내용 기반 검색을 위한 효율적인 동영상 데이터베이스 구성 기법

김유남, 김정림, 이건섭, 설상훈

고려대학교 전자공학과

Efficient Video Database Design Scheme for Content-Based Retrieval

Yunam Kim, Jung-Rim Kim, Keansub Lee and Sanghoon Sull

Dept. of Electronics Engineering, Korea University

요 약

최근 들어 내용 기반 검색에 관한 연구가 활성화되면서 이에 대한 많은 검색 기법들이 제안되고 있다. 대부분의 검색 기법에서는 초기에 구성된 데이터베이스에서 새로운 데이터를 추가하여 데이터베이스를 업데이트하게 될 때, 처음부터 모든 데이터에 대한 색인 처리 등의 재구성을 시행해야 했다. 본 논문에서는 이러한 데이터베이스 업데이트 시의 단점을 극복하고 사용자의 질의 영상에 대하여 데이터베이스에서 빠르게 검색하는 알고리즘을 제안한다. 이는 특징 공간상에 분포한 키프레임간의 거리 분포도를 고려한 거리 벡터와 사용자의 검색 임계치를 보장하는 다차원 쉘 구조를 적용하여 효율적인 데이터베이스 구성 및 빠른 검색에 관한 알고리즘으로써, 이를 내용 기반 검색에서 주로 사용되는 색인 트리 구조를 이용한 검색 기법과의 비교를 통해 장,단점을 비교 분석해 본다.

Ⅰ. 서 론

현대는 멀티미디어 시대다. TV 방송, 사진, 그래픽스, 애니메이션, 오디오, 비디오와 같은 멀티미디어 데이터가 하루에도 엄청난 양으로 새로이 생성되고 있으며, 우리는 이러한 멀티미디어 데이터를 매일 손쉽게 접하며 이를 통한 다양한 서비스를 누리고 있다. 전 세계적으로 인터넷의 급속한 발전으로 연간 기하급수적인 네티즌이 생기면서 이를 통해 서비스 업체는 사용자의 다양한 서비스 요구를 충족시키기 위한 기법이 끊임없이 연구, 개발되고 있다.

이러한 인터넷을 통한 멀티미디어 서비스의 하나로써 동영상 검색 서비스를 들 수 있다. 불과 4~5년 전에 생겨난 텍스트 기반의 정보 검색 서비스가 폭발적인 인기를 누리는 것과 더불어 최근에는 IBM Qbic[1], Virage[2], MediaSite[3] 등과 같은 동영상 검색 서비스가 제공되고 있으며 현재 활발히 운영되고 있다.

본 논문에서는 이러한 동영상 검색 서비스를 제공하는 방법에 관한 것으로써, 검색 엔진에서 구동되는 동영상 데이터베이스의 효율적인 구성 및 사용자가 원하는 질의 영상(query image)을 빠르게 검색하는 알고리즘을 제안한다.

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

즉, 기존에 제안된 대부분의 이미지/동영상 검색 시스템[4, 6, 7, 8, 9]의 단점은 새로운 이미지/동영상 데이터가 데이터베이스에 추가됨으로써 데이터의 색인 처리(indexing)를 새로 수행해야 한다는 점에 있다. 이는 대용량의 멀티미디어 데이터가 미리 색인되어 데이터베이스가 구성되어 있을 때, 상대적으로 아주 적은 양의 데이터가 추가된다 하여도 데이터베이스를 업데이트(update)하는 데에 있어서는 기존의 모든 데이터를 통해 색인 처리가 다시 이루어져야 하는 데에 기인한다.

