

# VQ를 이용한 영상의 객체 특징 추출과 이를 이용한 내용 기반 영상 검색

(Representative Feature Extraction of Objects using VQ and  
Its Application to Content-based Image Retrieval)

장 동 식 <sup>†</sup> 정 세 환 <sup>\*\*</sup> 유 현 우 <sup>\*\*\*</sup> 손 용 준 <sup>\*\*\*\*</sup>

(Dong-Sik Jang) (Seh-Hwan Jung) (Hun-Woo Yoo) (Yong-Jun Sohn)

**요약** 내용 기반 영상 검색을 위해 본 연구에서는 VQ(Vector Quantization)을 이용하여 영상을 구성하는 주요 객체들의 특징 추출 방법을 제안한다. 내용 기반 영상 검색 시스템에서 사용되는 영상의 주요 특징으로는 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 있다. 이 중 본 논문에서는 일반적인 색상 및 질감 특징 추출방법과 더불어 VQ 벡터 클러스터링 알고리즘을 이용하여 정지 영상을 구성하고 있는 객체들의 대표 색상과 질감 특징을 빠르게 추출하고 이를 내용 기반 검색에 이용함으로써 정지영상의 내용에 근거한 검색을 하였고 객체 단위 검색을 함으로써 객체의 위치, 회전 및 크기 변화에 무관한 검색을 가능케 했다. 연구의 실험 결과 VQ를 이용함으로써 대표특징치 추출시간을 줄일 수 있었고 검색시 색상과 질감 특징의 가중치를 각각 0.5, 0.5로 주는 것이 가장 높은 검출율을 보였으며, '사람' 영상에 제안한 방법을 적용한 경우 90%의 검출율을 보였다.

**Abstract** In this paper, a new method of feature extraction of major objects to represent an image using Vector Quantization(VQ) is proposed. The principal features of the image, which are used in a content-based image retrieval system, are color, texture, shape and spatial positions of objects. The representative color and texture features are extracted from the given image using VQ(Vector Quantization) clustering algorithm with a general feature extraction method of color and texture. Since these are used for content-based image retrieval and searched by objects, it is possible to search and retrieve some desirable images regardless of the position, rotation and size of objects. The experimental results show that the representative feature extraction time is much reduced by using VQ, and the highest retrieval rate is given as the weighted values of color and texture are set to 0.5 and 0.5, respectively, and the proposed method provides up to 90% precision and recall rate for 'person' query images.

## 1. 서론

최근 컴퓨터와 통신기술의 발달과 더불어 멀티미디어 정보 서비스에 대한 요구가 증가하고 있으며, 인터넷상

의 정지화상 및 동영상검색, 방송제작에 사용될 VOD(Video On Demand)서비스, 전자도서관(Digital Library), 의료분야(원격진료, 의료영상검색) 등 응용될 수 있는 멀티미디어 검색기술에 대한 연구의 필요성이 증가되고 있다.

연구 초기에는 검색의 대상이 되는 모든 멀티미디어 데이터에 사람이 직접 색인을 첨가하고, 사용자 또한 주제를 이용하여 원하는 정보를 검색하는 텍스트기반 검색이 사용되어왔다. 그런데 이 방법은 시간·비용이 많이 들며, 색인을 첨가하는 사람과 검색하는 사용자의 관점이 불일치할 경우 검색의 효율성이 크게 떨어지게 된다. 또한 멀티미디어 데이터가 가지는 복잡한 속성을

<sup>†</sup> 정 회 원 : 고려대학교 산업시스템정보공학과 교수  
jang@korea.ac.kr

<sup>\*\*</sup> 비 회 원 : LG 기술원 Innovation Center Digital Vision 그룹 연구원  
dustm35@korea.ac.kr

<sup>\*\*\*</sup> 학생회원 : 고려대학교 산업시스템정보공학과  
paulyh@korea.ac.kr

<sup>\*\*\*\*</sup> 비 회 원 : 고려대학교 산업시스템정보공학과  
marksohn@korea.ac.kr

논문접수 : 2000년 1월 11일

심사완료 : 2001년 9월 11일

텍스트만으로는 정확하게 표현할 수 없다는 단점을 가지고 있다. 이를 보완하기 위해 멀티미디어 데이터의 내용을 대표할 수 있는 특징을 추출하여, 이를 기반으로 색인과 검출을 수행하는 내용기반 검색방법이 필요하며, 이 방법은 멀티미디어 데이터로부터 특징을 자동으로 추출하여 색인과정에 사용함으로써 데이터베이스 구축에 필요한 시간 및 인력의 소모를 줄일 수 있다는 장점이 있어 현재 많은 연구가 이루어지고 있다. 대표적인 내용 기반 영상 검색 시스템으로는 IBM사의 QBIC[1], Virage사의 Virage[2], 콜롬비아 대학의 VisualSEEK[3], Chicago 대학교의 WebSeer[4], Excalibur사의 Excalibur Visual RetrievalWare (<http://www.excalib.com>), 등이 있다. 위에서 설명한 각각의 시스템들은 나름대로의 장점이 있지만, 각 시스템에 따라서 몇 가지 단점을 가지고 있다. QBIC의 경우, 사용자가 다양한 형태의 특징값을 사용할 수 있지만, 영상분할에 사용자의 수작업이 필요하고, Virage의 경우는 너무 일반적인 영상검색시스템이므로 구체적인 적용시에는 새로 방법론을 고려해야 하며, VisualSEEK의 경우는 영상내의 객체 또는 영역의 공간 정보를 이용한 검색에 적합하지만 색상 이외에 중요한 질감이나 모양정보를 사용하고 있지 않다. 그 외 WebSeer와 Visual RetrievalWare 등은 영상내부에 존재하는 객체에 대한 검색이 아닌 일반적인 전역적 특징을 이용한 검색만을 지원한다.

