다차원 쉘 구조에 기반한 내용 기반 검색 기법

김유남[0] 김정림 전승수 이건섭 설상훈
고 려 대 학 교 전 자 공 학 과
{kyn, jrkim, sschun, leeks, sull}@mpeg.korea.ac.kr

Scheme on Content-Based Retrieval using Multidimensional Shell-Based Algorithm

Yunam Kim[0] Jung-Rim Kim Seong Soo Chun Keansub Lee Sanghoon Sull
Dept. of Electronics Engineering, Korea University

요 약

최근 들어 인터넷의 급속한 발전과 엄청난 양의 멀티미디어 데이터가 다양한 형태로 생성, 제공되면서 이에 대한 서비스 및 관리 기법에 관한 문제가 주요 관심사로 대두되고 있다. 그 중에서도, 일반 인터넷 사용자들의 기본적인 요구로써 기존의 문자 기반 검색 서비스로써 제공될 수 없는 내용 기반 검색 기법을 들 수 있다. 사용자가 원하는 영상 혹은 임의의 동영상에서의 한 장면을 빠르고 정확하게 찾는 검색 기법이 최근 들어 연구가 활발히 진행되고 있는 가운데, 본 논문에서는 이러한 내용 기반 검색을 가능하게 하는 하나의 기법으로써 다차원 쉘 구조를 이용한 동영상에서의 영상 검색 알고리즘을 소개한다. 또한, 이를 내용 기반 검색에서 주로 사용되는 색인 트리 구조의 검색 기법과의 비교를 통해 장,단점을 비교 분석해 본다.

1. 서론

현대는 멀티미디어의 시대다. 현재의 인터넷 사용은 이미 학교, 기업은 물론 일반 가정에서도 보편화되었으며, 네트워크의 전송속도 향상 및 다양한 컨텐츠 보급, 확산으로 인하여 일반 인터넷 사용자들은 보다 다양하고 편리한 서비스를 요구하고 있다. 이러한 사용자들의 요구에 맞춰 최근에 연구 개발되는 분야로서 내용 기반(content-based) 검색 기법을 들 수 있다.

기존에 서비스되고 있는 검색 기법으로는 현재 폭발적인 인기를 누리는 야후(yahoo) 및 알타비스타(Alta Vista)등에서 제공되고 있는 문자기반(text-based) 검색 기법을 들 수 있다. 이 방식은 단순히 문자 데이터베이스 내에서 질의 문자(query text)에 매칭되는 문자를 검색함으로써 이와 관련한 URL을 제공해준다. 그러나 이러한 서비스는 멀티미디어의 컨텐트에 기반한 검색을 제공해주지 못하는 한계를 지니고 있다. 따라서 본 논문에서는 동영상에서의 내용 기반(content-based) 검색을 위한 다차원 쉘 구조를 이용한 알고리즘(Multidimensional Shell-Based Algorithm)을 소개하고 현재 많은 연구가 진행되고 있는 동영상 색인 트리를 이용한 검색 기법과의 비교를 통해 두 기법의 장, 단점을 제시한다.

본 논문의 구성은 다음과 같다. 우선 2장에서는 기존에 연구되고 있는 내용 기반 검색에 관한 관련 연구 및 응용 사례를 살펴보고, 3장에서는 본 논문에서 제안하는 검색 기법인 MSBA(Multidimensional Shell-Based Algorithm)를 소개한다. 4장에서는 제안하는 MSBA 검색 기법과 색인된 동영상 트리(tree-indexed video hierarchy) 기법과의 비교 실험을 하고, 마지막으로 최종적인 결론 및 향후 연구 과제를 5장에서 기술한다.

* 본 연구는 한국과학재단 주관의 특정기초연구과제(98-0102-04-01-3) 지원으로 수행되었음.

2. 관련 연구

CBIR(Content-Based Image Retrieval)에 관한 연구는 최근 2~3여년 전부터 시행되어 온 최신 연구 분야이다. 문자 기반의 검색의 한계성을 인식하고 멀티미디어 컨텐트에 기반한 검색 기법 연구가 새로운 이슈가 되면서 다양한 단체 및 대학에서 이에 대한 연구가 급속도로 이루어지고 있다. 내용 기반 검색에 관한 주요 응용 사례는 대표적으로 IBM의 QBIC Systmem[7]과 Virage[8], 그리고 MediaSite[9] 등을 들 수 있다. 이러한 영상/동영상 검색 서비스 시스템은 현재 인터넷을 통해 실험 단계를 거쳐 활발히 운영되고 있다. 이는 방대한 양의 멀티미디어 데이터 처리에 대한 다양한 기법 및 관련 서비스에 대한 제공이 수년 내에 빠르게 성장할 것임을 말해준다. 따라서 보다 효과적인 내용 기반 검색 기법에 관한 연구가 활발히 이루어지리라 기대한다.

