

# 실시간 광고 색인 및 검출

구영모, 김정림, 설상훈

고려대학교 전자컴퓨터공학과

ymgu@mpeg.korea.ac.kr, jrkim@mpeg.korea.ac.kr sull@mpeg.korea.ac.kr

## Real Time Advertisement Indexing & Detection

YoungMo Goo, Jung-Rim Kim, Sanghoon Sull

Department of Electronics and Computer Engineering, Korea Univ

### 요 약

최근 디지털 방송이 시작됨에 따라 방송 채널이 다변화하고 그 수도 많아지고 있다. 현재 방송되고 있는 디지털 방송에 지상파 방송과 케이블 TV의 방송까지 고려한다면 시청자가 접할 수 있는 방송은 다양하다. 기업들은 주요 홍보 매체인 방송을 통해 자신들의 상품을 광고로 내보낸다. 또한 자신들의 광고에 대한 모니터링을 통해 광고가 계약한 시간, 횟수, 광고의 길이에 맞게 방영되는지 알고 싶어 한다. 현재 이러한 모니터링은 사람에 의한 수작업으로 이루어지고 있으며, 여러 채널의 방송에 광고를 내보냈을 때에는 이러한 모니터링은 더욱 어려워진다. 본 논문에서는 디지털 방송뿐만 아니라 아날로그 방송에 대해서까지도, 현재 방송의 비주얼 리듬을 실시간으로 추출하고 이를 이용하여 장면 전환을 수행하고 특징 벡터를 추출해서 얻어진 정보들을 이용한 실시간 광고 검출 알고리즘을 제안한다.

### I. 서 론

최근까지는 방송이라고 하면 지상파 방송이 주가 되고 케이블 TV 정도가 우리가 접할 수 있는 방송의 전부였다. 하지만 디지털 방송이 시작되면서 수많은 채널들이 생겨나고 채널들마다 하루하루 수많은 프로그램들을 내보내고 있다. 현재 SkyLife는 디지털위성방송을 시작하였으며 140여 개의 채널을 가지고 있다. 이러한 가운데 광고는 그 프로그램들 사이사이에 항상 존재한다. 이러한 광고를 만들어 방송에 내보내는 광고주에게 자신의 광고 방송에 대한 정보를 제공할 수 있고, 광고에 관심 있는 사람에게 하루에 방송되는 광고에 대한 정보를 제공할 수 있는 응용 프로그램이 존재한다면 이들은 인력을 들이지 않고 손쉽게 원하는 정보를 제공 받을 수 있을 것이다.

광고 검출 방법으로 광고 내의 소리의 에너지 변화 특성을 이용한 방법이 있다[1]. 이 방법은 소리의 에너지 변화는 각각의 동영상마다 각자의 특성을 가지고 있고, 특히 광고의 경우에는 사람들의 주목을 끌기 위해 그 소리의 변화도 크다는 점을 이용하여 소리 에너지 변화의 특징을 이용하였을 때, 그 광고의 검출이 더 정확하다는 것을 보

이고 있다. 하지만 광고의 경우 모두 동일한 길이가 아니다. 짧은 것은 15초에서 긴 것은 30초까지 그 길이가 다양하기 때문에 이러한 1:1 matching의 방법으로는 질의와 길이가 다른 광고는 검출하기가 힘들뿐만 아니라 단순히 소리 에너지 하나만으로 동영상들을 특징짓지 못한다.

대부분의 동영상 검색은 대표 프레임을 이용한다[2-4]. 대표 프레임을 이용한 동영상 검색은 검색 대상이 많아 실시간 검색에 적용하기 어렵다. 또한 공간적 정보만을 이용하여 검색하므로 특징 벡터의 오차에 의한 애러가 존재할 수 있다.

다른 동영상 검색 방법으로는 시공간적 정보를 모두 사용하기 위해 MPEG로 압축된 동영상의 모든 프레임에서 DC 값만을 특징 벡터로 사용하는 방법이 있고[5], 시각적 정보와 청각적 정보를 이용하는 방법도 있다[6].