내용 기반 영상 검색을 위해 한 영상을 구성하는 객체들을 정확히 분할(segmentation)하여 분할된 각 객체들을 대표하는 대표특징을 추출한 후 이를 영상 검색에 이용한다면 영상에서 검색을 원하는 객체에 대한 명확한 질의가 가능할 것이다. 그런데 실제로 한 영상을 구성하는 객체들을 정확하게 분할하기란 쉽지 않다. 그래서 영상내의 한 객체는 유사한 색상과 질감특징을 갖는다고 보아 한 영상에서 추출한 특징치들을 EM (Expectation-Maximization)[5]이란 패턴 분류 알고리즘을 통해 클러스터링(clustering)하여 이를 검색에 이용하는 Region-Based Query[6]에 관한 연구가 진행 중이다. 그런데 EM 알고리즘은 클러스터링 될 클러스터 수를 미리 정해주어야 하며, 인의의 여러 클러스터 수에 대한 클러스터링 결과 중 영상을 구성하는 주요 객체 수만큼 가장 적절하게 분류된 클러스터 수를 선정해야하므로 객체의 대표 특징을 추출하는데 비교적 많은 시간이 소요되고 한 영상의 내용을 적절하게 표현하는 대표 특징치를 추출하기가 쉽지 않다. 본 연구에서는 패턴의 분포에 따라서 클러스터 수를 자동적으로 결정하는 VQ(Vector Quantization)알고리즘[7]을 이용하여

클러스터링 하여, 이 클러스터들 중 많은 화소멤버를 가진 것을 주요 특징치로 하여 같은 Region-based Query를 하는 EM알고리즘보다 빠르게 추출하였다. 그리하여 한 영상의 내용을 적절하게 표현하는 대표 특징치를 추출하고, 이를 영상검색에 이용하여 높은 검출율을 보였다. 2절에서는 색상 특징추출과 질감 특징추출에 대하여 설명하였고, 3절에서는 영상의 대표 객체 특징추출을 위해 사용하는 VQ알고리즘을 설명하였다. 그리고 4절에서는 대표 특징추출을 이용한 영상 검색을 설명하고, 5절에서는 실험 및 결과를 기술하였다.

## 2. 영상 특징 추출

내용 기반 영상 검색을 위해서는 멀티미디어 데이터의 내용을 대표할 수 있는 특징을 추출해야 하며, 이를 기반으로 색인과 검색을 수행해야 하는데 일반적으로 사용되고 있는 영상의 주요 특징으로는 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 있다.

내용 기반 영상 검색의 초기부터 사용되어온 영상의 색상 정보는 영상의 내용을 가장 잘 대표하는 특징이라 할 수 있으며, 영상내의 조명 변화나 관측 위치, 크기 변화 등에 어느 정도 무관하게 적용 가능하다는 장점을 갖고 있다. 색상 정보를 사용하는 검색 방법의 경우에는 색상 공간상에서의 거리 측정을 정확하게 하기 위하여 일반적인 색상 공간인 RGB모형을 사용하는 대신, HSV, YIQ, YUV 등과 같은 모형을 사용한다[8,9]. 이 중에서도 특히

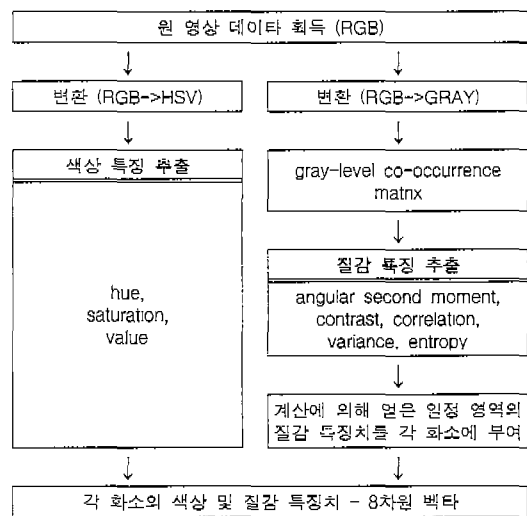


그림 1 색상 및 질감 특징 추출 과정

HSV모델은 인간의 시각 능력에 유사한 색상 모델로서, 영상의 색상 정보 추출에 많이 사용된다[10].

영상의 색상 정보와 더불어 매끄러운, 거친, 나뭇결 무늬, 벽돌 무늬 등과 같은 질감 정보를 특징으로 이용하면 더 좋은 검색 결과를 기대할 수 있다. 따라서, 화소단위 색상 정보 이외에 영역단위의 질감 정보를 추출하여 검색에 사용한다. 주로 이용되는 질감 특징으로는 Co-occurrence features[11], Gabor filters[12], Markov Random Field features, Fractal features 등이 있다[13].

본 연구에서는 색상(color) 특징 추출을 위해 HSV 색상 모델을 사용했으며, 질감(texture) 특징 추출을 위해서는 Co-occurrence matrix를 사용했다. 정지영상의 특징 추출과정은 그림 1과 같다.