3. MSBA를 이용한 검색 기법

3.1 영상의 특징

내용 기반의 영상 검색을 위해서 우선 효과적으로 영상의 특징을 추출하는 기법이 필요하다. 기존에 있는 영상 특징으로는 일반적으로 색상(color), 질감(texture), 형태(shape) 등이 있다. 본 논문에서는 영상의 색상 정보를 이용한 color correlogram

기법을 적용하였다[2]. 색상에서 주로 사용되는 칼라 히스토그램(color histogram)은 각각의 색상이 공간적으로 얼마나 분포한 정도를 나타내는 정보를 얻기 힘들다. 그러나 color correlogram은 한 영상에서 각각의 색상값이 거리값에 따라 얼마나 변화하는 정도를 나타내는 공간적 상관도(spatial correlation)를 나타내 줌으로써 영상의 특징을 보다 잘 표현한다.

영상의 color correlogram에 관한 식은 다음과 같이 나타낼 수 있다.

$$\gamma_{c_i,c_j}^{(k)}(I) \equiv \Pr_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_j} \mid\mid p_1 - p_2 \mid = k] \quad (1)$$

* Pr : 확률값, p: 픽셀 좌표값, I_{cj} : Cj 값을 갖는 전체 픽셀 집합, k : 거리값, Ci : 색상 값

일반적으로, 한 영상에서의 color correlogram을 적용한 특징 추출 정보량은 상당히 많기 때문에 동일한 색상에 관한 공간적 상관도를 추출하는 autocorrelogram을 이용한다. 본 연구에서도 영상의 특징 추출로써 autocorrelogram을 이용하였다. 이 때 추출된 영상의 특징 벡터에 관한 차원은 256으로 설정하였다.

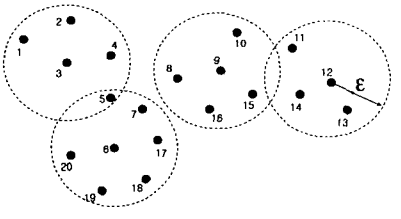
또한, 본 논문에서는 두 영상 간의 유사도(similarity)를 정량적으로 측정할 수 있는 거리 함수(distance function)로써 L1 norm을 적용하였다.

3.2 키프레임(keyframe) 추출

동영상은 사람의 눈으로 볼 때, 연속적이고 부드러운 장면을 연출하지만 그에 반하여 다수의 동일한 영상이 존재하게 된다. 따라서 내용 기반의 검색을 위해서는 우선적으로 내용의 중복성(redundancy)을 제거하기 위해 대표 프레임(representative frame)을 추출해야 한다. 이 때, 추출된 대표 프레임을 키프레임이라 하고 추출된 키프레임을 통해 내용 기반 검색을 위한 동영상 데이터베이스를 구성한다.

본 논문에서는 트리 구조를 이용한 전형적인 검색 기법의 하나인 [1]에서 제시한 동영상 검색 기법과의 비교를 위해 동일한 키프레임 추출 기법을 사용하였다. [1]에서는 키프레임 추출을 위해 그리디(greedy) 접근 방식을 사용하였다.

동영상의 임의 구역의 스트림(stream)이 특징 공간(feature space) 상에서 [그림 1]과 같이 분포되었다고 하자. 이 때, 키프레임 추출을 위한 사용자 임계값인 ε을 적절히 선정하여 모든 프레임에 대하여 ε을 반경으로 하는 원을 그린 후 고려해 볼 때, 최소 개수의 키프레임을 선택하는 조건에서 보면, [그림 1] 에서 프레임 3,6,9,12를 중심으로 하는 원을 그릴 때 최적의 해를 만족하게 된다.

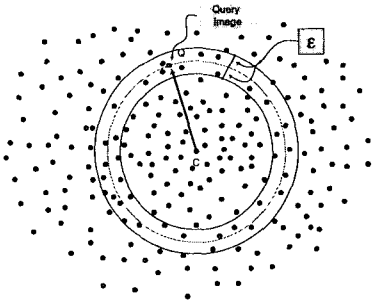


[그림 1] 특징 공간에서의 프레임 분포 예시도

이러한 최소 커버 문제는 Quine-McCluskey 알고리즘[3]을 적용하여 최적의 해를 얻을 수 있다.

