

멀티미디어 서비스를 위한 동영상 특징 정보 분석에 관한 연구

김용균*, 이창수, 오해석
송실대학교 컴퓨터공학과 멀티미디어 연구실
e-mail : se4870@multi.soongsil.ac.kr

The Study Of Video Feature Information Parsing For Multimedia Services

Yong-Gyun Kim, Chang-Soo Lee, Hae-Seok Oh
Dept. of Computer Science, Soongsil University

요약

급속한 컴퓨터 및 통신 기술의 발전으로 인해 동영상 정보는 인터넷 및 사회전반의 다양한 분야에서 활용되고, 기하급수적으로 증가되고 있으나, 기존의 정보 분석 시스템은 기본적으로 텍스트를 기반으로 한 것이기에, 동영상 정보가 가지는 애매성을 표현하기 곤란하며, 주석 작성에 따르는 과도한 작업부담 및 객관성 결여들의 문제점을 가지고 있다.

이에, 상기에 기술한 문제점을 해결하기 위하여 대용량의 동영상 정보를 효율적으로 분석 할 수 있는 동영상 분석 알고리즘을 제안하고자 한다. 본 논문의 주된 연구과제는 컷 검출을 위한 통합 특징량 적용법과 색상특징, 모양특징을 사용하여 이미지 내의 특징 정보를 분석하는 것이다.

1. 서론

최근의 기술은 방대한 양의 디지털 정보의 수준에 도달하였으나 이에 비해 낮은 사용률을 보인다. 동시에 디지털 저장 매체의 실행 비율은 꾸준히 증가하고 있다. 디지털 정보를 간단하고 낮은 비용으로 얻을 수 있는 것뿐만이 아니라 정보를 저장하고 모으는데 점점 더 쉽게 만들어 내고 한계성을 줄여가고 있다. 그러므로 전문가나 소비자의 환경에서 디지털 처리는 빠르게 성장하고 있다 [1].

디지털 비디오 압축분야의 발전은 또한 디지털 비디오의 생산을 가능하게 하였다. 집에서 비디오 매체의 디지털 성장은 현재 아날로그 비디오카세트 저장 시스템을 곧 바꾸어 놓을 것이라는 것을 기대할 수 있다.

디지털 라이브러리가 가지는 문제점은 방대한 정보의 양을 관리·보존하는 것이다. 즉 기존의 문자

기반(SQL)방식을 해결하는 것이다. 그러나 검색방법, 질의 방법과 시각 정보(이미지, 그래프, 비디오), 그리고 이러한 부분들의 문자정보를 연결하는 것은 여전히 초기수준밖에는 미치지 못한다[2].

동영상에서 시각 정보를 위한 표현을 얻기 위해서는 내용분석이 요구된다. 기존의 이미지에 몇 가지 내용분석 방법을 가지고 접근을 시도한다. 이러한 방법은 칼라, 텍스처, 모양을 비교하여 분석을 하게 된다. 연속된 비디오에서 내용 분석은 샷이라는 유사한 공간적인 세그먼트를 통해 처음에 연속성에서 분리하게 되고, 그때 그러한 세그먼트는 하나 또는 몇 개의 프레임으로 요약되고, 시각이나 음성적인 특징을 가지는 샷 사이에서의 관계성을 마지막으로 결정하게 된다[3].

비디오는 크게 프레임(Frame), 샷(Shot), 장면(Scene), 시퀀스(Sequence)의 4가지의 계층적이며 순차적인 데이터의 모임이다. 여기서 프레임이란 비디오를 구성하고 있는 계층 중 최하위 계층이며 하나의 정지

영상이다. 프레임의 상위계층인 샷은 하나의 카메라로 기록된 연속적인 프레임의 모임이며, 장면은 주체가 같은 내용을 가진 인접한 샷의 모임이며, 시퀀스는 비디오의 최상위계층이며 연관된 장면의 모

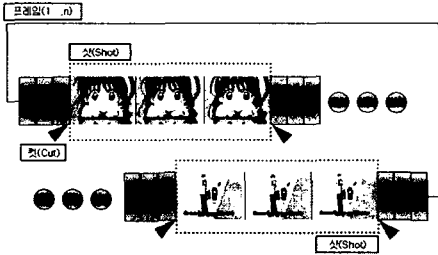


그림 1 비디오의 구성

임이다. 그림 1은 이러한 비디오의 구성을 보여준다.

