

MPEG-7의 충실도 기술법에 의한 뉴스 요약 기법

김정림, 전승수, 오석진, 설상훈

고려대학교 전자공학과

Scalable Hierarchical Summary of News using Fidelity in MPEG-7 Description Scheme

Jung-Rim Kim, Seong Soo Chun, Seok-jin Oh, Sanghoon Sull

School of Electrical Engineering, Korea Univ

요 약

인터넷이 빠르게 보급됨에 따라 많은 멀티미디어 콘텐츠를 접할 수 있게 되었다. 이에 따라 사용자가 쉽게 멀티미디어 콘텐츠에 접근하게 하기 위해 여러 가지 방법으로 이를 재구성하여 데이터베이스를 만드는 연구가 활발하게 진행되고 있다. 이러한 추세에 발맞추어 MPEG-7 표준이 진행 중에 있고, 사용자가 대용량의 멀티미디어 콘텐츠에 대한 접근을 용이하게 하기 위해 멀티미디어 요약과 같은 기법들이 표준화되고 있다. 본 논문에서는 MPEG-7에 제안된 충실도(fidelity) 기술법(description scheme)을 이용하여 뉴스를 효과적으로 요약할 수 있는 방법에 대해 제안한다.

I. 서 론

최근, 네트워크의 전송속도 및 대역폭이 증가함에 따라 많은 양의 멀티미디어를 인터넷을 통해 쉽게 접할 수 있게 되었다. 이와 같이 멀티미디어가 폭발적으로 증가함에 따라 자료의 저장이나 검색(retrieval), 탐색(browsing)등을 용이하게 할 수 없는 문제점을 내포하게 되었다. 이러한 문제점을 해결하기 위해 탐색이나 요약을 위한 효율적인 색인 기법들이 등장하기에 이르렀다.

최근에 이러한 흐름에 발맞추어 MPEG-7 표준안이 만들어지고 있으며, 특히 멀티미디어의 검색, 탐색 및 요약 등의 멀티미디어 기술법(Multimedia Description Scheme : MDS)에 대해 집중 논의되고 있다.

본 논문에서는 MPEG-7 MDS에서 사용자의 요청에 맞도록 비디오를 요약할 수 있도록 기술된 충실도(fidelity) 개념에 대해 알아보고, 이를 이용하여 뉴스에 맞는 충실도 개념을 제안하고자 한다.

본 논문의 구성은 다음과 같다. 제 II장에서는 관련 연구로 MPEG-7에서 논의되고 있는 것들과 충실도의 개념에 대해 알아보고 제 III장에서 본 논문에서 제안하는 뉴스에 알맞은 충실도를 계산하는 알고리즘 및 요약 기법을

소개하고 그 결과를 제 IV장에서 기술한다. 마지막으로 제 V장에서 최종적인 결론 및 향후 연구과제에 대해 기술한다.

II. 관련연구 및 충실도

1. 관련연구

비디오 요약에 관련된 연구는 최근에 연구되기 시작한 분야이다. 긴 시간의 비디오로부터 몇 개의 대표되는 정지 영상으로 비디오를 표현하는 것은 사용자가 그 비디오 전부를 보지 않고 내용을 쉽게 파악할 수 있도록 하는데 많은 도움을 준다. 대용량의 비디오를 낮은 주파수 대역폭에서 시청하기도 쉽지 않기 때문에 이러한 요약기법은 반드시 필요하게 되었다.

MPEG-7에 제안된 멀티미디어 콘텐츠 기술(Content Description)은 그 방법에 따라 크게 두 가지로 나눌 수 있다. 색상, 질감(texture), 소리 등의 시공간적 특징들로 구성된 구조적 기술(Description of the content structural aspects)과 사람의 인지적인 능력을 반영하는 개념적 기술(Description of the content conceptual aspects)로 나눌

수 있다[1]. 그리고 이러한 기술을 바탕으로 탐색 및 접근 (Navigation and Access)이 가능한 기술법을 제안하고 있다. MPEG-7에서 제안하고 있는 여러 탐색 및 접근 기술법 중에 사용자의 요청에 맞도록 요약이 가능한 충실도 개념의 요약 기법에 대해 알아본다.