본 논문에서는 시공간적 정보를 모두 이용하면서도 빠른 검색이 가능할 수 있도록 비주얼 리듬을 이용한 동영상 검색 방법을 사용한다.

본 논문에서는 비주얼 리듬 정보와 장면 전환 검출을 이용하여 각각의 장면의 특징 벡터를 구하여 특징 벡터로 이루어진 광고 데이터베이스를 구축한 후, 실시간으로 방

송되고 있는 스트림에서의 연속적인 장면들과 데이터베이스 내의 광고의 연속된 장면과의 유사도를 계산하여 스트림이 어느 광고에 해당하는지 아니면 광고가 아닌지를 알아내는 알고리즘을 제안하고자 한다. 시뮬레이션의 간략화를 위해 방송 스트림은 녹화된 방송을 사용한다.

본 논문의 구성은 다음과 같다. 제 II장에서는 검색 시스템과 색인하고 검색하는 일련의 과정에 대해 설명하고 III장에서는 유사 장면들로 선택된 장면들이 데이터베이스 내의 어떠한 광고의 장면들인지 아니면 광고가 아닌지에 대해 결정하는 검출 알고리즘을 제안한다. 제 IV장에서는 실시간으로 일련의 과정이 이루어질 수 있는지에 대한 검출 속도와 정확도에 대한 실험을 하고 그 결과를 기술한다. 마지막으로 V장에서는 실험 결과를 토대로 이 알고리즘의 효율성과 단점에 대해 논하고 향후 연구과제에 대해 기술한다.

## II. 검색 시스템

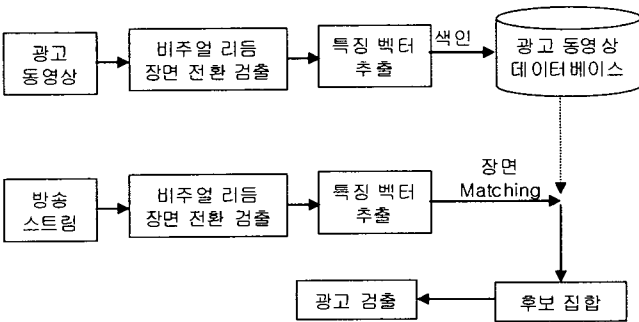


그림 1. 검색 시스템

그림 1의 검색 시스템 구성은 다음과 같다. 광고들의 비주얼 리듬 정보를 이용해 장면 전환 검출을 하고, 각 장면마다 양자화된 색상 정보를 특징 벡터로 사용하여 광고 동영상들의 특징 벡터들을 저장해 놓은 데이터베이스를 구축한다. 방송의 스트림을 계속 저장하면서 실시간으로 비주얼 리듬을 이용한 장면 전환 검출을 통해 특징 벡터를 구하고, 데이터베이스와의 장면 Matching과 새로운 알고리즘을 이용하여 방송 스트림이 광고인지 여부와 어떤 광고인지에 대한 결론을 내린다.

### 1. 비주얼 리듬 ( Visual Rhythm )

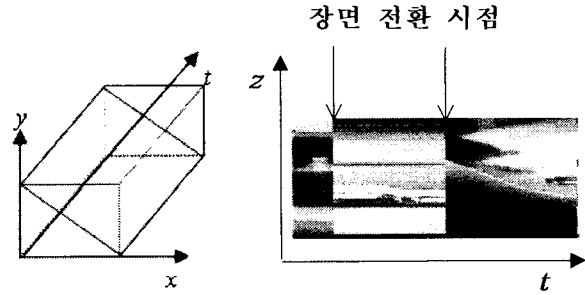
동영상의 시공간적 특징을 가지고 있는 정보로 동영상  $V$ 의 프레임  $t$ 에서, 좌표  $(x, y)$ 의 픽셀 값을  $f_i(x, y, t)$ 라고 하면, 비주얼 리듬,  $VR$ 은 아래와 같이 정의된다[7].

$$VR = \{f_{VR}(z, t)\} = \{f_{VR}(x(z), y(z), t)\} \quad (1)$$

여기에서,  $x(z)$ 와  $y(z)$ 는  $z$ 에 독립적인 1차 방정식이다.