샷을 검출하기 위한 기존의 제안된 연구방법으로는 공간영역의 특징인 각 픽셀의 차를 이용한 방법 [4], 픽셀값의 히스토그램의 차이법 [5], 분산과 표준편차 등의 통계적인 차를 이용한 방법 [6] 등에 관한 연구들이 있다.

위의 제안된 방법에서 통합 특징방법을 사용하여 샷이 되는 경계프레임을 검출하고 검출된 샷을 통하여 동영상 의미 분석을 한다.

본 논문에서는 좀더 효율적인 의미정보 추출을 위하여 색상정보와 모양정보를 통하여 동영상 데이터의 특징값을 추출해 낸다.

2. 기존의 장면전환 검출 방법

2.1. 픽셀 간의 차를 이용한 방법

인접한 프레임에 대응하는 픽셀간 차를 계산함으로써 두 프레임간의 샷 전환을 감지하는 방법으로 임계치(Threshold)를 계산해 계산된 임계치를 초과하는 프레임을 컷(샷 경계)으로 판단하는 방법이다.

$$DP_i(x, y) = |F_i(x, y) - F_{i+1}(x, y)| \quad (1)$$

$$\frac{\sum_{x,y=1}^{X,Y} DP_i(x, y)}{X \times Y} > T \quad (2)$$

식(1)에서 $F_i(x, y)$ 는 i 번째 프레임 (x, y) 의 밝기 값을 나타내며, $DP_i(x, y)$ 는 인접하는 프레임의 픽셀차를 의미한다. 식(2)의 $X \times Y$ 는 프레임의 크기를 나타내며, T 는 컷을 판별하기 위한 임계치이다. 이 방법은 구현이 아주 간단한 반면에 카메라

움직임에 민감하며, 영상 각각의 픽셀에 대해 비교를 해야하기 때문에 시간이 많이 걸린다는 단점이 있다 [7].

2.2 히스토그램의 차이를 이용한 방법

이 방법은 가장 일반적인 방법으로 연속되고 인접한 프레임간의 히스토그램 차이를 계산해 주어진 임계치와 비교해 임계치를 초과할 경우 샷으로 판단하는 방법이다. 히스토그램의 차를 이용한 방법은 카메라의 움직임에 강인하며, 연산속도가 빠르다는 장점이 있다 [5].

$$\sum_{j=1}^G |H_i(j) - H_{i+1}(j)| > T \quad (3)$$

식(3)의 G 는 칼라 레벨의 수이며, H_i 은 i 번째 프레임의 히스토그램을 의미한다.

2.3 통계적인 차를 이용한 방법

인접하는 프레임간의 통계적인 차를 이용한다. 즉 두 프레임간의 평균, 분산, 표준편차 등을 구한 후 임계치를 설정해 샷을 검출하는 방법이다 [6].

$$\frac{[(\frac{\sigma_i + \sigma_{i+1}}{2}) + (\frac{\mu_i - \mu_{i+1}}{2})]^2}{\sigma_i \times \sigma_{i+1}} > T \quad (4)$$

식(4)의 σ_i 는 i 번째 프레임의 분산이며, μ_i 는 프레임 평균을 나타낸다.

3. 동영상 특징 및 의미 정보 추출

샷의 경계가 되는 장면 전환 기법은 크게 두 가지로 나눌 수 있다. 그 중 하나는 급진적인 전환인데 이런 급진적인 전환의 컷 검출은 컷으로 판별되는 프레임과 인접하는 프레임과의 차이가 비교적 크기 때문에 샷을 나누기가 쉽다. 그러나 점진적인 전환은 영상의 시각적인 효과를 위한 카메라 조작으로 장면전환이 여러 프레임에 걸쳐 이루어지므로 컷 검출이 매우 어렵다. 이런 점진적인 전환은 카메라의 조작에 따라서 페이드(Fade), 디졸브(Dissolve)등이 있다.

일반적으로 카메라의 특수효과로 인해 컷 검출시 잘못된 컷을 감지하게 되는데 이러한 잘못된 컷의 감지를 보완하고, 급진적인 장면전환에 대해서도 좀더 정확한 컷을 검출하기 위해 본 논문에서는 기존에 연구되어진 칼라 히스토그램 방법, 픽셀간의 차분법, 통계적인 방법의 세 가지 특징량을 통합하여 컷을 검출하고 또한 샷에서 객체의 유사도를 측정하

여 객체를 분할하고 분할 트리를 구성한다. 유사도 측정에서 좀 더 정확한 측정을 위하여 칼라정보를 사용한다.

