

## 인덱싱을 위한 MPEG-7 시각 정보 기술자 분석

박주현<sup>0</sup> 남종호

서강대학교 컴퓨터학과

[parkjh@sogang.ac.kr](mailto:parkjh@sogang.ac.kr)<sup>0</sup>, [jhnang@ccs.sogang.ac.kr](mailto:jhnang@ccs.sogang.ac.kr)

### Analysis of MPEG-7 Visual Descriptors for Data Indexing

Joohyoun Park<sup>0</sup> Jongho Nang

Dept. of Computer Science, Sogang University

#### 요 약

검색된 정보의 정확성과 검색 속도는 정보 검색 시스템의 성능 평가에 있어서 매우 중요한 요소이다. MPEG-7 시각 정보 기술자는 예제 기반 이미지/비디오 검색 시스템을 구성할 때 사용할 수 있는 저급 수준 시각 정보에 대한 표준화된 기술을 의미한다. 제안된 기술자는 검색 결과의 정확성에 대해서는 검증받은 상태이지만, 고차원 데이터거나 시간 복잡도가 큰 매칭 함수를 가지고 있어 순차 검색을 할 경우 매우 많은 검색시간을 필요로 한다. 일반적으로 이러한 문제를 해결하기 위해 인덱스 정보를 구성하여 검색에 적용하는 방법이 주로 사용된다. 본 논문에서는 MPEG-7 시각 정보 기술자중 이미지 검색 시스템에 이용할 수 있는 색상 기술자와 텍스처 기술자, 그리고 외형 기술자를 인덱스 정보 구성 관점에서 분석하였다

#### 1. 서 론

MPEG-7은 멀티미디어 내용에 대한 기술 표준으로서 파트 3인 Visual 파트[1]에서는 이미지/비디오 검색에 이용할 수 있는 시각 정보 기술자(Visual Descriptor)가 정의되어 있다. 제안된 시각 정보 기술자들은 표준화 과정을 통해 동영상이나 이미지에 대한 표현 적합성이 이미 검증된 상태이며, 많은 관련 개발자들이 제안된 기술자를 사용하여 동영상/이미지 브라우저/검색 시스템을 연구 개발 하고 있다.

실용적인 면에 있어서 검색 시스템은 검색 결과의 정확성뿐만 아니라 검색 속도 또한 매우 중요한 요소임에 분명하다. 일반적으로 대량의 오브젝트를 가지고 있는 데이터베이스에서 질의내용에 부합하는 데이터를 순차 검색하는 것은 매우 많은 시간을 필요로 하는 일이다. 따라서 검색 속도 향상을 위한 특별한 방법이 필요하며, 그 방법의 하나로써 인덱싱을 생각해 볼 수 있다. 인덱스란 각 오브젝트의 특징을 효과적으로 표현 및 구성하여 검색에 활용되는 색인 정보이다. 오브젝트의 수가 매우 많고 고차원으로 구성되어 있거나 매칭 함수의 시간 복잡도가 매우 크다면 인덱스 구성은 선택이 아니라 필수적인 요소가 된다.

고차원인 MPEG-7 시각 정보 기술자를 이용하여 이미지 검색 시스템을 구성할 경우 속도 문제는 반드시 해결해야 할 문제이다. 본 논문에서는 MPEG-7 시각 정보 기술자에 대한 인덱스를 구성하여 검색 속도 향상을 주고자 했을 경우 발생할 수 있는 문제점에 대해서 논의하고, 그 해결 방안을 제시하고자 한다.

#### 2. 관련 연구

##### 2.1 인덱스에 관한 기존의 연구

인덱스를 구성하는 방법은 크게 좌표 기반 인덱싱 (Coordinates Based Indexing)과 거리 기반 인덱싱 (Distance Based Indexing)으로 나누어 생각해 볼 수 있다. 전자는 공간상의 데

이터 위치를 기반으로 해서 인덱싱 하는 방법으로서, 데이터가 벡터공간(Vector Space)에서 표현할 수 있을 경우에만 사용 가능하다. R-Tree[2] 계열이 대표적인 벡터 공간에 존재하는 다차원 데이터에 대한 인덱싱 방법들이다. 하지만 이 방법들은 10차원 이하의 다차원 데이터에 대해서만 좋은 성능을 보여주고 있기 때문에, MPEG-7 시각정보 기술자와 같은 고차원 데이터에 적용하기에는 문제가 있다. 이는 차원이 커짐에 따라 검색 공간이 기하 급수적으로 커지게 되며, 검색 공간상에 오브젝트들을 산재시키는 소위 curse of dimensionality문제에 기인한다. VA-File[3]등은 고차원 공간에서의 데이터의 분포 특성이나, 고차원 데이터가 가지고 있는 특성에 기반하여 인덱스를 구성하는 방법을 제안하고 있다.