2. 충실도(Fidelity)

충실도는 목구조의(tree-structured) 대표 프레임(key frame) 계층에서 상위 대표 프레임이 하위 대표 프레임들을 어떠한 정도로 표현하고 있는가를 나타내는 개념이다.[2, 3, 4]

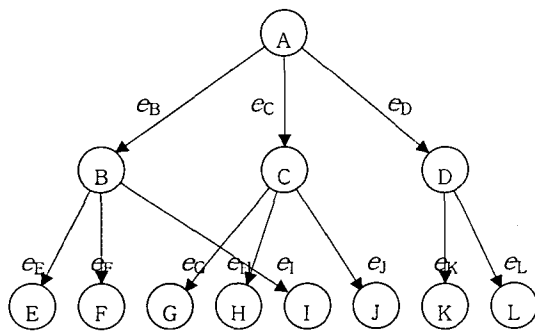


그림 1. 대표 프레임 계층에서의 충실도 예제 (A, B, C 등은 대표 프레임을 나타내고, e_B, e_C 등은 각 대표 프레임들의 충실도를 나타낸다.)

대표 프레임들을 그림 1과 같이 계층적인 구조로 표현할 수 있고 그 상위 대표 프레임과 하위 대표 프레임들간의 상관관계를 나타낼 때 충실도를 사용하여 표현하게 된다. [2]에서 제안된 충실도는 그 상위의 대표 프레임들간의 공간적 특징의 차이를 이용하여 계산하였다. 즉 어떤 대표 프레임 S 의 충실도(e_S)는 다음과 같은 식으로 표현하게 된다.

$$e_S = 1 - \max_{S_i \in S} (d(P, S_i)) \quad (1)$$

여기에서 $d()$ 는 $[0, 1]$ 로 정규화된 거리 함수를 나타내고 P 는 상위 대표 프레임, S_i 는 S 의 하위 대표프레임들을 나타낸다.

III. 충실도를 이용한 뉴스의 요약

이 장에서는 뉴스라는 특별한 장르에 대해 충실도를 구하는 방법과 그를 이용한 요약에 대해 기술한다.

1. 뉴스의 요약을 위한 계층화 알고리즘

[2]에서 제안된 충실도 계산은 단순히 대표 프레임들간의 공간적 특징을 사용하여 충실도가 식(1)과 같이 계산되었다. 이러한 계산법은 의미론적인 개념이 약하게 반영되어, 때에 따라 어떠한 비디오들에 대해서는 요약 결과가 좋지 못하게 나타나는 경우가 발생할 가능성이 있다.

특히 뉴스라는 특별한 경우는 그 구조가 명확하고, 시간에 따른 중요도가 비례하는 경우가 많다. 즉 뉴스는 앵커 장면(Anchor Shot)이라는 특별한 장면과 그 장면을 시작으로 해당 기사가 방송되는 구조를 갖고 있으며, 그 날 중요한 사건이 대체로 시간의 순서에 따라 배치되는 경우가 많다. 따라서 이러한 앵커 장면과 시간을 고려한 요약을 함으로써 보다 사용자에게 내용이나 의미 전달을 쉽게 할 수 있게 된다. 따라서 (1)의 식을 아래와 같이 시간에 대한 충실도($e(t)$)를 같이 고려하여 다음과 같이 표현한다.

$$e_S' = we_S + (1-w)e(t) \quad (2)$$

여기에서 w 는 공간적 충실도에 대한 가중치를 나타내며, $0 \leq w \leq 1$ 의 범위에 있다.

또한 [2]에서는 시간적인 순서를 고려하지 않은 군집화(clustering)로 인해 상위 대표 프레임과 하위 대표 프레임간의 시간적인 상관관계가 무시되는 경우도 발생한다. 예를 들면 그림 1에서 최하위 계층의 대표 프레임이 {E, F}, {G, H, I, J}, {K, L}과 같이 3개의 장면(shot)으로 구성되어 있다면 상위 계층의 대표 프레임 B가 I라는 시간적으로 멀리 떨어진 대표 프레임을 대표할 수 있게 된다. 그러나 뉴스라는 특별한 구조에 적용하게 될 경우 앵커 장면이 모두 하나의 대표 프레임으로 대표되는 경우가 발생하여 의미 전달을 위한 효과적인 요약이 어렵게 된다. 우선 앵커 장면들을 검출하고 그 사이에 들어가는 사건들을 요약하는 것이 보다 의미 전달을 위한 효과적이라 볼 수 있다. 따라서 다음과 같은 방향식 계층화 알고리즘을 이용하여 뉴스를 계층화 시켰을 경우 효과적인 결과를 얻을 수 있다.

1. 먼저 장면 전환들을 먼저 검출한다.
2. 각각의 장면에서 공간적인 특징 벡터를 이용하여 몇 개의 대표 프레임을 추출한다.
3. 각 장면이 앵커 장면인지 아닌 지를 검출한다.
4. 앵커 장면일 경우 각각을 상위 계층으로 올리고 앵커 장면이 아닐 경우(이를 사건 장면이라 가정한다.) 연속된 몇 개의 장면을 군집화 하면서 상위 계층으로 올린다.
5. 원하는 계층이 완성될 때까지 과정 4를 반복한다.