비주얼 리듬은 각 프레임에서 대각선 방향의 픽셀들만 샘플링한 정보로 그림 2.(a)와 같고, 동영상에서 비주얼 리

듬을 추출한 모양은 그림 2.(b)와 같다.



(a) 대각선 픽셀 샘플링 (b) 비주얼 리듬  
그림 2. 비주얼 리듬의 구성

### 2. 장면 전환 검출 ( Shot Detection )

그림 2.(b)에서 비주얼 리듬의 불연속적인 부분에서 장면의 전환이 이루어진다[8]. 즉, 색상의 급격한 변화는 내용이 많이 변했다는 것을 의미하므로 하나의 장면으로 인정될 수 있다. 본 논문에서의 장면 전환은 색인과 검색에서 같은 방식으로만 이루어지면 검색에는 문제가 없기 때문에 간단한 수식을 통해 계산의 복잡도를 줄였다. 장면 전환 검출 방법은 다음과 같다.

$$a * |f_{VR,R}(z, t) - f_{VR,R}(z, t-1)| + b * |f_{VR,G}(z, t) - f_{VR,G}(z, t-1)| + c * |f_{VR,B}(z, t) - f_{VR,B}(z, t-1)| > \epsilon \quad (2)$$

여기에서  $f_{VR,R}$ ,  $f_{VR,G}$ ,  $f_{VR,B}$ 는 각각 Red, Green, Blue의 비주얼 리듬 값을 의미한다. 비주얼 리듬의 R, G, B의 색상 정보의 거리값이 어떤 임계치  $\epsilon$ 을 넘으면 장면 전환이 이루어졌다고 판단한다. 임계치  $\epsilon$ 과 Scaling Factor인  $a$ ,  $b$ ,  $c$ 는 실험에 의해 결정되어진다.

### 3. 특징 벡터 추출 ( Feature Vector Extraction )

특징 벡터는 한 장면 내의 비주얼 리듬들의 색상 정보를 이용하여 추출된다. 색상 정보를 표현하는 방법 중, 자주 사용되는 방법은 히스토그램이다. Red, Green, Blue의 세 가지 색깔을 모두 사용하게 되면 각각 8bit씩 24bit의 히스토그램이 필요하게 된다. 이런 경우 너무 많은 정보를 특징 벡터로 사용하게 되어 검색의 속도가 느려지게 된다. 특징 벡터의 정보를 줄이기 위해 R, G, B의 정보 중 중요한 정보를 담고 있는 3bit씩만 색상 정보로 사용하여 히스토그램의 bin으로 사용하면 512개의 bin만 사용하게 된다. 이 히스토그램의 bin을 특징 벡터로 사용하게 되면 특징 벡터의 정보량도 줄어들고, 색상의 중요 정보도 보존할 수 있다.

## III. 검색 알고리즘

### 1. 장면의 검색 및 후보 집합 구성

광고의 실시간 검출을 위해서는 방송 스트림이 들어왔

을 때, 즉시 처리되어 검색 결과를 알려주어야 한다. 이를 위해 본 논문에서는 방송 스트림을 장면 단위로 나누어진 장면이 만들어지면 검색을 수행하는 방식으로 실시간 검색을 구현한다. 방송 스트림을 받아 비주어리듬을 생성하면서 장면 전환 시점에서 하나의 장면을 구성하여 특징 벡터를 추출하게 된다. 추출된 특징 벡터를 광고 동영상 데이터베이스 내의 특징 벡터들과 비교하여 후보 집합을 구성한다. 이때, 유사도 검색은 Normalized Histogram 간의 유클리디안 거리 값들로 이루어지며, 방송 스트림의 장면  $q$ 와 데이터베이스의 장면  $i$ 와의 거리 값  $d(q, i)$ 는 다음과 같이 정의 된다.

$$d(q, i) = \sqrt{\sum_{m=0}^M [H_q(m) - H_i(m)]^2} \quad (3)$$

여기에서  $M$ 은 히스토그램의 전체 bin수를 나타내고,  $H_q(m)$ 과  $H_i(m)$ 은 각각  $q$ 와  $i$ 의  $m$ 번째 히스토그램의 값을 의미한다.