2.1 색상 특징 추출

색상 특징으로는 RGB모델을 HSV모델로 변환하여 얻은 색도(hue),채도(saturation),명도(value)를 이용했다[8,9]. 그림 2의 HSV 색상 공간에서 보는 바와 같이 색도는 0°~360°의 육십분법으로 나타내며 채도와 명도는 0~1사이의 값으로 나타내어진다. 채도는 순수도를 나타내며 HSV 색상 공간에서 쌍원뿔 표면에 가까이 위치할수록 색상은 순수하며, 반대로 명도 표시축에 가까이 위치할수록 탁하다. 명도는 밝기정도를 나타내며 쌍원뿔의 아래방향으로 내려갈수록 어둡고 위로 올라갈수록 밝다.

색도 18개, 채도와 명도는 각각 3개씩으로 하며 그레이(gray) 5개를 추가하여 색상 특징치를 167개로 양자화(quantization)하여 사용함으로써 색상특징치를 단순화하였다. RGB모델에서 HSV모델로의 변환식은 식 (1) -식 (3)과 같다.

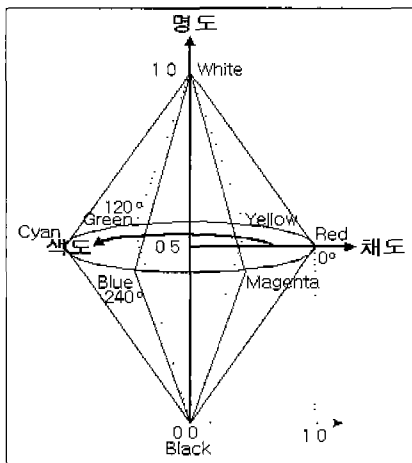


그림 2 HSV 색상공간

$$H = \begin{cases} \theta, & G \geq B \\ 2\pi - \theta, & G < B \end{cases} \quad (1)$$

여기서  $\theta = \cos^{-1} \left[ \frac{\frac{1}{2} [(R-G) + (R-B)]}{\left[ \frac{(R-G)^2 + (R-B)(G-B)}{2} \right]^{\frac{1}{2}}} \right]$

$$S = 1 - \frac{3}{R+G+B} [\min(R, G, B)] \quad (2)$$

$$V = \frac{1}{3} (R + G + B) \quad (3)$$

2.2 질감 특징 추출

일반적으로 영상의 질감은 그레이 화소값 들의 모멘트(moment), 대비(contrast), 연관성(correlation), 분산(variance), 혼잡도(entropy)를 이용해서 가장 잘 표현할 수 있으며, 이러한 특징치 들은 Th. Hermes[14]의 영상시스템에 적용되어 좋은 결과를 얻은 바 있다. 그러므로, 이 연구에서는 co-occurrence matrix를 이용하여 ASM(Angular Second Moment), contrast, correlation, variance, entropy의 5개의 질감특징을 계산한다. ASM과 entropy는 영상의 균일성(homogeneity)을, contrast는 영상 안에서 나타나는 대조성과 지역적 변동성을, correlation은 gray-tone의 선형적 의존성(linear-dependency)을, variance는 인접 화소의 gray-level 차에 대한 산포도를 나타낸다[6] [9]. 질감 특징을 얻기 위해 한 영상을 7×7, 즉 49개의 영역으로 나누어 각 블록에 대해 먼저 정규화된 4방향(0°, 90°, 45°, 135°) gray-level co-occurrence matrix를 계산한 후 이들 matrix로부터 질감 특징치를 추출하여 평균함으로써 회전전에 무관한 질감 특징치를 얻었다.

먼저 정규화(normalization)되기 전 matrix의 각 엔트리(entry)의 값(frequency)은 아래와 같은 정의에 의해 얻을 수 있다.

$$P(i, j, d, 0^\circ) = \frac{\text{num} \{ ((k, l), (m, n)) \in (Ly \times Lx) \times (Ly \times Lx) \mid k-m=0, |l-n|=d, I(k, l)=i, I(m, n)=j \}}{\text{total}}$$

$$P(i, j, d, 45^\circ) = \frac{\text{num} \{ ((k, l), (m, n)) \in (Ly \times Lx) \times (Ly \times Lx) \mid (k-m=d, l-n=-d) \text{ or } (k-m=-d, l-n=d), I(k, l)=i, I(m, n)=j \}}{\text{total}}$$

$$P(i, j, d, 90^\circ) = \frac{\text{num} \{ ((k, l), (m, n)) \in (Ly \times Lx) \times (Ly \times Lx) \mid |k-m|=d, l-n=0, I(k, l)=i, I(m, n)=j \}}{\text{total}}$$

$$P(i, j, d, 135^\circ) = \frac{\text{num} \{ ((k, l), (m, n)) \in (Ly \times Lx) \times (Ly \times Lx) \mid (k-m=d, l-n=d) \text{ or } (k-m=-d, l-n=-d), I(k, l)=i, I(m, n)=j \}}{\text{total}}$$

$P(i, j, d, \text{방향})$ : 각 방향에 대한 matrix i행 j열 entry의 값  
 $Lx = \{1, 2, \dots, Nx\}$ ,  $Ly = \{1, 2, \dots, Ny\}$ ,  $G = \{1, 2, \dots, NG\}$   
 image I:  $Ly \times Lx \rightarrow G$

$Nx$ : 영상의 가로방향 픽셀수

$N_y$ : 영상의 세로방향 픽셀수

$N_g$ : gray levels

$d$ : 계산에 고려할 이웃픽셀과의 거리

$num$ : 정의된 만족하는 개수

matrix의 각 엔트리 값을 계산한 후 각 방향에 대해 아래와 같은 R값으로 각 엔트리를 나눠줌으로써 정규화 하여 4방향( $0^\circ$ ,  $90^\circ$ ,  $45^\circ$ ,  $135^\circ$ )에 대한 각각의 gray-level co-occurrence matrix를 얻을 수 있다.

