

비디오 영상의 계층적 요약 정보 생성 방법

이광국*, 이상엽**, 김희율*

*한양대학교 전자통신전파공학과

**삼육대학교 경영정보학과

e-mail : nohoho@vision.hanyang.ac.kr

A method of generating summerized video in heirarchy

Gwang-Gook Lee, Sang-Yeob Lee, Whoi-Yul Kim
Divison of Elecronic Computer Engnerring, Hanyang University

요약

최근 비디오 데이터의 급격한 증가에 따라 원하는 동영상 정보를 검색하고자 할 때 막대한 시간과 노력의 낭비가 생긴다. 비디오 데이터의 효율적인 관리와 검색을 위해서는 동영상 데이터의 색인이 필요하다. 일반적으로 비디오 색인을 하기 위한 시작점은 장면전환 검출이 된다. 본 논문에서는 장면 전환을 검출한 후, 검출된 장면들 간의 변화를 참조하여 계층적 구조로 설정하여 동영상의 인덱스를 생성하는 방법을 소개한다. 컷으로 검출된 장면들을 비교하는데에는 전역적인 방법과 Bipartite 매칭을 이용한 국부적인 방법이 비교되었다. 컷 사이의 비교에 국부적인 방법을 사용함으로써 씬이 잘못 구분되는 오류를 줄일 수 있었다.

1. 서론

컴퓨터의 성능, 네트워크, 데이터 압축기술 등의 발달로 현재 멀티미디어 데이터의 양은 엄청나게 증가하고 있다. 그리고, 이러한 데이터 양의 증가와 함께 내용기반 비디오 검색, 비디오 인덱싱 등 멀티미디어 데이터를 관리, 사용하기 위한 다양한 요구 역시 증가하고 있다. 자동화된 컷 검출은 이러한 다양한 요구를 충족시키기 위한 출발점이 된다.

비디오를 구성하는 최소 단위는 한 장의 영상을 나타내는 프레임(frame)이다. 하나의 카메라에 의해서 연속적으로 얻어진 프레임의 시퀀스를 샷(shot)이라고 하며, 내용상으로 연관이 있는 샷들의 집합을 씬(scene)이라고 한다. 컷(cut)은 장면과 장면 사이의 경계에 해당하는 장면 전환점을 의미한다.

지금까지 컷 검출을 위한 다양한 방법들이 제안되었다. 이러한 방법들은 압축되지 않은 비디오 시퀀스를 위한 방법과 압축된 비디오 시퀀스 위한 방법의 두 종류로 분류될 수 있다. 압축되지 않은 비디오 시퀀스에 적용되는 방법으로는 연속된 두 프레임

입간에 픽셀의 차이 비교[1], 히스토그램의 차이 비교,[2], 움직임 벡터의 비교[3] 등이 있다. 압축된 비디오 시퀀스에 사용될 수 있는 방법은 DCT 압축 계수를 비교하는 방법, I,P,B 프레임을 이용한 방법 등이 있다[4].

컷을 자동으로 찾아낸다고 하더라도 검출해낸 컷을 비디오 인덱싱 등의 응용에 그대로 이용하기에는 무리가 따른다. 같은 샷 안에서도 여러 컷이 존재할 수 있는데, 사람이 보기에 의미를 갖는 비디오의 단위는 컷이 아닌 씬이기 때문이다. 따라서, 검출된 컷을 각 씬에 따라서 계층적, 구조적으로 저장할 필요가 생기게 된다.

본 논문에서는 히스토그램 및 에지(edge) 정보를 이용해서 컷 검출을 하고, 그 결과를 의미있는 비디오 구성단위의 씬에 따라 인덱스를 생성하였다. 씬의 검출을 위해서 검출된 컷 사이의 차이를 비교하는 방법이 사용되었는데, 컷 사이의 비교를 위해서는 프레임 전체의 전역(global) 정보를 이용하는 방법과 하나의 프레임을 여러 블록으로 나누어 국부(local) 정보를 이용하는 방법을 비교하였다. 각 블록

화된 프레임 사이의 비교를 위해서 Bipartite 매칭 (matching) 이 사용되었다.