따라서, 본 논문에서는 이러한 데이터 추가 시, 추가된 데이터만의 색인을 거쳐 기존의 데이터베이스 구조와 단순한 병합(merge) 처리만 수행하여 새롭게 데이터베이스를 구성할 수 있는 기법을 제안한다. 이러한 기법으로써 데이터베이스 구성에 드는 시간적 복잡도(time complexity)를 상당히 경감시킬 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 본 연구와 관련된 멀티미디어 데이터의 내용 기반 검색(content-based retrieval)을 위한 다양한 기법을 소개하고, 본 논문에서 제안하는 효율적인 데이터베이스 구성 기법 및 검색 알고리즘을 3장에서 기술한다. 아울러 4장에서는 본 논문에서의 제안 알고리즘과 기존의 검색 기법과의 비교 실험을 통해 성능을 평가, 분석해 보고, 마지막으로 5장에서는 제안 알고리즘에 대한 최종적인 결론 및 향후 연구 과제를 고찰해 본다.

Ⅱ. 관련 연구

CBIR(Content-Based Image Retrieval)에 관한 연구는 최근 2~3여년 전부터 주로 시행되어 온 최신 연구 분야이다. 이러한 내용 기반 검색 기법에 있어서 현재까지 다양한 알고리즘이 제안되었는데 주로 이미지/동영상 데이터를 색인하는 과정에서 다양한 기법들이 제안되었다.

기존에 제안된 멀티미디어 색인 기법으로써 전형적이고 가장 많이 사용되는 방식으로써 트리 구조를 이용한 색인 알고리즘이 있다[4, 6, 7, 8, 9]. 이러한 알고리즘에서는 데이터베이스에 있는 이미지들로부터 특징 벡터를 추출해 낸 후 각각의 특징 벡터를 통해 유사도(similarity)에 따른 트리를 구성하는 방식을 사용하였다. [4]에서는 각 영상의 특징 벡터로부터 그리디(greedy) 접근 방식의 군집화(clustering) 알고리즘을 적용하여 유사도에 기반한 트리를 구성하는데 이때 계층적으로 형성된 영상의 트리 구조에서 각 노드에 있는 키프레임은 자신이 포함하는 하위 키프레임을 얼마나 잘 표현하는가 하는 정량적 정보를 계산하여 검색 과정에 적용됨으로써 검색 속도 및 효율을 높이고 있다. 또한, [7]에서는 TSVQ(Tree-Structured Vector Quantization) 알고리즘을 이용하여 이미지 데이터베이스를 coarse-to-fine 형태로 표현되는 top-down 트리로 구성하여 삼각 부등식(triangle inequality)을 만족하면서 빠른 branch and bound 검색을 유도하고 있다. [8]에서는 SS-Tree(Similarity Search Tree)라는 데이터 구조를 새롭게 제안하며 R*-Tree와의 검색 성능 비교를 통해 저장공간 및 검색 과정의 비교회수에서 보다 나은 성능을 보이고 있으며, [9]에서는R-트리를 이용하여 다차원 특징 벡터 공간에서 질의된 객체와 가장 가까운 객체의 점을 찾는 방법을 제시하고 있다.

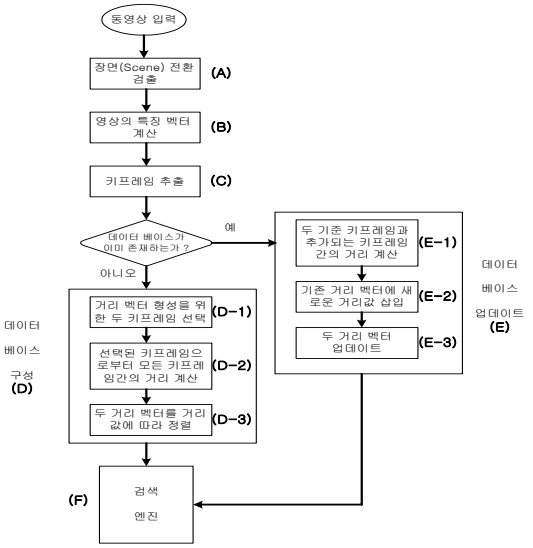
그러나 위의 색인 기법들은 새로운 영상 데이터가 데이터베이스에 추가되어 새롭게 데이터베이스를 구성할 때 제안하는 각 기법들을 처음부터 다시 시행해야 하는 단점이 있다. 즉, 데이터베이스 업데이트에 대한 효과적인 기법이 소개되지 않았다. 또한, [7, 8, 9]에 사용된 색인 기법들은 비록 검색면에 있어서는 빠른 검색을 유도하지만 정확 탐색(exact search)이 아닌 근사 탐색(approximate search)을 목적으로 한 알고리즘이다.