$$0^\circ : R = 2Ly(Lx - 1)$$

$$90^\circ : R = 2Lx(Ly - 1)$$

$$45^\circ : R = 2(Ly - 1)(Lx - 1)$$

$$135^\circ : R = 2(Lx - 1)(Ly - 1)$$

다음은 위에서 얻은 matrix를 이용하여 질감특징을 얻는다. 각 질감 특징치를 계산하는 식은 식 (4)-식 (8)과 같다.

$$Angular\ Second\ Moment = \sum_i \sum_j \{p(i, j)\}^2 \quad (4)$$

$$Contrast = \sum_{n=0}^{n-1} n^2 \left\{ \sum_{i=1}^n \sum_{j=1}^n p(i, j) \right\}, li - j = n \quad (5)$$

$$Correlation = \frac{\sum_i \sum_j \{ij\} p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (6)$$

$$Variance = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (7)$$

$$Entropy = - \sum_i \sum_j p(i, j) \log p(i, j) \quad (8)$$

여기서  $p(i, j)$  : gray-level co-occurrence matrix의  $(i, j)$ 엔트리

$N_g$  : gray level

$\mu_x, \mu_y, \sigma_x, \sigma_y$  :  $p_x$ 와  $p_y$ 의 평균과 표준편차,

$\mu$  : gray-level co-occurrence matrix의 평균

$p_x$  :  $p(i, j)$ 의 행을 합하여 얻은 주변 확률 matrix

$p_y$  :  $p(i, j)$ 의 열을 합하여 얻은 주변 확률 matrix

$$\sum_i = \sum_{i=1}^{N_g} \sum_j = \sum_{j=1}^{N_g}$$

색상특징은 각 화소(pixel)로부터 추출하며, 질감특징은 인접화소간의 관계를 통해서 추출하므로 색상특징은 화소단위 특징인 반면, 질감특징은 지역단위 특징이다. 따라서 VQ를 이용하기 위해서는 특정영역의 질감특징치를 영역 안의 모든 화소에 부여하는 방법을 통하여 질감특징도 화소단위 특징치를 갖도록 변환한다.

본 연구에서는 각 블록의 특징치 5개를 그 블록 안의 모든 화소(pixel)에 부여함으로써 색상 특징치와 더불어 각 화소는 8차원(색상 특징치 3개, 질감 특징치 5개) 특징치 벡터를 갖게되며 이는 VQ를 이용한 클러스터링 과정에서 입력 데이터로 이용된다.

### 3. 대표 특징치 추출

#### 3.1 VQ(Vector Quantization)

VQ(Vector Quantization)[7]는 비지도 클러스터링 알고리즘(unsupervised clustering algorithm)으로서 입력 데이터들을 여러 개의 의미 있는 클래스로 나눈다. K-means algorithm[7]과 같이 클러스터링(clustering)하기 전에 분류될 클러스터(cluster)의 개수를 미리 정해줄 필요 없이 데이터의 분포에 적합하게 자동적으로 데이터를 분류한다. 즉, 기존에 Region-Based Query[6]를 위해 클러스터링 알고리즘으로 이용된 EM(Expectation Maximization)[3,4] 알고리즘이 분류될 클러스터의 개수를 미리 정해주어야 하기 때문에 EM을 영상 검색을 위한 데이터 분류 알고리즘으로 사용하기 위해서는 여러 클러스터 개수의 클러스터링 결과 중 영상에서 추출한 데이터를 가장 적절하게 클러스터링한 결과를 선정해야 하는데 이런 어려움을 VQ를 이용함으로써 해결할 수 있다. ART (Adaptive Resonance Theory)[7]가 VQ와 마찬가지로 클러스터를 동적으로 할당하면서 클러스터링 하지만 neural approach인데 비해 VQ는 nonneural method이다. VQ 알고리즘에는 두 가지의 단점이 있다. 먼저 입력되는 데이터의 순서에 민감하다는 점이다. 즉, 클러스터링을 위해 입력되는 데이터의 순서가 바뀔 경우 클러스터링의 결과에 다소 차이가 있을 수 있다. 그러나 영상에서 추출한 데이터를 화소(pixel)의 위치 순서대로 VQ를 이용하여 클러스터링 해본 결과, 다른 순서를 이용하여 클러스터링 한 결과와 큰 차이 보이지 않았고, 영상을 대표하는 객체의 특징치를 잘 추출하고 있었다. 두 번째로 VQ를 이용하기 전에 클러스터링 과정에 사용할 클러스터간의 한계 거리를 임계값(threshold)으로 미리 정해주어야 하는데, 데이터의 차원(dimension)과 특성, 그리고 데이터가 갖는 범위 등을 고려하여, 적절한 임계값을 테스트를 통하여 결정하여 준다. 이는 임계값을 너무 작게 하면 한 객체에 대해 여러 개로 클러스터링 되고, 임계값을 너무 크게 하면 여러 객체의 특징이 한 클러스터로 클러스터링 되는 좋지 않은 결과를 낳는다. 그래서 가장 적절한 임계값을 설정하는 것이 중요한데, 각 영상을 구성하는 객체들마다 필요로 하는 임계값은 다르기 때문에 이를 모두 만족하는 임계값 선정은 불가능하다. 본 연구에서는 테스트를 통해 가장 좋은 결과를 보이는 0.4를 임계값으로 통일하여 사용하였다.