본 논문의 구성은 다음과 같다. 2장에서는 사용된 컷 검출의 방법을 설명하고, 3장 검출된 컷으로부터 씬의 인덱스를 생성하는 방법을 설명한다. 4장에서는 실험 결과를 보이고, 5장에서는 결론 및 향후 과제에 대하여 기술하였다.

2. 컷 검출 방법

본 논문에서 사용된 컷 검출 방법은 컬러 히스토그램과 에지 정보를 이용하였다. 연속된 두 프레임 사이에 히스토그램 및 에지 정보의 변화를 관찰하여 변화 값이 주어진 임계값을 초과하는 경우 컷으로 판단했다.

컬러 히스토그램의 계산에는 HSV 컬러 모델이 사용되었다. HSV 컬러 모델은 RGB 컬러모델에 비해 색상을 표현하는 방법이 인간이 인식하는 방법과 비슷하며, 따라서 색상사이의 거리도 인간이 느끼는 것과 좀 더 비슷하다는 장점이 있다.

컬러 히스토그램에는 H, S, V 성분이 모두 사용되었으며 각 히스토그램 구간을 적절히 양자화하였다. 이는 계산량을 줄이기 위한 것이기도 하지만, 히스토그램을 양자화 할 경우 작은 변화에는 덜 민감해지는 장점이 있기 때문이기도 하다. H, S, V 성분의 양자화 레벨은 사람이 중요하게 느끼는 정도에 따라 각각 다르게 정해주었는데, 각각 18, 12, 16 레벨이 사용되었다.

연속한 두 프레임 M 와 M-1 사이의 각 컬러 히스토그램의 변화량의 총 합을 인접 프레임 사이의 거리(distance)중에서 컬러에 의한 성분, $Dist.Color$ 라고 하였으며 수식으로는 다음의 식과 같다.

$$Dist. Color = \sum_{i=1}^N |H_M(i) - H_{M-1}(i)|$$

(for each H, S, V)

$H_M(i)$: M번째 프레임의 i 번째 히스토그램 값

프레임 내에서 에지 성분을 추출하기 위해서는 소벨 연산자를 사용하였다. 소벨 연산자를 그레이 영상에 적용한 후 임계치 이상의 그레이 값을 갖는 픽셀의 숫자를 구하여 연속된 두 프레임간의 그 차이를 컷 검출을 위한 에지에 의한 성분, $Dist. Edge$ 로 사용하였다.

$$Dist. Edge = \sum_{x=0}^M \sum_{y=0}^N E_M(x, y) - \sum_{x=0}^M \sum_{y=0}^N E_{M-1}(x, y)$$

$$E(x, y) = \begin{cases} 0, & \text{for } S(x, y) < Th \\ 1, & \text{for } S(x, y) \geq Th \end{cases}$$

$S(x, y)$: Sobel 경계선 영상

연속된 두 프레임 사이의 차이는 위에서 얻은 $Dist. Color$ 와 $Dist. Edge$ 의 조합으로 구해졌으며, 이 값이 임계치 이상이 되면 장면전환이 발생한 것으로 판단했다. 임계치는 실험에 의해 적절한 값을 설정해 주었다.

3. 인덱스 생성 방법

앞에서 설명한 컷 검출의 방법으로 검출된 컷을 그대로 비디오 데이터의 인덱스로 사용하기에는 많은 무리가 따른다. 이는 하나의 씬 안에도 여러 개의 컷이 존재할 수 있기 때문이다. 이 컷들은 단지 장면의 전환일 뿐 사람이 보기에 어떤 의미를 가지고 나누어진 것은 아니다. 따라서, 사람이 보기에 의미를 가진 비디오 구성단위인 씬 단위로 인덱스를 구성하는 것이 유리하다. 그러므로 검출된 각 컷들이 같은 씬에 해당하는 것인지 아닌지를 판단하여, 검출된 컷을 씬 단위로 구조화하여 저장할 필요가 있다.