3.1 통합 특징량을 이용한 컷 검출

기존에 연구된 각 특징량은 자기 서로 다른 장점과 단점을 모두 가지고 있다. 따라서 이와 같은 각각의 특징량을 통합하여 좀 더 효율적인 결과가 얻어질 것으로 기대할 수 있다[2].

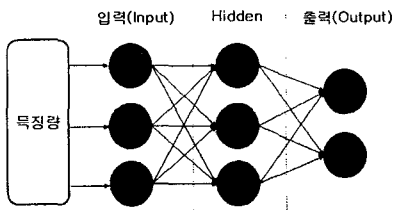


그림 2 통합 특징량 구조

이 논문에서 히스토그램 방법, 픽셀간의 차분법, 통계적인 방법의 세 가지 특징량을 통합 특징량으로 사용한다. 이러한 통합 특징량은 대부분이 중복되는 값이 많기 때문에 처리시간이 상대적으로 줄어들게 되는 장점을 가지고 있다.

통합 특징량을 이용한 방법은 그림 2에서 보는 바와 같이 간단한 인공신경망을 사용하였고 특징량을 정규화하고 입력받는 입력층과 출력층은 컷인지 아닌지를 판단하는 두 개의 뉴런으로 구성되어 있다. 히든층은 입력층과 같은 뉴런의 수를 배치하였는데 이는 특징량의 중복되는 과정을 계산과정에서 배제하기 위해서 이다.

3.2 샷(Shot) 내용의 표현

검출된 컷(Cut)과 컷(Cut) 사이의 샷은 색상, 텍스처, 모양, 경계(edge)등으로 표현을 하게된다. 이러한 특징 중 색상정보와 모양정보를 사용하여 동영상 특징점과 의미정보를 추출한다.

3.2.1 색상특징

원본과 대상 이미지에서 색상은 가장 뛰어난 구별 능력을 가지고 있다. 완전히 다른 객체의 두 이미지가 유사한 색상을 가지는 경우는 매우 드물다. 색상에 대한 근본적인 표현은 평균밝기, 색상 히스토그램, 지배적 색상 그리고 통계적 모멘트를 포함하게 된다.

색상 분포 히스토그램은 분할 기법에 의해 잘 표현되지 않는 텍스처된 이미지에 특히 유용한 양적인 표현이다. 또한, 색상 히스토그램은 변형이나 관점축회전에서 변화가 없으며 관점의 변화나 크기 변화, 폐쇄에서 천천히 변화하게 된다.

$$D^2_{his}(I, Q) = \sum_i^N \sum_j^N a_{ij}(I_i - Q_i)(I_j - Q_j) \quad (5)$$

식(5)에서 N개의 bin으로 각각 구성되는 I 와 Q 두 개의 색상 히스토그램 사이의 거리를 표현하고 있다. 그림 3은 a_{ij} 가 단위행렬(identity matrix)일 때 이 측정법은 유클리드 거리(Euclidean distance)를 나타내고 있다[8].

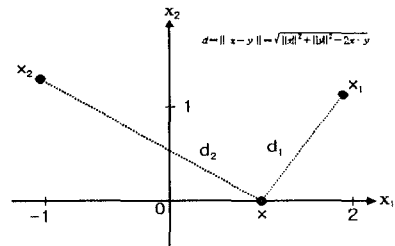


그림 3 유클리드 거리

3.2.2 모양특징

모양정보를 이용하는 방법은 윤곽선 정보를 추출하여 매칭에 이용하거나 불변 모멘트를 계산하여 대표값으로 사용한다. 이미지 내에 지배적 객체를 색상 특징정보와 윤곽선 정보를 사용하여 얻어 낼 수 있다. 모양의 유사성을 비교할 때, 대수적 모멘트(algebraic moment)나 인수곡선(parameter curve)보다 더 인간의 모양 감지에 가까운 측정법을 제공하기 때문에 누적 전환각(cumulative turning angle)를 사용한다. 이러한 값들은 회전이나 이동에도 변하지 않으며, 불록하거나 불록하지 않은 다면체에 대하여 변하지 않고 상대적으로 계산이 쉽다(그림 4).

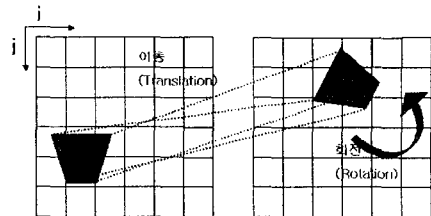


그림 4 불변모멘트의 회전과 이동

4. 실험 결과

본 논문에서 사용된 데이터는 AVI형식의 동영상 파일이며 펜티엄 350Mhz PC를 사용하였고, Visual c++ 6.0을 사용하여 프로그램하고 동영상데이터에서 프레임별로 저장하여 디스플레이 하였다.