VQ에 의한 데이터의 클러스터링 알고리즘은 다음과 같다.

step 0: 임계값이 미리 정하여 진다.  
 step 1: 클러스터링 될 n-차원 벡터 데이터가 존재하면 step 2로 진행한다.  
 그렇지 않으면 클러스터링 알고리즘을 끝낸다.  
 step 2: 여러 클러스터 중에서 가장 가까운 거리를 갖는 클러스터를 찾기 위하여 식 (9)와 (10)에 의하여 계산된 k번째 클러스터를 결정한다. k번째 클러스터와의 거리가 임계값보다 적으면 k번째 클러스터에 포함시킨다. 그리고 k번째 클러스터의 중심값을 식 (11)에 의하여 새롭게 계산한다.  
 만약 임계값보다 크면 클러스터가 존재하지 않으므로 입력 데이터를 중심값으로 하는 새로운 클러스터를 생성한다. 즉, 첫 번째 입력 데이터는 첫 클러스터의 중심값이 된다. step 1로 이동한다.  
 위의 알고리즘은 그림 3에 나타나 있다.

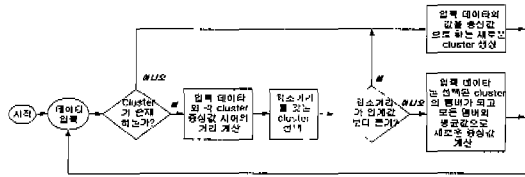


그림 3 VQ 알고리즘

$$d = \|X^{(n)} - C_j\| = \left[ \sum_{i=1}^N (x_i^{(n)} - c_{ij})^2 \right]^{1/2} \quad (9)$$

$$\|X^{(n)} - C_k\| < \|X^{(n)} - C_j\|, \quad \text{여기서 } \begin{cases} j=1, \dots, M \\ j \neq k \end{cases} \quad (10)$$

$$C_k = \frac{1}{N_x} \sum_{x \in S_k} X \quad (11)$$

여기서  $X^{(n)}$ : p번째 입력벡터,  $N$ : 벡터 차원

$C_j$ : j번째 cluster의 중심값

$C_k$ : 입력 데이터와 가장 가까운 cluster의 중심값

$N_x$ : 클러스터 포함 멤버(데이터)수,

$S_k$ : k번째 클러스터 집합

이와 같이 VQ는 분류될 데이터의 클러스터링 개수를 미리 정하여 주지 않아도 적절한 클러스터링 개수만큼 데이터를 자동적으로 분류한다. 이런 특징을 본 연구에 사용함으로써 EM을 비롯한 다른 알고리즘에 비해 영상을 구성하는 주요 객체의 수만큼 비교적 유연하게 대표 특징을 추출할 수 있었고 특징 추출시간을 줄일 수 있었다.

**3.2 VQ를 이용한 객체의 대표 특징치 추출**

한 화소당 3가지 색상 특징치(색도, 채도, 명도)와 5가지 질감 특징치(angular second moment, contrast,

correlation, variance, entropy), 즉, 8차원 벡터의 특징치를 추출하였는데 이것은 객체의 대표 특징치 추출을 위한 데이터로 사용된다. 즉, 그림 4와 같이 한 영상의 각 화소가 갖는 특징치 벡터가 입력데이터가 되고 이 데이터는 VQ에 의해서 클러스터링 되어 최종적으로 영상을 구성하는 각 객체들의 대표 특징치가 얻어진다. 이는 영상 검색에 활용된다.

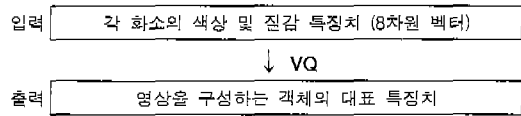


그림 4 VQ 이용시 입력 데이터와 출력값

영상 데이터로 가로 192 화소, 세로 128 화소 영상을 사용했다. 따라서 영상이 갖는 총 화소 수는  $192 \times 128 = 24,576$ 개이므로 VQ 알고리즘 이용을 위한 입력 벡터 수는 24,576개이다. 영상에서는 인접 화소간의 중복성(redundancy)에 의하여 질감 특징과 색상특징이 비슷하므로, 모든 화소의 특징치 벡터를 사용하는 것은 VQ를 알고리즘을 사용하는데 있어 많은 입력 벡터수로 인해 불필요하게 소요시간만 늘릴 뿐이다. 그래서 영상에서 수평과 수직방향으로 두 화소씩 건너 뛰면서 입력 벡터로 사용할 화소의 특징치를 샘플링(sampling)하여 한 영상 당 2,752개의 입력 벡터만 사용함으로써 VQ 이용을 위한 입력 벡터 수를 약 1/9로 줄였다. 그림 5 영상에서 VQ를 이용하여 추출된 주요 객체 4개의 대표 특징치는 표 1과 같다. 그림 5 영상을 구성하는 주요 객체 하늘, 집, 짚, 잔디에 관한 대표 특징치가 8차원 벡터(색상 특징치 3차원, 질감 특징치 5차원)로 나타남을 볼 수 있다.

표 1 객체의 대표 특징치 추출 예

객체	하늘	집	짚	잔디	
클러스터	1	2	3	4	
멤버수	748	422	370	290	
중심값	색도	0.5653	0.0576	0.0904	0.3319
	채도	0.3820	0.3025	0.4357	0.2440
	명도	0.4300	0.4014	0.6291	0.3396
	ASM	0.0151	0.0017	0.0013	0.0030
	Contrast	0.0065	0.0294	0.0399	0.0193
	Correlation	0.7690	0.6932	0.5969	0.6505
Variance	0.0256	0.0995	0.1006	0.0601	
Entropy	0.4184	0.5901	0.6014	0.5779	



그림 5 대표 특징치 영상

#### 4. 객체의 특징치를 이용한 영상검색

각 정지 영상에서 추출된 객체의 대표 특징치는 DB에 저장되어 영상 검색 과정에서 이용된다. 객체의 대표 특징치를 이용한 영상 검색 과정은 그림 6과 같다.