거리 기반 인덱싱 방법은 데이터들간의 거리를 기준으로 하여 인덱스를 구성하는 방법으로서 VP-Tree[4]등이 대표적이다. 이 방법을 사용하기 위해서는 데이터가 메트릭 공간(Metric Space)에 존재해야 한다.

##### 2.2 MPEG-7 시각 정보 기술자

MPEG-7 시각 정보 기술자는 색상(Color), 텍스처(Texture), 외형(Shape), 움직임(Motion), 그리고 그외(eg: face descriptor) 총 5개의 그룹으로 이루어져 있다. 이 중 이미지 검색에 이용할 수 있는 것은 색상, 텍스처, 외형 그룹이다. 색상 그룹에는 Dominant Color(DC), Scalable Color(SC), Color Layout(CL), Color Structure(CS), 4개의 기술자가 있으며, 텍스처 그룹에는 Homogeneous Texture(HT), Edge Histogram(EH), 외형 그룹에는 Region Shape(RS), Contour Shape(CS)등 총 8개의 기술자가 포함되어 있다. 표준안에는 각 기술자들에 대한 기술 방법은 정의되어 있지만, 비교 방법은 정의되어 있지 않다. 따라서 각 기술자의 비교 방법은 사용자의 목적에 따라 달라질 수 있다.

하지만 XM[5]문서에는 실험 모델로서 비교 방법을 상세히 제시하고 있으며, 이러한 실험 모델은 표준화 과정에서 사용된 것으로서 각 기술자의 특성을 가장 잘 반영하고 있다고 볼 수 있다. 따라서 본 논문에서는 XM에 제시된 비교 방법에 준하여 각 기술자를 분석하였다.

**3. 인덱싱을 위한 MPEG-7 시각 정보 기술자 분석**

<표 1>은 인덱스 구성 관점에 있어서 각 기술자의 특징을 보여 주고 있다.

**표 1 MPEG-7 시각 정보 기술자의 특징**

Descriptor	Dimensions	Matching	Space
DC	6~48	Like quad dist	No space
SC	16~256	L1	Vector
CL	2~192	weighted L2	Vector
CS	32~256	L1	Vector
HT	62	weighted L1	Vector
EH	150	weighted L1	Vector
RS	35	L1	Vector
CS	2~130	Like L2	No Space

<표 1>에서 보는바와 같이 이미지 검색을 위해 사용할 수 있는 대부분의 MPEG-7 시각정보 기술자들은 수십 차원 이상의 고차원으로 구성되어 있으나 벡터 공간에 정의되어 있어 기존의 인덱스 방법들을 적용하기에 적합하다. 하지만 Dominant Color와 Contour Shape는 벡터 공간이나 메트릭 공간에 표현할 수 없기 때문에 기존의 인덱싱 방법을 바로 적용할 수 없다는 문제점이 있다. 따라서, MPEG-7 시각 정보 기술자에 대한 인덱스를 구성하기 위해서는 Dominant Color와 Contour Shape를 벡터 공간이나 메트릭 공간으로 옮겨놓는 일이 반드시 선행되어야 한다.

**3.1 Dominant Color 비교 방법의 변형**

Dominant Color 기술자는 임의의 영역에 대한 주요 색들의 집합을 의미한다. 최대 8가지 색으로 표현할 수 있으며,  $F = \{(c_i, p_i, v_i), s\}$  ( $i = 1, 2, \dots, N$ )로 정의할 수 있다. N은 색상의 개수를 의미하며,  $c_i$ 는 색상 값을 의미한다.  $p_i$ 는 이미지나 이미지 영역에서  $c_i$ 가 차지하고 있는 영역의 비율을 의미하며,  $v_i$ 는 옵션 정보로서 해당 색상 영역내에서의 분산을 의미한다. s는 이미지 전체에서 Dominant color들의 공간적 일관성을 나타낸다. 두 기술자  $F_1, F_2$ 에 대한 비유사도는 <식 1>과 같이 표현된다. 이를 벡터공간에서 표현할 수 없음을 분명하다.