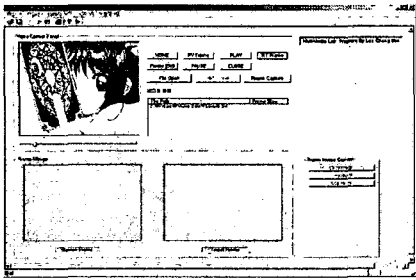


그림 5 동영상 편집기 화면

그림 5는 동영상 화면을 디스플레이하고 로드 된 동영상 파일에서 통합 특징량을 이용하여 장면전환 되는 시점부터 다음 시점까지의 프레임을 BMP파일 형식으로 저장한다. 이러한 통합 특징량을 사용하여 검출한 컷은 각각의 특징량으로 얻어지는 컷보다는 2.5배 이상의 효율을 얻을 수 있었다(그림 6).

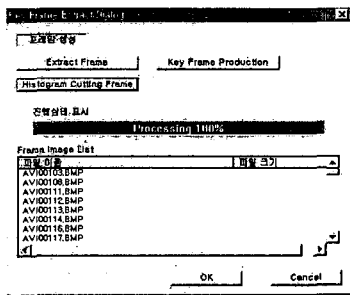


그림 6 컷(Cut) 검출 화면

생성된 프레임은 색상특징 정보와 모양정보를 측정하여 최종 동영상 특징 정보 추출을 얻어 낼 수 있다(그림 7).

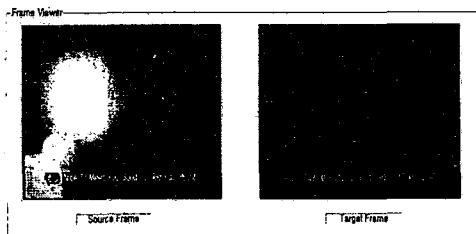


그림 7 특징정보 추출

5. 결론 및 향후 연구과제

본 논문에서 제안한 방법인 통합 특징량을 적용하여 컷을 검출한 결과 각각의 특징량 추출보다는 향상된 컷 검출을 얻을 수 있었다.

검출된 컷은 다시 영상 정보를 얻기 위해 모양특징정보와 색상특징 정보를 사용하여 이미지에서 자막을 추출하는 과정을 간단히 보여주고 있고(그림 7), 색상 특징정보에서 대표 색상을 선출하는 경우와 모양정보에서 불변모멘트를 선출하는 방법을 사용하여 다른 프레임과의 유사성 비교에 사용하여 복잡하지 않은 연산결과로 시간 단축과 양질의 특징정보를 얻어내었다.

향후 과제로는 비디오 내에서 검출되는 특징값을 이용하여 인덱싱 하고, 동영상 브라우저에서 인덱싱된 값만을 이용하여 사용자가 원하는 부분을 정확히 빠르게 탐색할 수 있는 연구가 진행되어야 한다.

[참고 문헌]

- [1] Alan Hanjalic, Reginald L, Lagendijk, "Automated High-level Movie Segmentation For Advanced Video-retrieval Systems", IEEE
- [2] 한국전자통신연구원 정보통신부, "내용기반 동화상 자동분할 S/W 개발", 1999,12.30
- [3] P.Salembier, "Region-Based Representation of Image and Video: Segmentation Tools for Multimedia Services", IEEE (Invited Paper), 1999
- [4] J. S. Boreczky and L. A. Rowe, "Comparison of video shot boundary detection techniques," SPIE Storage and Retrieval for Image and Video Databases, Vol. 2670, pp.170-179, 1994.
- [5] N. V. Patel and I. K. Sethi, "Video Shot Detection and Characterization for Video Databases," Pattern Recognition: Special issue on multimedia, 1996.
- [6] I. K. Sethi and N. Patel, "A Statistical Approach to Scene Change Detection," SPIE Storage and Retrieval for Image and Video Databases, Vol. 2420, 1995.
- [7] 나재형, "내용기반 검색을 위한 뉴스비디오에서 자막추출", 숭실대학교 석사논문, 1997.12.
- [8] Hongjiang Zhang, Et Al, "Video Parsing, Retrieval and Browsing : An intergrated and Content-Based Solution", IEEE