본 논문에서는 데이터베이스 업데이트 수행 시, 기존에 구성된 데이터베이스와 무관하게 새로 추가된 데이터만에 대한 색인 절차를 통해 기존 데이터베이스와 단순한 병합 처리만으로 업데이트가 수행하게 하며 또한 사용자 임계 범위내에서 빠른 검색을 유도하도록 질의 이미지에 대하여 유사도가 적정 범위에 드는 후보 키프레임들(candidate keyframes)을 미리 선별하는 알고리즘을 고려해 보았다.

Ⅲ. 제안 알고리즘

본 장에서는 제안하는 데이터베이스 구성 기법 및 이를 이용한 동영상 검색 기법을 상세히 소개한다.

동영상 검색 엔진을 구동시키기 위해서는 우선 데이터베이스를 구성시켜야 하는데 이를 위해서는 동영상으로부터 장면 전환 검출, 영상의 특징 추출 및 키프레임 추출과 같은 선행 과정이 필요하다. [그림 1]은 본 논문에서 제안하는 알고리즘에 대한 전체적인 플로우챠트를 나타낸다.



[그림 1] 데이터베이스 구성 알고리즘 흐름도

각 단계별로의 세부 설명은 다음과 같다.

1) A 단계 : 장면전환 검출

본 논문에서는 MPEG-1 비디오 스트림에 대해 실험하였으며 샷(shot) 단위로 처리하기 위하여 dissolve, fade in/out, wipe 등의 장면전환 검출 기법을 사용하였다.

2) B 단계 : 영상의 특징 벡터 계산

J.Huang이 제안한 color correlogram[5] 기법을 적용하여 영상의 특징 벡터를 추출하였다. 본 실험에 사용된 특징 벡터는 256 차원(dimension)의 요소를 갖는다. 또한 본 논문에서는 두 영상 간의 유사도(similarity)를 정량적으로 측정할 수 있는 거리 함수(distance metric)로써 L1 norm을 적용하였다.

3) C 단계 : 키프레임 추출

본 논문에서는 트리 구조를 이용한 전형적인 검색 기법의 하나인 [4]에서 제시한 동영상 검색 기법과의 비교를 위해 [4]에서 사용된 그리디(greedy) 접근 방식을 동일하게 적용하였다.