위의 과정을 수행하기 위해서, 이미 검출된 컷 가운데서 연속된 두 컷간의 차이를 계산하였다. 연속된 두 컷의 차이가 임계치 이하이면 두 컷은 서로 같은 씬에 해당하는 것으로 판단하여 인덱스에서 같은 씬의 아래에 포함시키고, 그 차이가 임계치보다 크다면, 새로운 씬이 발생한 것으로 판단하여 해당 컷을 인덱스 상에서 새로운 씬으로 구분한다. 컷간의 차이를 계산하는 데는 컷 이미지 전체의 전역 정보를 이용하는 방법과 컷 이미지를 블록화해서 국부 정보를 이용하는 두 가지 방법이 비교되었다.

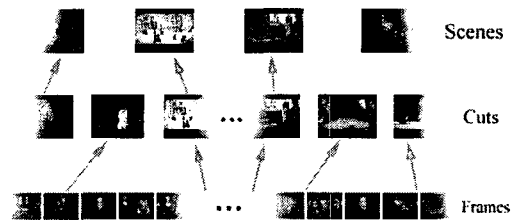


그림 1 . 인덱스의 생성

3.1 전역 정보의 사용

검출된 각 컷을 씬 별로 구분하기 위해서 우선 컷의 영상 전체를 비교하는 방법이 검토되었다. 이 방법은 뒤에서 제시된 방법과 비교하여 전역 정합이라고 부른다. 두 컷간의 차이를 계산하는데 사용된 방법은 컷 검출에 사용된 방법과 같은 방법이 사용되었다. 즉 이미 검출된 연속된 두 컷의 컬러 히스토그램 및 에지 정보를 이용하여 두 컷의 차이를 계산하였고, 이 값이 임계치 이상이면 두 컷이 서로 다른 씬에 포함된다고 결정하였다.

3.2 국부 정보의 사용

컷 간의 비교에 국부 정보를 사용하기 위하여 우선 비교 대상인 두 컷을 3x3 의 구획으로 나누었고, 각 블록에 대하여 컷 검출에 사용되었던 것과 같은 특징값을 추출하였다.

이 특징값의 비교를 위해서는 Bipartite 매칭(matching)을 사용하였다.[5,6] Bipartite 매칭은 Bipartite 그래프(graph)에서 각 노드(node)들을 최적으로 매칭하는 조합을 찾아준다

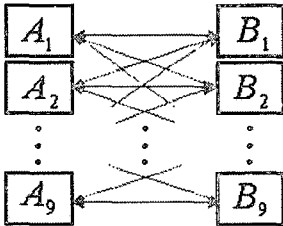


그림 2 . 이미지 블록으로부터 Bipartite 그래프 생성

비교 대상인 두 컷의 각 블록별 히스토그램과 에지 성분을 추출한 후 각 블록끼리의 거리를 계산하여, 그 거리를 가중치(weight)로 갖는 Bipartite 그래프를 구성한다. 이렇게 구성된 그래프가 그림 2 에 나타나 있다. 그림에서 A_i 는 컷 A 를 3x3 으로 블록화했을 때 i 번째 블록을 나타낸다. 이 그래프에 Bipartite 매칭을 이용하면 가능한 모든 조합 중에서 두 컷 사이의 거리가 최소가 되는 조합을 얻을 수 있으며, 이 조합을 통한 두 컷 사이의 거리를 씬 검출에 사용하였다.

4. 실험 결과

실험에 사용한 비디오 영상은 4개의 MPEG 영화

에서 임의의 지점에서부터 5000 프레임 길이의 클립을 추출한 것이다. 사용한 실험영상에 대한 정보는 표 1에 나타나 있다. 각 실험영상에 대하여 두 가지 방법 모두를 사용하여 인덱스를 생성한 후에 그 결과를 사람이 인지하는 결과와 비교하여 잘못 인식된 씬의 숫자를 기록하였다. 실험 결과는 표 2 에 나타나 있다.