이 때 연속된 장면의 군집화 과정은 [5]에서 제안한 알고리즘을 사용하며, 보다 효과적인 결과를 얻기 위해서 마지막 계층화 과정에서 앵커 장면을 제외하면 사건들만을 모은 결과를 얻을 수 있다. 위의 알고리즘을 도식화 한 것

이 그림 2이다.

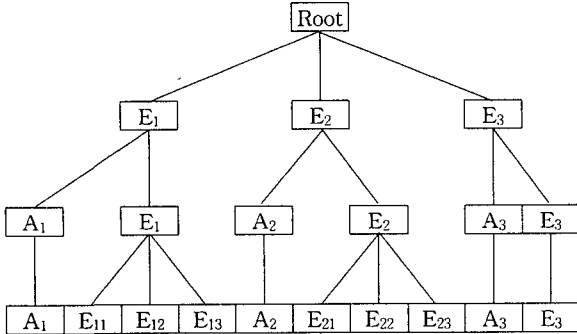


그림 2. 계층화의 예제 (An : 앵커 장면, En : 사건 장면)

2. 요약 기법

충실도를 사용한 계층적인 요약기법은 [2]에서 제안된 다음과 같은 알고리즘을 사용한다.

우선 처음에 루트 노드를 요약 집합 K 에 넣고 집합 K 에 있는 노드 α 와 이에 연결된 집합 K 에 속해 있지 않은 노드 β 중에서 가장 작은 충실도를 갖는 노드 β 를 집합 K 에 넣고 원하는 요약 개수가 될 때까지 반복한다.

```

add root_node to K;
while ( card(K) < N ) {
  let <math>\alpha, \beta</math> be a minimum fidelity
  such that <math>\alpha \in K</math> and <math>\beta \in K^c</math>;
  add <math>\beta</math> to K;
}
    
```

IV. 실험결과

본 장에서는 위에서 제안한 알고리즘을 사용하여 목구조의 대표 프레임 계층을 생성하고 이를 이용한 뉴스의 요약에 대해 기술한다.

1. 대표 프레임 계층화

본 실험에서 대표 프레임 추출, 앵커 장면 검출 및 군집화에 사용된 공간적 특징 벡터는 비디오의 DC 영상에서의 luminance projection을 사용하였다. [6]에서 제안한 $M \times N$ DC 영상 f 의 n 번째 줄과 m 번째 열의 luminance projection(l_n^r, l_m^c)은 각각 다음의 식 (3)과 같이 정의된다.

$$l_n^r(f) = \sum_{m=1}^M Lum\{f(m, n)\} \quad (3)$$

$$l_m^c(f) = \sum_{n=1}^N Lum\{f(m, n)\}$$

이러한 공간적 특징 벡터를 사용하여 추출된 대표 프

임들을 총 4단계로 계층화하였고, 표 1과 같은 비디오 자료를 사용하였다.

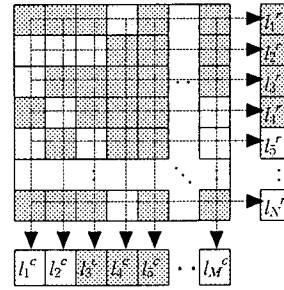


그림 3. Luminance Projection

표 1. 실험에 사용된 비디오 자료

비디오	길이	대표 프레임 개수
News 2000년 7월 5일자	27분 42초	271
News 2000년 7월 6일자	27분 42초	254
News 2000년 7월 7일자	27분 43초	338
News 2000년 7월 8일자	27분 33초	283

그리고 앵커 장면을 검출하기 위해 다음과 같은 식을 사용하였다.

$$A = \{f_k \mid d(f_r, f_k) \leq \epsilon, f_k \in R\} \quad (4)$$

f_r 은 사용자가 지정한 참조 앵커 장면이고, ϵ 은 앵커 장면의 추출을 위한 사용자 임계치이며, R 은 대표 프레임 집합을 의미한다.