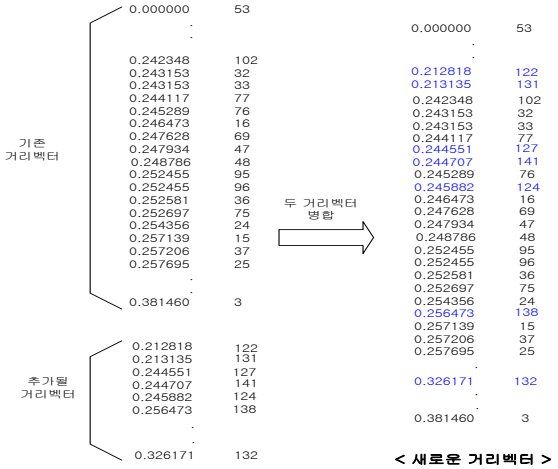
4) D 단계 : 데이터베이스 구성

윈 동영상으로부터 장면전환 검출, 각 프레임의 특징 벡터 계산 및 이를 이용한 키프레임 추출 과정이 끝나면 추출된 키프레임을 갖고 데이터베이스를 구성하게 된다. 특징 공간 상에 분포하는 모든 키프레임 중에서 거리차가 가장 큰 두 키프레임 한쌍을 계산해 구한 후(D-1), 구해진 두 키프레임을 기준 키프레임(Reference Keyframe)으로 하여 그 두 키프레임으로부터 나머지 모든 키프레임간의 거리값을 구함으로써 두 개의 독립적인 거리 벡터(distance vector)를 생성시킨다(D-2). 이때 특징 공간 상에서 거리차가 가장 큰 두 키프레임을 선정하는 이유는 후에 설명할 두개의 다차원 쉘을 이용하여 교차 영역에 대한 검색을 시행할 때 확률적으로 교차 영역을 좁힐 수 있는 중요한 요인이 되기 때문이다. 여기서, 거리 벡터는 기준 키프레임으로부터 나머지 키프레임에 대한 거리 값 및 그 값에 해당하는 해당 키프레임 ID에 대한 정보를 포함하고 있는데,

거리 벡터의 거리 값들을 다시 오름차순(혹은 내림차순)으로 정렬함으로써(D-3) 최종적인 두 개의 거리 벡터를 생성하여 검색 엔진에 이용하게 된다. 두 개의 거리 벡터가 검색 엔진에 이용되는 방식은 후에 설명하기로 한다.

5) E 단계 : 데이터베이스 업데이트

이미 데이터베이스가 형성된 상태에서 새로운 데이터를 데이터베이스에 추가할 때 동작되는 루틴이다. 기존에 제안된 방식에서는 새로운 데이터가 추가될 때마다 기존에 있는 데이터도 같이 고려하여 색인 과정이 이루어 지는데 반하여 본 알고리즘에서는 새로 추가된 키프레임들에 대한 거리 벡터를 형성하여 기존의 거리 벡터와 새로 형성된 거리 벡터를 병합(merge)하여 데이터베이스를 업데이트 하게 된다. 거리 벡터간의 병합 과정은 [그림 3]과 같다.



[그림 2] 두 거리 벡터 병합

이때 두 거리 벡터의 빠른 병합 처리를 위해서 두 거리 벡터의 거리 값들을 각각의 힙(Heap)에 저장하여 거리값의 대소에 따라 새로운 거리 벡터에 대한 힙에 삽입하면 빠른 병합을 유도할 수 있다.

6) F 단계 : 검색 엔진 구동

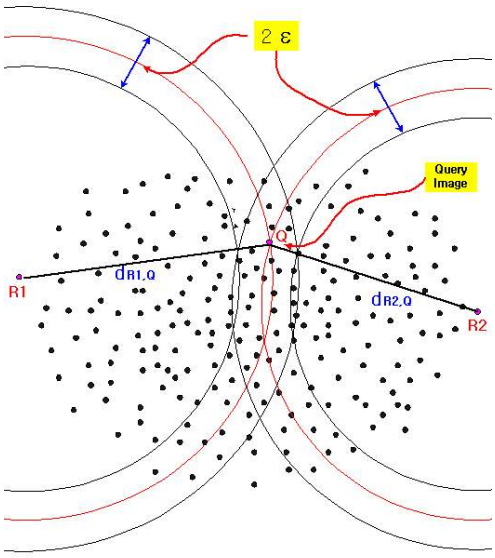
D(E) 단계에서 형성된 두 거리 벡터를 이용하여 검색에 적용하게 되는데 실제 검색에 응용되는 과정은 다음과 같다. 우선, [그림 3]은 제안하는 검색 기법을 설명하기 위한 키프레임의 다차원 특징 벡터를 2차원 평면에서 고려해 본 예시도이다. 그림에서 보는 바와 같이, 질의 이미지(query image)가 사용자에 의해 질의되면 우선 질의 이미지(Q)와 기준 키프레임(R1, R2)간의 거리값 $d_{R1,Q}$, $d_{R2,Q}$를 계산한다.