$$D^2(F_1, F_2) = \sum_{i=1}^{N1} p_{1i}^2 + \sum_{j=1}^{N2} p_{2j}^2 - \sum_{i=1}^{N1} \sum_{j=1}^{N2} 2a_{i,2j} p_{2i} p_{2j} \tag{식 1}$$

$$a_{k,l} = \begin{cases} 1 - \|c_k - c_l\| / d_{max} & d_{k,l} \leq T_d \\ 0 & d_{k,l} > T_d \end{cases}, \text{ where } d_{max} = \alpha T_d$$

또한  $D^2(F_1, F_2)$ 은 음수항이 양수항보다 더 커져서, 메트릭 공간 존재의 조건 중 하나인 non-negativity를 만족하지 않을 수 있다. 일반적으로 벡터 공간 조건을 만족하면 메트릭 공간 조건도 만족하므로, 먼저 메트릭 공간 조건을 만족하도록 변형해본다. 우선 시도할 수 있는 것은 식 자체를 변형하지 않고 추출 과정의 조건이나 임계값들을 합리적으로 변형하여 메트릭 공간 조건을 만족시킬 수 있는지의 여부를 살피는 것이 중요하다.

만일 하나의 기술자안에서 서로 다른 색상들간 거리가  $2T_d$  이상 떨어져 있으면,  $F_1$ 의 각 색상에 대해서  $F_2$ 의 색상중 최대 1개의 색상만이 매칭이 되므로 non-negativity 조건을 충족시킬 수 있다. 두 번째로  $D^2(F_1, F_2)$ 는  $p$  벡터간의 Euclidean 거리 형태를 띄고 있으나, 음수항의 가중치인  $a_{k,l}$ 는  $p$  벡터와는 독립된 의미를 가지고 있는  $c$  벡터를 사용하여 계산되기 때문에 임계값등의 변형을 통해서 Triangle Inequality를 만족하도록 하는 것은 매우 어려운 일이다. 또한 매칭 함수 자체를 변형하더라도 원 함수의 의미를 유지하자면 색상에 대한 고려가 필요하다. 때문에 역시 같은 문제에서 벗어나기는 쉽지 않다. 따라서 매칭 함수의 변형이 아닌 매칭 함수의 의미를 유지하면서 전혀 다른 형태로의 변형을 고려해야 한다.

전술한 바와 같이  $D^2(F_1, F_2)$ 는 각 기술자의 색상들간에 거리가 가까운  $p$  쌍에 대한 L2 Distance의 합의 형태로 표현되어 있음을 관찰할 수 있다. 즉, 색상  $c$ 는  $p$ 의 위치를 결정 지으며, 같은 곳에 위치한  $p$  간의 L2 Distance의 합을 의미한다. 따라서 Dominant Color는  $c$ 를 bin(bin)의 구간으로,  $p$ 를 bin의 값으로 사용하는 히스토그램 형태로 만들어, 두 히스토그램간의 L2 Distance를 구하는 형태로 변형이 가능하다. bin의 개수는  $c$ 의 형태에 따라 달라질 수 있으며 다음과 같이 표현할 수 있다.  $c$ 를 표현하기 위해  $n$  차원을 사용한다고 했을 때, 각 차원을  $c_{d1}, c_{d2}, \dots, c_{di}, \dots, c_{dn}$ 이라 표현하자.  $di$  차원의 최소값을  $\min(c_{di})$ , 최대값을  $\max(c_{di})$ 라 하고, 색상에 대한 양자화(Quantization) 값을  $Q(c_{di})$ 라 했을 때,  $di$  차원을 위한 bin의 개수는 <식 2>와 같이 결정할 수 있다. 각 차원은 서로 독립이므로 전체 bin 개수는 <식 3>와 같이 표현할 수 있다.

$$\left\lceil \frac{\max(c_{di}) - \min(c_{di})}{Q(c_{di}) \times T_d} \right\rceil \tag{식 2}$$

$$\prod_{i=1}^n \left\lceil \frac{\max(c_{di}) - \min(c_{di})}{Q(c_{di}) \times T_d} \right\rceil \tag{식 3}$$

예를 들어, RGB 3차원으로  $c$ 를 표현하고 양자화 값을 4,  $T_d$ 를 10로 정한다면, 각 차원을 위한 bin은 7개이고, 전체 bin의 개수는  $7^3$ 이 된다.  $c_i$ 에 해당되는 bin에  $p_i$  값을 할당하여 변형을 마무리 한다. 하지만 서로 인접한 bin에 존재하는  $c_i$  간의 거리도  $T_d$ 보다 작을가능성이 있기 때문에  $c_i$ 에 해당되는 bin의 인접 bin에도  $p_i$  값을 확장 할당하여 오차의 폭을 줄일 수 있다. 예를 들어, 3차원의 색상일 경우  $(x, y, z)$ 에  $c_i$ 가 해당된다면,  $(x+1, y, z)$ ,  $(x, y+1, z)$ ,  $(x, y, z+1)$ ,  $(x+1, y+1, z)$ ,  $(x, y+1, z+1)$ ,  $(x+1, y, z+1)$ ,  $(x+1, y+1, z+1)$ 에도  $p_i$  값을 할당한다. 최종 거리