광고 데이터베이스의  $i$  번째 광고의 장면들로 이루어진 집합  $W_i$ 는 다음과 같이 정의한다.

$$W_i = \{w_{i1}, w_{i2}, w_{i3}, \dots\} \quad (4)$$

후보 집합은 거리 값이 가장 작은 데이터베이스의 광고 장면들로 이루어진다.  $k$  번째 방송 장면  $b_k$ 의 후보 집합  $C_k$ 는 다음과 같이 정의할 수 있다.

$$C_k = \{w_{ij}^k | d(b_k, w_{ij}^k) \leq T, w_{ij}^k \in W_j\} \quad (5)$$

여기에서  $w_{ij}^k$ 는  $i$  번째 광고의  $j$  번째 장면을 나타내고,  $T$ 는 검색 임계치를 나타낸다. false alarm을 고려하여  $C_k$ 의 원소 개수는 5개를 넘지 않도록 하였다. 또한 가장 작은 거리 값만을 가진 것을 선택하여 검색하지 않는 이유는 밝기가 다르거나 광고의 길이가 다른 경우 유사 장면이 검색되지 않는 것을 방지하기 위해서이다.

## 2. 광고 검출 알고리즘

광고의 검출을 위해서는 후보들의 연속성과 장면 배열의 순서가 가장 중요한 요소이다. 후보들의 연속성이 있다고 할지라도 이들의 배열 순서가 올바르지 않다면 이것은 false alarm으로 볼 수 있다. 이 두 가지를 모두 고려하기 위하여 우선 광고마다 하나씩의 큐(Queue)  $S_i$ 를 만들어 두고, 방송 스트림의 장면이 들어오면 후보 집합을 구하여 후보들을 식 (6)과 같이 각각에 맞는 큐( $S_i$ )에 순서대로 채워 넣는다.

$$S_i = \{v_1, v_2, \dots, v_n \in W_i\} \quad (6)$$

이때,  $S_i$  내의 원소들은 연속성을 가져야하는데 이는 이전에 들어온 원소와 현재 들어온 원소의  $k$  값에 따라 결정된다.

$$v_{l-1} = w_{ij}^k \in S_i, \quad v_l = w_{i'j'}^{k+1} \in S_i, \quad k' < k + \lambda \quad (7)$$

$\lambda$ 는 장면 검출 예러와 편집된 광고를 고려하기 위해 정한 값이다. 이를 만족시키지 못하면 이전 원소들은 큐에서 지운다. 간단한 개념도를 그리면 그림 3과 같다.

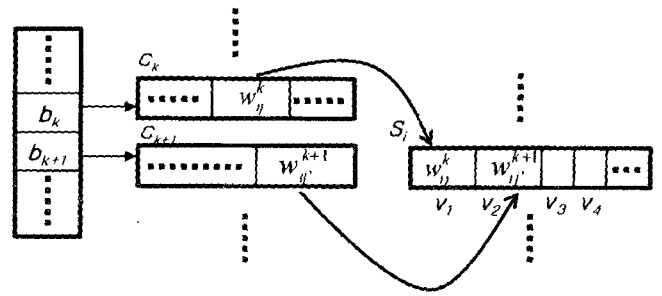


그림 3. 광고 검출 개념도

$S_i$ 의 크기( $n$ )는  $i$  번째 광고 장면의 총 개수( $N_i$ )에 따라 정해지며 이를 넘지 못한다. 큐  $S_i$ 가 모두 채워지면 원소들의 배열 순서가 올바른지 판단하여야 한다. 이 과정은 선택되어진 큐  $S_i$ 가 false alarm인지를 확인하는 과정으로 광고 검출의 정확도를 높여준다. 이는 선택된 큐  $S_i$  내의 후보들의 방송 장면 색인 번호  $k$ 와 광고 세그먼트 색인 번호  $j$ 에 의해 결정되며 방법은 아래 식과 같다.