그 다음으로, 사용자 오류 허용 임계치인 ε을 보장하면서 질의 이미지에 대한 유사 이미지로써의 후보 프레임들로 선택될 영역을 결정한다. 즉, 사용자 오류 허용 임계치(ε)를 만족시키는 다차원 쉘 구조에 존재하는 키프레임은, d를 n차원 공간 R^n의 거리함수라 할 때, 한 실수(사용자 오류 허용 임계치) $\varepsilon > 0$ 에 대하여

$$A = A(R1, \varepsilon) = \{x \in R^n \mid |d_{R1,Q} - d(R1, x)| < \varepsilon\}$$

$$B = B(R2, \varepsilon) = \{y \in R^n \mid |d_{R2,Q} - d(R2, x)| < \varepsilon\} \quad (1)$$

$$S = S(R1, R2, \varepsilon) = A \cap B$$



[그림 3] 특징 공간에서의 검색 기법 예시도

를 만족하는 집합 S에 속한 키프레임들이 후보 키프레임들로 미리 선정하게 된다. 두 기준 키프레임으로부터 형성된 두 다차원 쉘의 교차 영역에 속한 키프레임은 각각의 다차원 쉘에 속한 키프레임을 힙(Heap)에 저장하여 두 다차원 쉘에 중복하여 존재하는 키프레임을 새로운 힙에 삽입하면 교차되는 키프레임을 빠르게 추출해 낼 수 있다.

이렇게 미리 선별한 집합 S 안에 존재하는 키프레임들과 질의 이미지와의 유사도 비교를 통해 최종적으로 사용자가 원하는 최적의 유사 프레임에 대한 검색을 가능하게 할 수 있다.

Ⅳ. 실험 및 결과

1. 실험 환경

본 실험은 펜티엄 III(450MHz), 256 Mb RAM의 기본 하드웨어와 Microsoft Visual C++ 6.0을 컴파일러로 사용하였다. 또한 본 실험에 사용된 동영상 데이터는 다음과 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

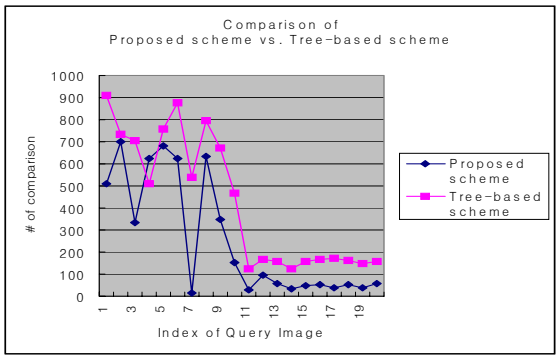
2. 트리 구조 기법과의 비교

본 논문에서는 동영상을 트리 구조를 이용하여 색인처리 후 검색에 응용되는 방식중에 [4]에서 제안한 색인 기법과 제안 알고리즘과의 비교를 수행하여 보았다.

데이터베이스 구성을 위한 색인 처리에 있어서 [4]에서는 그리디 접근 방식을 사용함으로써 시간 복잡도가 $O(N^2)$에 이르는 반면 제안 알고리즘은 특징 벡터간의 거리 계산

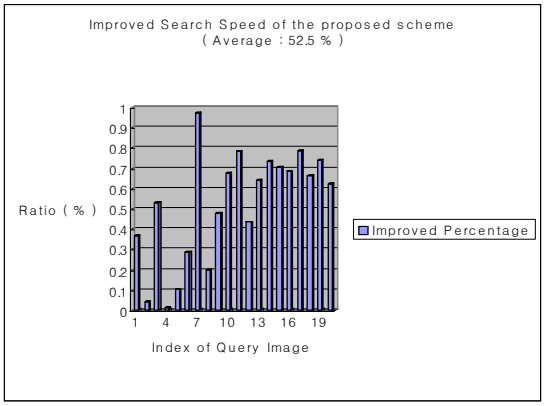
후 거리 벡터를 형성하는 과정에서 정렬에 드는 시간 복잡도로써 O(NlogN)이 보장되었다. 비록 데이터베이스 구성이 off-line에서 수행되는 것이라 할지라도 데이터베이스 구성 측면에서 볼 때 상당한 계산량을 줄일 수 있는 이점을 갖는다. 또한, 하나의 동영상에서 일부 동영상 클립(clip)이 제거된 상태에서 데이터베이스가 구축되었을 때 나머지 클립을 새로 데이터베이스에 추가시키고자 할 때, [4]에서는 추가되는 동영상 클립 및 원래 데이터베이스에 존재하는 키프레임 전체에 대한 색인 과정을 수행해야 하는 반면, 본 논문에서 제안하는 기법에서는 추가되는 클립에 대한 데이터베이스 구성이 완료된 후 기존 데이터베이스와 단순한 병합 처리만 하여도 새롭게 데이터베이스 업데이트가 가능하였다.