를 구할 때에는,  $2^2$ 으로 L2 Distance 결과를 나누어 주어야 한다. 정리하면, 각 Dominant Color를 히스토그램으로 변형한 후 L2 Distance를 취하면  $D(F_1, F_2)$ 에 근사한 값을 산출하면서 벡터 공간에서 표현이 가능하다. <식 4>는 이를 표현한 것이다. 물론 변형된 비교 방법은 원 함수와 동일한 결과를 만들어내지 않는다. 하지만 기대값을 사용한 확률적인 분석을 해보면 평균

적으로 원함수 결과와 약 0.03정도의 차이가 발생할 수 있음을 증명할 수 있다.

$$\frac{L2(Histogram\_F^1, Histogram\_F^2)}{\frac{n}{2^2}} \cong D(F^1, F^2) \quad \text{<식 4>}$$

### 3.2 Contour Shape 비교 방법의 변형

이 기술자는 전역 파라미터 (Global Parameter)인 Circularity와 Eccentricity, 그리고 CSS(Curvature Scale Space) peak 정보로 구성되어 있다. 두 Contour Shape 기술자간의 비교는 두 단계로 수행되는데, <식 5>을 만족하는 것들에 대해서만 <식 6>을 사용하여 거리를 구한다.  $c_q[0]$ 와  $c_r[0]$ 은 각각 질의 기술자와 대상 기술자의 Eccentricity를 의미하고,  $c_q[1]$ 와  $c_r[1]$ 은 Circularity를 의미한다.  $Th\_e$ 와  $Th\_c$ 는 각 파라미터의 임계값이다.

$$\frac{|c_q[0]-c_r[0]|}{MAX(c_q[0], c_r[0])} \leq Th\_e \quad \frac{|c_q[1]-c_r[1]|}{MAX(c_q[1], c_r[1])} \leq Th\_c \quad \text{<식 5>}$$

$$D(q,r) = 0.4 \times \frac{|c_q[0]-c_r[0]|}{MAX(c_q[0], c_r[0])} + 0.3 \times \frac{|c_q[1]-c_r[1]|}{MAX(c_q[1], c_r[1])} + Mcss$$

$$Mcss = \sum_1 ((xpeak[i]-xpeak[j])^2 + (ypeak[i]-ypeak[j])^2) + \sum_2 (ypeak[i])^2$$

<식 6>

$Mcss$ 는 CSS peak 사이의 L2 Distance를 의미하며 첫 번째 항은  $xpeak$ 의 거리가 0,1 이하인 peak들을 위한 계산이고, 두 번째 항은 매칭 되지 않은 peak들에 대한 페널티를 의미한다.

Dominant Color와 마찬가지로 각 차원의 데이터에 대한 비교 방법이 서로 다르기 때문에 Contour Shape는 벡터공간에 존재하지 않음은 분명하다. 또한 <식 6>을  $Mcss$ 를 제외한 부분과  $Mcss$ 의 두 부분으로 나누어 분석하면, 전자는 메트릭 공간에 존재함을 증명할 수 있으나,  $Mcss$ 는 매칭 되는 것과 그렇지 않은 것에 대한 계산 방법이 다르기 때문에 Triangle Inequality 조건에 위배되어 메트릭 공간에 존재하지 않음을 알 수 있다. 따라서 인덱싱을 위해서는  $Mcss$ 를 메트릭 공간이나 벡터 공간으로 이동해야 한다. 먼저 생각할 수 있는 방법은 Dominant Color와 마찬가지로, 임계값이나 각 변수들에 합리적인 제한을 두어서 메트릭 공간 조건을 만족하게끔 하는 것이다. 하지만  $Mcss$ 가 이러한 조건을 만족하기 위해서는 기술자의 핵심인 peak 값 자체의 범위를 상당 부분 제한해야 하기 때문에 이를 합리적인 방법이라 하기는 어렵다. 두 번째로 생각할 수 있는 방법은 매칭 함수를 변형하는 것이다.  $Mcss$ 에서 매칭의 의미를  $xpeak$ 의 거리가 0,1 이하인 것이 아닌,  $(xpeak, ypeak)$ 쌍의 거리가 최소인 것으로 정의해보자. 매칭되는 쌍들에 대하여 L2 Distance를 사용하면 페널티 항이 없어지기 때문에 Triangle Inequality 조건을 만족시킬 수 있다. 하지만 이 방법은 실험 결과로 봤을 때 사용자 만족도에 있어서는 의미있는 검색 결과가 나오지만, 원 함수 결과와 상당히 다른 결과를 산출하는 문제점이 있다. 마지막 방법은 벡터 공간으로의 변형이다.  $xpeak$ 을 빈 구간으로 하고,  $ypeak$ 을 구간 값으로 정의하여 히스토그램 형태로 표현할 수 있다. 이렇게 하면 매칭이 되지 않는  $ypeak$ 들에 대한 페널티의 의미를 살릴 수 있지만  $xpeak$  관련 항이 없어지기 때문에 역시 어느정도의 오차는 존재한다.