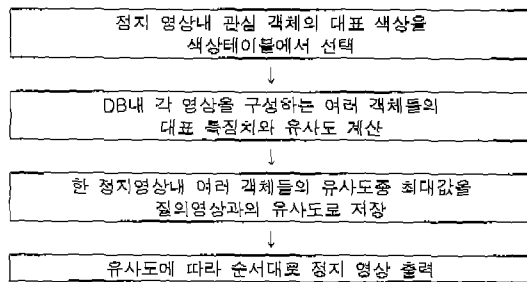


그림 6 영상 검색 과정

본 연구에서는 그림 7과 같이 각 영상마다 정지 영상 내 객체들의 대표 특징치중 주요 대표 특징치 4개를 선정하여 색상테이블에 표현하였다. VQ를 이용한 클러스터링 결과 중 가장 많은 멤버를 포함하는 클러스터 4개를 선정하여 각 클러스터 중심값을 DB에 저장하고 8차원 특징치중 색상특징인 색상, 채도, 명도를 이용하여 색상 테이블에 표현한다. 대표색상 중 검색을 원하는 객체의 대표 색상을 선택함으로써 검색을 시작한다. 대표색상이 선택되면 각 정지영상마다 계산해 놓은 객체들의 대표 특징치들과 유사도 계산을 한다. 유사도 계산을 위해서는 각 특징치를 간의 거리계산이 필요한데 본 논문에서는 식 (12)과 같이 가중치를 준 유클리디언 거리계산법이 이용되었다. 가중치의 결정은 사용하는 DB안의 여러 객체에 대한 테스트를 거쳐 가장 효과적인 것으로 결정한다. 이 연구에서는 여러 테스트를 한 후에 가장 적절한 것으로써, 거리 계산을 위해 가중치로 색상과 질감특징에 각각 0.5로 동일한 가중치를 주었으며, 색상특징에 대해서 색도는 0.44, 채도와 명도는 0.28로 색도에 더 많은 중요도를 부과했고, 질감특징은 다섯 개 모두 0.2로 동일하게 가중치를 주었다. 유사도  $S_{QB}$ 는 식 (13)에 의해서 두 객체 특징치간의 유사도를 0-1 사

이의 값으로 표현했다. 색상과 질감이 전혀 다른 객체인 경우는 0에 가까운 값으로, 동일한 객체간의 유사도는 1로 표현했다.

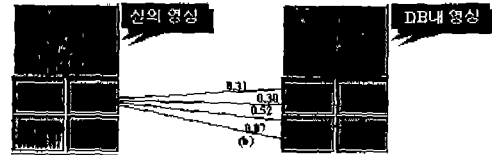


그림 7 영상 검색 과정. (a)질의 영상의 관심 객체 특징치를 색상으로 표현, DB내 저장 특징치는 0.4233:0.2366:0.5722:0.0019:0.0054:0.5864:0.0156:0.5762; (b) DB내 영상의 각 객체들과의 유사도 비교, 3번째 객체가 0.87로 최대값을 나타냄

$$d_{QB} = w_9 [w_1 (H_Q - H_D)^2 + w_2 (S_Q - S_D)^2 + w_3 (V_Q - V_D)^2]^{1/2} + w_{10} [w_4 (ASM_Q - ASM_D)^2 + w_5 (CONT_Q - CONT_D)^2 + w_6 (COR_Q - COR_D)^2 + w_7 (VAR_Q - VAR_D)^2 + w_8 (ETRP_Q - ETRP_D)^2]^{1/2} \quad (12)$$

여기서 Q : 질의영상에 관한 인덱스, D : DB내 영상에 관한 인덱스, H : hue, S : saturation, V : value, ASM : angular second mement, CONT : contrast, COR : correlation, VAR : variance, ETRP : entropy,  $w_1-w_8$  : 각 특징치에 대한 가중치,  $w_9$  : 색상 특징에 대한 가중치,  $w_{10}$  : 질감 특징에 대한 가중치, 실험에서  $W_1=0.44, W_2=W_3=0.28, W_4=W_5=W_6=W_7=W_8=0.2, W_9=W_{10}=0.5$ 로 설정

$$S_{QB} = e^{-\frac{d_{QB}}{2}} \quad (13)$$

#### 5. 실험 및 결과

##### 5.1 실험환경

실험을 위해 펜티엄II333, RAM 64M 사양의 PC를 이용했으며, 실험 데이터로 192×128 크기의 24bit BMP, GIF, JPG 포맷의 영상 600개를 사용하였다.

##### 5.2 실험결과

본 연구에서 세 가지 실험을 하였다. 그 첫 번째로 VQ와 EM의 clustering 소요시간을 비교하여 보았다. EM의 경우 각 영상을 구성하는 객체 수에 알맞게 cluster의 개수를 자동적으로 결정짓지 못하기 때문에 각 영상마다 2~10개의 cluster로 분류하는데 걸리는 시

간을 계산하였다. EM이 VQ에 비해 대략 1.4배정도의 시간이 소요됨을 알 수 있었다. 소요되는 시간은 아래의 그림 8과 표 2와 같다.