$$d_{idx}(v_l) = k - j, \text{ for } v_l = w_{ij}^k \in S_i \quad (8)$$

$$\sum_{l=1}^n |d_{idx}(v_l) - d_{idx}(v_{l-1})| < T_{idx}$$

여기에서  $T_{idx}$ 는 큐  $S_i$  원소들의 배열 순서가 올바르지 않은 임계치이다.

이 두 조건을 모두 만족하는  $S_i$ 가 광고로 검출되며, 어떤 광고인지도 결정되어지게 된다. 방송 스트림이 들어올 때마다 이러한 일련의 과정을 거쳐 광고를 검출하게 된다.

## IV. 실험결과

이 실험에서는 1주일간의 방송을 녹화하여 방송 중에 있는 광고를 검출하였다. 이 실험에 사용된 데이터는 총 1800분의 방송 분량이며, 데이터베이스를 구성하고 있는 광고의 개수는 110개를 사용하였다. 실험을 위해 사용된 PC는 PIII-1GHz, 1G RAM 이다.

실시간 광고 검출이 가능한지 알아보기 위해 방송을 색인하는데 걸리는 시간과 검출하는데 걸리는 시간을 조사하였다. 색인 시간과 검출 시간은 모든 방송(방송은 60분 간격으로 잘라서 사용)을 색인, 검출하는데 걸린 시간을 평균하여 구한 것이다. 실험 결과는 표 1과 같다.

표 1. 색인 및 검출 시간

	색인 및 검출 시간(초)
3600초 분량의 방송	2007

표 1에서의 결과를 보면, 3600초 분량의 방송을 색인하고 검출하는데 걸리는 시간은 2007초로 실제 분량의 반 정도의 시간 밖에 걸리지 않는다. 이 결과로 본 논문의 방법이 실시간으로 수행될 수 있음을 알 수 있다.

광고의 검출을 위해 필요한 집합  $S_i$ 의 원소 개수,  $n_i$ 와 원소들의 배열 순서가 올바르지 판단하기 위한  $T_{idx}$ 가 변함에 따라 precision과 recall의 값이 어떻게 변화하는가에

대한 실험을 한다.

$i$  번째 광고의 총 장면의 개수  $N_i$ 에 따라  $S_i$ 의 크기  $n_i$ 를 얼마로 정할지와 또,  $n_i$ 에 따라 배열 순서 판별을 위한 임계치  $T_{idx}$ 의 값을 얼마로 정해야 하는지를 표 2와 표 3에서와 같이 3가지의 경우로 나누어 실험해 보았으며,  $n_i$ 와  $T_{idx}$ 의 값에 따라 precision과 recall의 변화를 보고 실험을 통해 가장 적합한  $n_i$ 와  $T_{idx}$ 를 정하게 된다.

표 2. Precision

$T_{idx} \backslash n_i$	$0.7 \cdot N_i$	$0.8 \cdot N_i$	$0.9 \cdot N_i$
$0.8 \cdot n_i$	1.00	1.00	1.00
$1 \cdot n_i$	0.98	0.98	1.00
$1.5 \cdot n_i$	0.89	0.80	0.86

표 3. Recall

$T_{idx} \backslash n_i$	$0.7 \cdot N_i$	$0.8 \cdot N_i$	$0.9 \cdot N_i$
$0.8 \cdot n_i$	0.57	0.75	0.50
$1 \cdot n_i$	0.75	1.00	0.63
$1.5 \cdot n_i$	1.00	1.00	0.75