### 4 실험 및 결과

Berkeley에서 제공하는 25,000 장의 풍경 이미지를 사용하여 실험하였다. Dominant Color는 7bit로 양자화 된 LUV 3차원 색상 공간을 사용하였고,  $T_d$ 는 10을 사용하였다. Contour Shape는 오브젝트 추출 과정이 선행되어야 하기 때문에 대량의 테스트 기술자를 만들기에 큰 어려움이 있다. 랜덤한 데이터 셋을 사용하여 실험하였으나 실험 결과의 신뢰도가 떨어지기 때문에 실험 결과에 포함하지 않았다. 임의로 1000개의 질의 이미지를 선택하여 실험한 결과는 다음과 같다. 원 함수에서 100위까지의 결과는 변형된 비교 방법을 사용하여 10배수를 뽑았을 경우 그 안에 평균 93%이상 존재했으며, 평균 차이는 거리를 1로 정규화 했을 경우 약 0.035에 불과하였다. <그림 1>은 100번의 실험에 대한 결과를 보여주는 그림이다. 가로 축은 실험 횟수를 의미하며, 위쪽에 있는 그래프가 포함 비율, 밑에 있는 그래프가 오차의 크기를 의미한다.

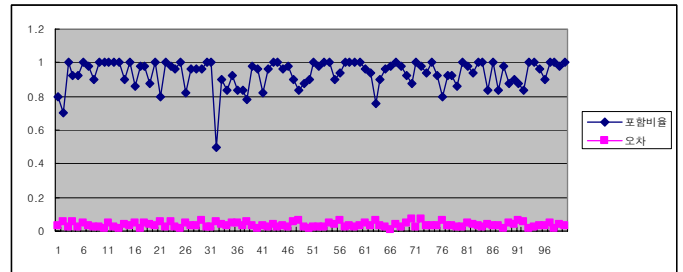


그림 1 원 함수와 히스토그램 방법과의 비교

### 5 결론 및 앞으로의 연구 방향

본 논문에서는 기존의 여러 인덱싱 방법을 MPEG-7 시각 정보 기술자에 적용할 수 있도록 하기 위해서 Dominant Color와 Contour Shape의 매칭 함수를 변형하였다. 변형된 두 기술자에 대해서는 결과면에 있어서 어느정도의 오차는 존재하나, 그 결과 이미지 검색시 필요한 모든 MPEG-7 시각 정보 기술자에 대해서 인덱싱이 가능해졌다. 다음 단계로 각 기술자의 특성을 분석하여, 기존의 방법중 최적의 인덱싱 방법을 찾거나 새로운 인덱싱 방법을 개발하여 실질적으로 검색 속도를 향상시키고자 한다.

### 6. 참고 문헌

- [1] ISO/IEC JTC1/SC29/WG11 *Information Technology Mul-timedia Content Description Interface-Part3: Visual*, 2001.
- [2] A.Guttman, " R-Trees: A Dynamic Index Structure for Spatial Searching," *Proc. of the ACM SIGMOID Conf.*, pp.47- 57, 1984.
- [3] R.Weber, H.Schek, and S.Blott, " A Quantitative Analysis and Performance Study for Similarity-Search Methods in High-Dimensional Spaces," *Proc. of VLDB Conf.*, pp.194-205, 1998.
- [4] P.N.Yianilos, " Data Structures and Algorithms for Nearest Neighbor Search in General Metric Spaces," *Proc. of the ACM-SIAM Conf.*, pp.311-321, 1993.
- [5] ISO/IEC JTC1/SC29/WG11 *MPEG-7 Visual part of eXperiment Model Version 11.0*, 2001.