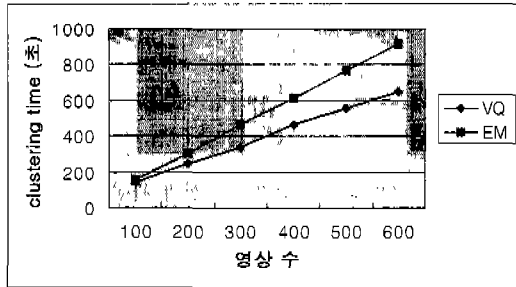


그림 8 clustering 소요시간 비교

표 2 clustering 소요시간 비교

영상 수	VQ(초)	EM(초)
100	139.0	152.9
200	245.5	305.4
300	336.7	462.6
400	467.9	614.9
500	556.5	766.3
600	647.4	915.4

두 번째 실험으로 색상과 질감의 가중치를 변화하여 질의를 해보았다. 아래 표 3에서 보는 바와 같이 색상 특징만을 혹은 질감 특징만을 이용하여 검색한 결과보다 색상과 질감 특징을 혼합하여 이용한 것이 좋은 검색 결과를 보였다. 그 중에서도 색상과 질감의 가중치를 동일하게 0.5로 하는 것이 가장 좋은 결과를 나타냈다.

표 3 색상과 질감 특징의 각 가중치에 대한 검색결과 (테스트 영상: '말' 관련 영상, C:색상, T:질감)

검색 영상수	검색된 관련 영상 수			
	C:1.0 T:0.0	C:0.0 T:1.0	C:0.5 T:0.5	C:0.7 T:0.3
5	4	4	5	4
10	8	4	9	9
15	12	9	14	12
20	12	10	16	14

마지막 실험으로 총 600개의 영상을 DB에 저장해 놓

고 그 중 질의로 사용할 영상(장미, 말, 계곡, 일몰, 꿈, 사람 영상)을 각각 20개씩 포함시킨 후 각 질의 영상에 대해 검색해 보았다. DB에 저장하여 사용하는 데이터에 따라 검색율은 다소 차이가 있겠으나 본 실험에서는 대체로 높은 검색율을 얻을 수 있었다. 600개 중 20개 영상을 검색한 결과, 표 4와 같이 '사람' 영상의 경우 관련 영상이 18개가 검색되어 0.9의 높은 검색율을 보였으며, '일몰' 영상의 경우 0.75로 다소 낮은 검색율을 보였다.

표 4 6종류 영상에 대한 검색결과

질의 영상수	검색된 관련 영상 수						검색율					
	장미	말	계곡	일몰	꿈	사람	장미	말	계곡	일몰	꿈	사람
5	5	5	5	5	5	5	1	1	1	1	1	1
10	9	9	10	8	10	9	0.9	0.9	1	0.8	1	0.9
15	13	11	15	11	13	14	0.87	0.93	1	0.73	0.87	0.93
20	16	16	16	15	17	18	0.8	0.8	0.8	0.75	0.85	0.9

내용 기반 영상 검색에서 통상 precision과 recall을 이용하여 검색성능을 판단하는데 본 연구에서는 DB에 저장된 관련영상과 검색영상 수를 20개씩으로 동일하게 함으로써 precision과 recall에 의한 결과 수치를 동시에 나타내는 검색율을 이용했다. 계산식은 식 (14)와 같다.

$$\text{검색율} = \frac{\text{관련 영상 수}}{\text{총 검색 영상 수}} \quad (14)$$

검색 데모는 그림 10~11과 같으며, 그림 9는 검색시스템의 초기화면을 나타낸다.

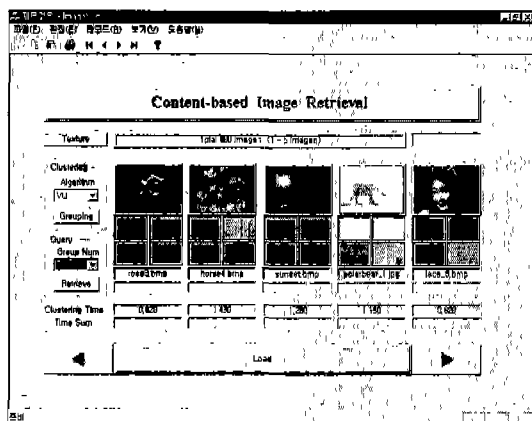


그림 9 검색시스템 초기화면

'꿈' 관련 영상의 검색 화면은 그림 10, '사람' 관련

영상의 검색 모습은 그림 11과 같다.

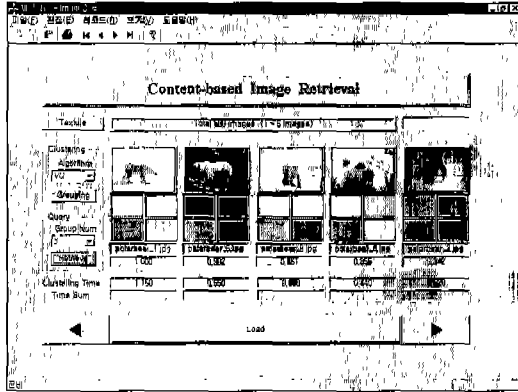


그림 10 '곰' 관련 영상 검색 결과

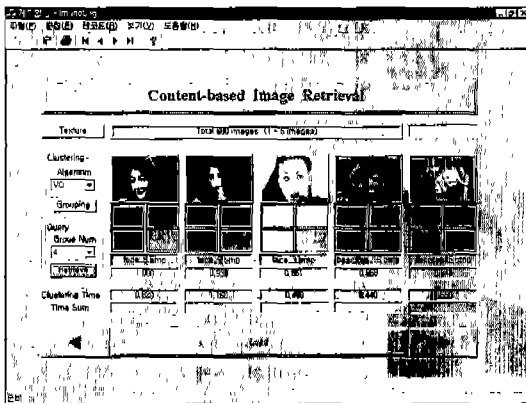


그림 11 '사람' 관련 영상 검색 결과

### 6. 결론 및 향후과제

본 논문에서는 정지 영상을 구성하고 있는 주요 객체들의 대표 색상 및 질감 특징치를 VQ(Vector Quantization) 알고리즘을 이용하여 빠르게 추출하였으며 한 영상의 내용을 적절하게 표현하는 대표 특징치를 내용 기반 검색에 이용함으로써 정지영상의 내용에 근거한 검색을 가능하게 했고 객체 단위 검색을 함으로써 객체의 위치, 회전 및 크기변화에 무관한 검색은 가능하게 했다.