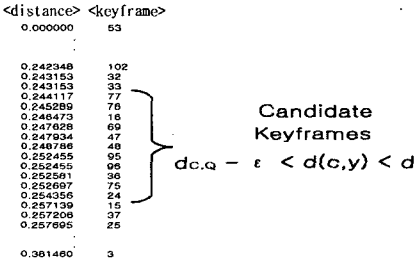
3.3 제안 알고리즘

내용 기반 검색을 위해서는 실제 검색 엔진에 이용되기 위한 정보를 위한 부호화(encoding) 과정을 거치게 된다. 즉, 추출된 키프레임들을 가지고 효율적인 검색을 유도하기 위한 효과적인 데이터베이스를 구성하게 된다. 본 논문에서 제시하는 기법은 특징 공간에서 키프레임 전체의 중심 프레임을 구한 후, 그 중심 프레임으로부터 나머지 프레임들 간의 거리 값을 계산해 낸 후 이를 하나의 거리 벡터로써 표현한다. 이때의 거리 벡터는 거리값에 대해 정렬(sorting)이 된 후 검색 엔진에 적용하게 된다.



[그림 2] 특징 공간에서의 MSBA 설명 예시도

부호화 과정에서 얻은 중심 키프레임에 대한 거리 벡터를 실제 검색에 응용되는 과정은 다음과 같다. 우선, [그림 2]는 MSBA 기법을 설명하기 위한 다차원 특징 벡터를 2차원 평면에서 고려해본 예시도이다. 그림에서 보는 바와 같이, 질의 이미지(query image)가 사용자에 의해 질의되면 우선 질의 이미지(Q)와 중심 키프레임(C)간의 거리값 $d_{C,Q}$ 를 계산한다. 그 다음으로, 사용자 임계치 값을 보상시키는 영역을 결정해야 하므로 거리값이 $d_{C,Q} - \varepsilon$ 과 $d_{C,Q} + \varepsilon$ 구간에 존재하는 키프레임들을 [그림 3]과 같이 거리 벡터로부터 찾아내고, 그러한 구간에 존재하는 키프레임들이 결국 사용자의 질의 이미지에 대한 유사 이미지로써의 후보 프레임들로 선택하게 된다.



[그림 3] 거리 벡터로부터의 검색 대상 키프레임 선정 예

즉, 사용자 임계치(ε)를 만족시키는 다차원 쉘 구조에 존재하는 키프레임은, d를 n차원 공간 R^n의 거리함수라 할 때, $C \in R^n$ 와 한 실수(사용자 임계치) $\varepsilon > 0$ 에 대하여

$$S_d(C,\varepsilon) = \{y \in R^n \mid\mid d_{C,Q} - d(C,y) \mid < \varepsilon\}$$
$$= \{y \in R^n \mid d_{C,Q} - \varepsilon < d(C,y) < d_{C,Q} + \varepsilon\} \quad (2)$$

를 만족하는 키프레임 y 들의 집합을 의미한다. 이렇게 미리 선별한 집합 Sd 안에 존재하는 키프레임들과 질의 이미지와의 유사도 비교를 통해 최종적으로 사용자가 원하는 최적의 유사 프레임에 대한 검색을 가능하게 할 수 있다.

4. 실험 및 결과

4.1 실험 환경

제안하는 MSBA 검색 기법과 전형적인 트리 구조를 이용한 검색 기법의 하나인 [1]과의 비교를 위해 동일한 동영상 데이터와 동일한 키프레임을 가지고 실험하였다. 실험에 사용된 테스트 동영상은 다음 표와 같다.

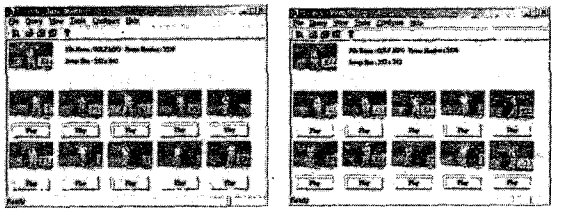
Info / Video	Source	Length	# of keyframe
KBS Golf	MPEG-7 Content(V18)	10' 04"	1170
NHK Documentary	MPEG-7 Content(V10)	9' 04"	198

[표 1] 테스트 동영상

표에서 키프레임 추출에 사용된 ε 값은 0.1이며 이는 각각의 프레임 간의 거리를 0과 1사이의 값으로 정규화 과정을 거친 후, 많은 실험을 거쳐 적절한 값으로 선택된 값이다. 또한, 실제 실험에 적용된 영상의 특징은 256 차원(dimension)을 갖는다.