실험결과 표에서 Miss 는 같은 씬에 포함된 두 컷을 다른 씬으로 구분한 경우이며, False 는 다른 씬에 포함된 두 씬을 같은 씬으로 처리한 경우이다. 실험 결과 Bipartite 매칭을 이용하여 국부 정보를 이용한 방법이 전역 정보만 이용한 방법보다 전반적으로 더 나은 성능을 보임을 알 수 있다.

실험영상	총 프레임 수	검출된 컷 수
1	5000	70
2	5000	107
3	5000	54
4	5000	64

표 1. 실험에 사용된 동영상 데이터

실험영상	방법 1			방법 2		
	Miss	False	정확도	Miss	False	정확도
1	15	3	74%	8	4	83%
2	16	0	85%	6	0	93%
3	6	2	85%	2	4	89%
4	4	6	84%	5	4	86%

표 2. 실험 결과



그림 3. 같은 씬에 포함된 연속된 두 컷

그림 3 은 같은 씬에 포함된 연속된 두 컷이다. 그림은 두 컷은 방법 1 에 의해서 다른 씬으로 인식되었으나, 방법 2 에 의해서는 같은 씬으로 인식되었다. 그 이유는 두 개의 컷에서 거리차의 값을 크게 하는 인자가 특정 블록에 설정되어 있기 때문이다. Bipartite 매칭은 변화가 발생하는 영상에서 특정 영역의 변화가 클 뿐 다른 영역에서의 변화가 그

렇게 크지 않는 형태의 정합에서 특정 영역의 변화를 무시하거나 최소화 할 수 있다. 따라서 같은 것에서 새로운 결과의 삽입에 의해서 발생하는 거리차에 강인하게 대응하게 된다.

5. 결론 및 향후과제

본 논문에서는 동영상의 인덱스를 생성하기 위해 검출된 컷을 통해 씬을 찾아내는 방법을 보였으며, 씬의 검출에 있어서 전역적인 방법과 국부적인 방법을 비교하였다. Bipartite 매칭을 이용한 국부적인 방법을 통해 씬 검출의 효율을 좀 더 높일 수 있음을 보였다.

본 논문에서는 씬 검출을 위해 고정된 임계점을 이용하였는데, 임계점의 수준에 따라 실험의 결과가 큰 차이를 보였다. 향후 적정 임계치를 자동으로 설정할 수 있는 방법 등에 대한 연구가 이루어져야 할 것으로 보인다.

* 본 연구는 교육부 두뇌한국21(BK21) 정보기술사업단의 지원에 의한 것입니다.

6. 참고문헌

- [1] H. J. Zhang, A. Kankanhalli and S. W. Smoliar, "Automatic partitioning of full-motion video," *Multimedia Systems*, VOL.1, No.1, pp.10-28, 1993
- [2] A. Nagasaka, Y. Tanaka, "Automatic Video Indexing and Full-Video Search for Object Appearances," *Visual Database System II*. Eds.E, Knuth and L.M. Wegner. Elsevier Science Publishers B.B., IFIP. pp.113-127, 1992
- E. Ardizzone, M. L. Cascia, "Automatic Video Database Indexing and Retrieval", *Multimedia Tools and Applications*, Vol.4, No.1, pp.29-56, January. 1997
- [3] H. Ueda, T. Miyatake, and S. Yoshizawa, "Impact : An Interactive Natural-motion-picture Dedicated Multimedia Authorizing System," *In Proc. of Human Factors in Computing Systems(CHI91)*, pp.343-354, New Orleans, Louisiana, 1991
- [4] Boon-Lock Yeo and Bede Liu, "Rapid Scene Analysis on Compressed Video," *IEEE Trans. on CIRC UTS and SYSTEM for VIDEO TECH.*

vol.5 No.6 December 1995

[5] Whoi-Yul Kim and C. Kak, "3-D Object Recognition Using Bipartite Matching Embedded in Discrete Relaxation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 3, pp. 224-251, March. 1991

[6] 이상엽, 김희율, "Bipartite Matching를 이용한 캐릭터 영상 검색," 제13회 영상처리 및 이해에 관한 워크샵 발표 논문집 pp489-573, 2001.1