이와 같은 데이터베이스 구성에 관한 이점 뿐만 아니라 검색 성능면에서도 우수한 효율을 나타내었다. [그림 4]는 트리 구조 기법의 검색과 제안 알고리즘에 대한 검색 성능을 비교한 그래프이다.



[그림 4] 제안 알고리즘과 트리 기법과의 검색 성능 비교

사용된 질의 이미지 중 1~10은 KBS Golf에서, 11~20은 NHK Documentary에서 무작위로 추출한 영상으로써 똑같은 질의에 대한 두 알고리즘의 검색 성능을 비교하였다. 그림에서 보듯이 제안 알고리즘을 이용하였을 때 트리 구조를 이용한 방식보다 검색시 요구되는 비교회수가 적게 들었다. 평균적으로 제안 알고리즘은 색인 트리 검색 기법보다 52.5% 정도의 향상된 성능을 가짐을 확인할 수 있었다. [그림 5]는 제안 기법의 성능 향상 정도를 나타낸다.



[그림 4] 제안 알고리즘의 검색 성능 향상도

V. 결론 및 향후 과제

본 논문에서는 내용 기반 검색을 위한 효율적인 데이터베이스 구성 및 사용자의 질의에 빠른 검색을 유도하는 기법에 관한 연구를 수행하였다. 기존에 주로 사용되는 트리 구성 방법보다 데이터베이스의 효과적인 업데이트가 가능하고 색인 과정에 드는 시간 복잡도가 훨씬 적게 나타남을 확인할 수 있었다. 또한 검색 성능면에서도 색인 트리를 이용한 기법보다 대략 52.5% 정도의 빠른 검색율을 나타내었다.

그러나 본 연구에서 실험한 데이터베이스는 전체 동영상이 약 20분 가량으로써 추출된 전체 키프레임은 1300여개 수준으로 설정되었다. 이론적으로 키프레임 수가 많아지면 두 다차원 쉘의 교차 영역에 분포하는 키프레임 수 또한 많아질 것이므로 사용자 임계값을 조절하면서 이에 대한 많은 실험이 필요하다. 따라서 본 논문에서 제안하는 알고리즘을 장시간의 동영상 데이터베이스에 대해서도 적용해 보아야 할 것이며 이를 향후 과제로 남겨둔다.

참고 문헌

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

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

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

[4] 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.

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

[6] Jau-Yuen Chen, Charles A. Bouman, John C. Dalton, "Hierarchical Browsing and Search of Large Image Databases," *IEEE Trans. Image Processing*, vol. 9, no. 3, March 2000.

[7] Jau-Yuen Chen, Charles A. Bouman and Jan P. Allebach, "Fast Image Database Search using Tree-Structured VQ," in *Proc. of International Conference on Image Processing*, vol. 2, pp.827-830, 1997

[8] D.A.White and R.Jain, "Similarity Indexing with the SS-tree," in *Proc. 12th Intl. Conf. On Data Engineering*, pp.516-523, 1996.

[9] Roussopoulos N., Kelley S., Vincent F., "Nearest Neighbor Queries," in *Proc. ACM SIGMOD Int. Conf. on Management of Data*, pp.71-79, 1995.

[10] 김유남, 김정림, 전승수, 이건섭, 설상훈, "다차원 쉘 구조에 기반한 내용 기반 검색 기법," 제27회 한국정보과학회 추계학술발표회, Oct, 2000.