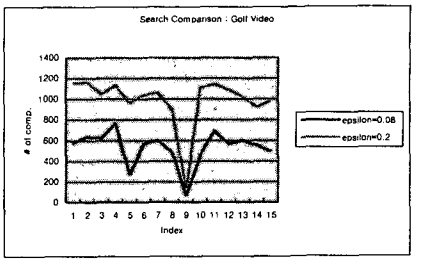
4.2 트리 구조 기법과의 비교

[1]에서 제안하는 색인된 트리 구조를 이용한 내용 기반 검색에 있어서, 실제 검색 엔진을 구동시키기 위한 데이터베이스 부호화 과정에서 추출된 키프레임 수 N에 대하여 $O(N^2)$의 시간 복잡도(time complexity)를 갖는다. 이는 트리를 구성하는 과정에서 모든 키프레임들간의 거리 관계를 이용하여 3.2에서 제시한 기법으로 유사 키프레임들끼리 군집화(clustering)과정을 거치므로 $O(N^2)$의 복잡도를 갖게 된다. 반면에, 제안하는 MSBA 검색 기법은 중심 키프레임으로부터의 거리 벡터를 구한 후, 이를 실제 검색 과정에서 빨리 수행하기 위해 정렬을 취했다. 본 연구에서는 Merge Sort 알고리즘을 적용하였다. 결국, 항상 O(NlogN)을 보장하므로 동영상 데이터베이스를 증가시킬수록 부호



[그림 4] 검색 결과의 한 예(질의:Golf #3578 frame)
(좌:MSBA 기법 이용시, 우:Tree-Based 기법 이용시)

화 과정에 드는 복잡도는 상대적으로 줄어들게 된다. 이는 비록 부호화 과정이 off-line에서 실행되는 것이라 할지라도 데이터베이스 구성 측면에서 볼 때 상당한 계산량을 줄일 수 있는 이점을 갖는다. 또한, [그림 4]에서 보는 바와 같이, 질의에 대한 결과는 두 기법에 대해 거의 유사한 만족도를 제공해준다. 그러나 MSBA의 단점으로는 사용자 임계치에 상당한 의존성을 갖는다는 점이다. 사용자 임계치(ε)가 각각 0.08, 0.2 인 경우에 대해 실험한 결과, 0.08인 경우에는 평균적으로 순차검색보다 55% 정도 빨랐으나, 0.2인 경우에는 겨우 15% 수준에 그쳤다. [그림 5]는 Golf



[그림 5] 검색 비교 회수

동영상에 대한 실험을 한 결과이다. 트리 구조를 이용한 검색에서는 사용자 임계치에 대한 의존성(dependency)이 크지 않았다. 평균적으로 순차 검색보다 30~40% 정도의 빠른 검색율을 보였는데 이는 트리 구성시에 충실도(fidelity) 값을 미리 계산함으로써 검색 과정에 충실도 값에 기반하여 중도 차단(pruning)을 가능하게 함으로써 이와 같은 결과를 보였다.

5. 결론 및 향후 과제

본 논문에서는 내용 기반 검색을 위해 MSBA 검색 기법을 제안하였고 이를 트리 구조를 이용한 검색 기법과의 비교를 실행하였다. 그 결과, 검색 엔진을 구동하기 위한 전 단계인 부호화 과정에서는 MSBA 기법이 O(NlogN)을 보장함으로써 보다 나은 결과를 나타내었고, 적절한 사용자 임계치에 대해서 대체로 비슷한 검색 결과를 보이면서 좀 더 빠른 검색을 유도하였다.

그러나 사용자 임계치에 상당한 의존성이 있음을 확인할 수 있었다. 현재 제안하는 MSBA 기법에서 좀 더 개량된 것으로서, 특징 공간상에서 가장 먼 거리에 있는 두 키프레임에 대한 거리 벡터 두개를 사용하여 두 다차원 쉘의 교차지역에 대한 검색을 시행한다면 이와 같은 문제는 어느정도 해결되리라 생각하며 이를 향후 과제로 남겨둔다.

6. 참고 문헌

[1] H.Chang, S.Sull and S.Lee, " Efficient Video Indexing Scheme for Content-Based Retrieval," *IEEE Trans.Circuits and Systems for Video Technology*, vol.9,pp1269-1279, Dec, 1999.

[2] J.Huang, S.R.Kumar, M.Mitra, W.J.Zhu, and R.Zabih, " Image indexing using color correlograms," *Proc.of 16th IEEE Conf. on Computer Vision and Pattern Recognition*, pp.762-768, 1997.

[3] J.P.Hayes, *Introduction to Digital Logic Design*, Reading: Addison Wesley, 1993.

[4] Wayne Wolf, " Key frame selection by motion analysis," in *Proc. IEEE Int. Conf. Acoust, Speech, and Signal Proc*, 1996

[5] Jau-Yuen Chen, Charles A. Bouman and John C. Dalton, " Active Browsing using Similarity Pyramids," *SPIE/IS&T conference on Storage and Retrieval for Image and Video Database VII*, vol.3656, pp.144-154, San Jose CA, January 26-29, 1999

[6] Yueting Zhuang, Yong Rui, Thomas S. Huang, and Sharad Mehrotra, " Adaptive key frame extraction using unsupervised clustering," in *Proc. IEEE Int. Conf. On Image Proc.*, 1998

[7] http://wwwqbic.almaden.ibm.com

[8] http://www.virage.com

[9] http://www.mediasite.net