Precision이 작으면 false alarm이 많다는 것을 의미하고, recall이 작다는 것은 miss가 많다는 것이다. Precision과 recall의 값이 모두 1에 가까운 값일 때, 광고 검출이 가장 잘 되었다고 할 수 있다. 그러므로 표 2와 표 3에서 보듯이  $n_i$ 는  $0.8 \cdot N_i$ 일 때,  $T_{idx}$ 는  $1 \cdot n_i$ 일 때 광고의 검출이 잘 되었다고 말할 수 있다. 또한 이때, recall의 값은 1로 miss가 나지 않았지만 precision의 값은 1보다 작은 값으로 false alarm이 생겼음을 알 수 있다. 이것은 그림 4와 같이 광고의 대각선 부분이 일정한 색상으로 계속되어 지면, 특히 흰색과 같은 색상으로 이어지면 이와 유사한 경우가 생기게 되어 false alarm이 생기게 된다.



그림 4. 길이가 다른 광고 검출

그림 4는 광고 데이터베이스의 15초 광고에 의해 방송에서 검출된 광고가 30초인 경우를 보여준다. 두 광고는 밝기도 서로 조금 다르지만, Normalized Histogram을 특징 벡터로 사용하였기 때문에 다른 특징 벡터를 사용했을 때와 비교해 길이와 밝기에 크게 영향을 받지 않는다.

## V. 결 론

본 논문에서는 방송이 되고 있는 스트림의 비주얼 리듬을 이용하여 장면 전환과 특징 벡터를 추출한 후 실시간으로 광고를 검출하는 것을 목표로 알고리즘을 구성하였다. 실험에서는 실제 방송되고 있는 스트림을 분석하지는 않았지만, 본 논문에서 제안한 방법을 이용하면 충분히 실시간으로 광고 검출이 가능함을 실험을 통해 보여주었다.

그 결과 또한, false alarm은 존재하였지만 miss는 생기지 않았기 때문에 방송에서 광고를 찾아내는데 도움이 될 수 있다. 다만, 광고의 장면 수가 적은 것은 false alarm 때문에 사용할 수 없는 단점이 있다. 이 문제를 해결하기 위해서는 한 광고에 대한 특징 벡터 정보를 더 많게 하는 방법이 있겠지만 색인 및 검출 속도에 영향을 줄 수 있기 때문에 이에 대한 더 많은 연구가 필요하다.

## 참고문헌

- [1] Lourens, J.G, "Detection and logging advertisements using its sound", IEEE Transactions on Broadcasting, vol. 36, pp 231-233, Sept. 1990.
- [2] A. Hampapur and R. Bolle, "Feature based indexing for media tracking", International Conference on Multimedia and Expo, vol. 3, pp. 1709-1712, 2000.
- [3] A. Jain, A. Vailaya, and W. Xiong, "Query by video clip", International Conference on Pattern Recognition, vol. 1, pp. 909-911, 1988.
- [4] S. H. Kim and R. Park, "Efficient matching and clustering of video shots", IEEE Transaction on Circuits and Systems for Video Technology, vol. 12, pp. 592-596, 2002.
- [5] Y. Tan, S. Kulkarni, P. Ramadge, "A framework for measuring video similarity and its application to video query by example", International Conference on Image Processing, vol. 2, pp. 106-110, 1999.
- [6] V. Kulesh, V. A. Petrushin, and I. K. Sethi, "Video clip recognition using joint audio-visual processing model", International Conference on Pattern Recognition, vol. 1, pp. 500-503, Aug. 2002.
- [7] H. Kim, J. Lee, J. Yang, S. Sull, W. Kim and S. M. Song, "Visual rhythm and shot verification," Multimedia Tools and Application, 2001.
- [8] M. G. Chung, H. Kim, S. M. Song, "A scene boundary detection method", International Conference on Image Processing, vol 3. pp. 933-936, Sept. 2000.