 간의 비교 결과를 보여준다. 비슷한 개수의 대표 프레임을 선택하였을 경우, 인간이 선택한 대표 프레임들 중에서 60% 이상의 같은 프레임들이 제안된 알고리즘에 의해 선택되었고, 나머지 프레임들도 주관에 따라 대표 프레임으로 선택될 수 있는 것이었다. 오직 5% 정도 내외만 차이가 있을 뿐이다. 2 장에서 기술한 기존의 대표 프레임 선택 방식과 비교할 때 제안된 방식은 컨텐트의 중요

도를 주 영역의 크기, 중심 위치, 밝기등의 공간적인 중요도와 시간 축상의 변화량에 의존하여 처리함으로서 인간이 느끼는 세만틱(semantics) 개념에 가까운 방식이다. 또, MPEG의 압축 데이타를 완전히 복호화하지 않고 DC 데이타에서 주 영역을 추출하여 영역에 관련된 컨텐트의 중요도를 계산함으로써 많은 오버헤드를 줄일 수 있다.

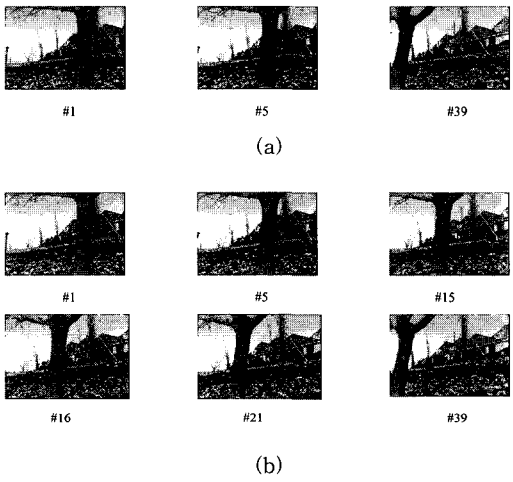


그림 13 다양한 상세도에 의한 대표 프레임 선택: (a) COARSE degree, (b) FINE degree

표 2 대표 프레임 검출 결과

카메라 모션	사람이 검출한 개수	제안된 알고리즘 적용(degree=COARSE)	사람이 검출한 프레임이 제안된 알고리즘에 의해 선택된 개수(선택 율)
mjackson1.mpg (100 frame)	15	19	11 (73%)
wg_cs_5.mpg (130 frame)	5	6	4 (80%)
flower.mpg (150 frame)	8	10	5 (62.5 %)

6. 결 론

본 논문에서는 컨텐트의 중요도 및 시간 축 상의 변화량을 영역 정보, 모션 정보및 퍼지 이론을 바탕으로 계산하여 대표 프레임을 선택하는 방법을 제안하였다. 대표 프레임을 선택하기 위해, 먼저 비디오 셧을 카메라 모션에 따라 "팬", "틸트", "줌" 및 "카메라 모션 없음"으로 구분하였고, 분류된 각 비디오 셧에 컨텐트의 중요도를 계산하기 위해 제안한 규칙들을 적용하여 대표 프레임을 선택하였다. 또, 사용자가 원하는데로 브라우징의 상세도를 결정하여 대표 프레임의 수를 조절할 수 있게 하였다. 이것은 다양한 비디오 브라우징과 제한된 저장 장치에 유용하게 사용될 수 있다.

제안된 방식은 비디오 데이타에 있어서의 시공간 상에 존재하는 중요한 특징들을 이용하였으므로, 이러한 특징들을 바탕으로 고급 개념들을 결합시켜 세민틱을 반영하는 비디오 구조를 추출하는 연구가 바람직하다. 다시말해서, 선택된 대표 프레임들로부터 사용자가 원하는 방식의 클러스터링을 수행하여 비구조화된 비디오 데이타에서 의미있는 구조를 추출하여, 사용자로 하여금 다양한 내용 기반의 비디오 인덱싱 및 검색(content-based video indexing and retrieval)을 가능케 하는 것이다. 이렇게 함으로서 사용자는 궁극적으로 비디오 데이타로부터 다양한 브라우징 및 네비게이션을 수행할 수 있다. 이러한 기술들은 디지털 라이브러리(digital library) 분야, 방송 프로그램 검색 및 편집, 주문형 비디오, 엔터테인먼트 및 의료 분야에 이르기까지 광범위한 응용 분야에 적용할 수 있다.

참 고 문 헌

[1] W. Grosky, R. Jain and R. Mehrotra, *The Handbook of Multimedia Information Management,* Prentice Hall, 1997.

[2] B. Furht, S. Smoliar and Zhang, *Video and Image Processing in Multimedia Systems,* Kluwer Academic Publishers,1995.

[3] M. Cascia, E. Ardizzone,"JACOB: Just A Content-Based Query System for Video Databases," *Proc. ICASSP '96,* pp. 1216-1219, 1996.

[4] W. Wolf, "Key Frame Selection by Motion Analysis," *Proc. ICASSP '96,* pp. 1228-1231, 1996.

[5] H. Zhang, J. Wu, D. Zhong and S. Smoliar, "An Integrated System for Content-based Video Retrieval and Browsing," *Pattern Recognition,* 30(4), pp. 643-658, 1997.

[6] R. L. Lagendijk, A. Hanjalic, M. Ceccarelli, M. Coletic, and E. Person, "Visual Search in a SMASH System," *ICIP '96,* pp.671-674, 1996.

[7] X. Sun, M. Kankanhali, Y. Xhu, J. Wu, "Content-Based Representative Frame Extraction for Digital Video," *Proc. IEEE Multimedia System'98,* pp.190-193, 1998.

[8] B. D. Lucas, T. Kanade, " An Iterative Technique of Image Registration and Its Application to Stereo," *Proc. IJACI*, pp. 674-679, 1981.
[9] M. Pardas and P. Salembier, "3D Morphological Segmentation and Motion Estimation for Image Sequences," *Signal Processing*, Vol. 38, pp. 31-43, 1994.
[10] L. Vincent and P. Sollie, "Watersheds in Digital Spaces: An Efficient Algorithms based on Immersion Simulation," *IEEE Trans. Pattern and Mach. Intell.*, Vol. 13, No. 6, pp. 583-598, June 1991.
[11] J. Jang, C. Sun and E. Mizutani, *Neuro-Fuzzy and Soft Computing*, Prentice Hall, 1997.
[12] S. Katz, *Shot by Shot*, 1998.
[13] 강행봉, "영역 흐름 및 칼라 정보를 이용한 MPEG 데이타의 내용 기반 셧 경계 검출", *정보과학회 논문지*, 제 27권 4호, 2000.

강 행 봉
정보과학회논문지 : 소프트웨어 및 응용
제 27 권 제 4 호 참조