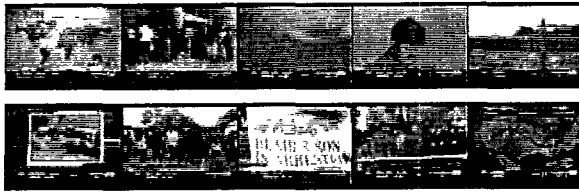
또 각각의 앵커 장면의 공간적인 충실도 값이 항상 1이 되므로 이러한 충실도는 의미가 없어진다. 따라서 공간적 충실도의 가중치를 0으로 두어 시간적인 충실도만 사용하게 하였고, 사건 장면들은 공간적인 특징을 사용한 군집화를 통해 상위 계층으로 올라가나 그 시간적인 중요성에 더 많은 가중치를 부여하여 공간적 충실도 가중치를 $w < 0.5$ 로 하였다.

2. 요약 실험

위의 대표 프레임 계층을 가지고 제 III장 2절에서 설명한 알고리즘을 이용하여 News 2000년 7월 6일자와 7일자에 대해 각각 10개의 프레임과 20개 프레임에 대해 요약해 보았다. 그림 4는 각각의 결과를 나타낸 것이다.

그림 4에서의 결과와 같이 10개의 프레임으로 요약했을 경우 중요 사건들만 기술되는 프레임으로 나타나며, 계속 증가했을 경우 그 사이에 앵커 장면들이 들어오게 되는

결과를 알 수 있다.



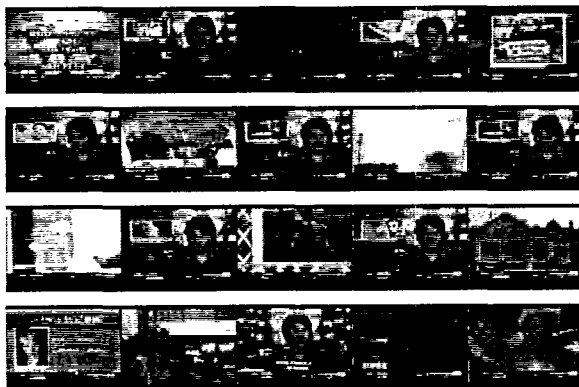
(a) News 7월 6일자 10개 프레임으로 요약



(b) News 7월 6일자 20개 프레임으로 요약



(c) News 7월 7일자 10개 프레임으로 요약



(d) News 7월 7일자 20개 프레임으로 요약
그림 4. News 요약 결과

V. 결론 및 향후과제

본 논문은 뉴스라는 구조적으로 잘 정의된 장르를 가지고 충실도를 사용한 요약 기법에 대해 제안하고 그 기법을 이용한 실험 결과를 도출하였다. 앵커 장면과 사건 장면들을 다른 방법으로 각기 계층화를 시키고, 시간이라는 개념을 넣어 충실도를 계산하는 방법에 대해 제안하였다.

그리고 이 요약 결과를 이용하여 각 대표 프레임에 문자 주석(text annotation)을 첨가하였을 경우, 보다 효율적인 의미 전달을 할 수 있는 방법으로 발전시킬 수 있다.

본 연구에서 참조가 되는 앵커 장면을 검출하는 방법에 대해서 기술하지 않았다. 현재 그러한 장면은 제작자가 직접 넣어주는 수준에서 이루어지고 있으나, 앞으로 연구를 통해 앵커 장면을 자동으로 검출할 수 있도록 할 것이다. 또한 음성 인식이나 문자 인식을 통한 문자 주석 첨가를 향후 과제로 남긴다.

참고문헌

- [1] "Overview of the MPEG-7 Standard," *ISO/IEC JTC1/SC29/WG11 N4031*, Singapore, March 2001.
- [2] Sanghoon Sull, Jung-Rim Kim, Yunam Kim, Hyun Sung Chang, and Sang Uk Lee, "Scalable Hierarchical Video Summary and Search," *Proc. of SPIE 2001*, Vol. 4315 Storage and Retrieval for Media Database 2001, pp553-561, Feb. 2001
- [3] Sanghoon Sull, Jung-Rim Kim, and Yunam Kim, "Efficient and effective search and browsing using fidelity," *ISO/IECC/JTC1 SC29/WG11 M5101*, 54th MPEG Meeting, La Baule, Oct. 1999.
- [4] Sanghoon Sull, Jung-Rim Kim, and Yunam Kim, "Improved notion of the fidelity for efficient browsing," *ISO/IEC JTC1/SC29/SG11 M5442*, 55th MPEG meeting, Maui, Dec. 1999.
- [5] Minerva M. Yeung and Bede Liu, "Efficient matching and clustering of video shots," *Proc. of IEEE International on Image Processing '95*, vol. 1, pp. 338-341, Oct. 1995.
- [6] Hyung Sung Chang, Sanghoon Sull, and Sang Uk Lee, "Efficient Video Indexing Scheme for Content-based Retrieval," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, No. 8, pp. 1269-1279, Dec. 1999.