영상의 특징으로 색상과 질감 특징을 이용하였다. 색상 특징으로는 RGB 색상 모델을 HSV 색상 모델로 변환함으로써 얻을 수 있는 색도, 채도, 명도를 이용하였고, 질감 특징으로는 gray-level co-occurrence matrix를 이용하여 얻을 수 있는 ASM, contrast, correlation,

variance, entropy를 이용했다.

추출된 색상과 질감 특징은 화소단위 특징으로 변환되며 각 화소는 8차원(색상특징치 3차원과 질감특징치 5차원)의 벡터 특징치를 갖게 되었다. 각 화소의 8차원 벡터 특징치는 VQ 적용을 위한 입력 데이터로 사용되었으며, VQ를 이용하여 한 영상의 특징치들을 클러스터링 함으로써 각 영상을 구성하는 주요 객체들의 대표 특징치를 얻을 수 있었으며 다른 클러스터링 알고리즘에 비해 적은 시간 안에 클러스터링 결과를 볼 수 있었다.

VQ를 이용하여 얻은 대표 특징치는 영상 검색에 활용되었다. 영상의 대표 특징치 중 색상 관련 특징치를 이용하여 각 영상의 주요 객체들에 대한 대표 색상을 색상 테이블에 표현하였다. 영상 안의 관심 객체를 나타내는 색상을 색상 테이블에서 선택함으로써 DB에 저장되어 있는 모든 영상의 각 객체들과 유사도 계산을 하여 검색 결과를 보여주었다. 유사도 계산에 있어 색상과 질감 특징에 대한 기중치를 0.5로 동일하게 주는 것이 가장 좋은 검색 결과를 보였으며, 6종류의 영상을 검색해 본 결과 최고 0.9의 높은 검출율을 보였다.

본 연구에서는 VQ를 이용한 객체의 대표 특징치 추출을 위해 임계값으로 0.4를 고정하여 사용하였다. 몇 가지 임계값의 변화에 따른 클러스터링 결과 중 본 연구에서 사용한 데이터에는 0.4의 임계값이 일반적으로 가장 좋은 클러스터링 결과를 보였기 때문이다. 그런데 영상에서 추출된 특징치들의 분포의 특성에 따라 임계값을 조정하여 각 영상의 특징치들을 클러스터링 한다면 더욱 정확하게 각 객체들을 대표하는 특징치를 얻을 수 있고, 따라서 더욱 좋은 검색결과를 얻을 수 있을 것이다. 이와 더불어 본 논문에서 제안한 객체 단위 검색 이외에 검색 성능을 높이기 위해 전체적인 영상 정보로서 전역적인 색상 및 질감 정보를 추가하는 방법을 고려할 수 있다. 또한 trademark나 symbol 등과 같이 색상 및 질감 특징보다는 형상 정보가 유용한 영상의 검색을 위해서 Fourier descriptor나 Canny-edge 등과 같은 형태 정보를 추가하여 이용하고 질의자의 주관적 의사를 반영하기 위해서 질의자가 검색 결과의 만족 여부를 시스템에 피드백(feedback)하는 relevance feedback을 이용하면 더 나은 검색 성능을 갖는 검색 시스템을 구현할 수 있을 것이다.

### 참고 문헌

[1] W. Niblack, R. Barber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic, C. Faloutsos, and G. Taubin, "The QBIC project : Querying images by



- content using color, texture, and shape," In Proc. SPIE vol.1908 : Storage and Retrieval for Image and Video Databases, pp.173-181, February 1993.
- [2] A. Hampapur, A. Gupta, B. Horowitz, C. Fuller, J.R. Bach, M. Gorkani, R. C. Jain, Virage Inc., "Virage Video Engine," In Proc. SPIE Vol.3022:Storage and Retrieval for Image and Video Databases, pp.188-198, February 1997.
- [3] J. R. Smith and S. F. Chang, "VisualSEEK : A Fully Automated Content-Based Image Query System," in Proc. ACM Int'l. Conf. Multimedia, pp.87-98, 1996.
- [4] M. J. Swain, C. Frankel, Vassilis Athitsas, "Webseer: An Image Search Engine for the World Wide Web," Chicago Univ. Technical Report 96-114.
- [5] Tom M. Mitchell, Machine Learning, The McGraw-Hill Companies, Inc., 1997.
- [6] Serge Belongie, Chad Carson, Hayit Greenspan, and Jitendra Malik, "Color and Texture-Based Image Segmentation Using EM and Its Application to Content-Based Image Retrieval," Sixth International Conference on Computer Vision, pp.675-682, January. 1998.
- [7] Abhijit. S.Pandy, Pattern Recognition With Neural Networks in C++, IEEE Press, 1995.
- [8] Ioannis Pitas, Digital Image Processing Algorithms, Prentice Hall, England Cliffs, NJ, 1993.
- [9] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, 1993.
- [10] P.Aigrain, H. Zang and D. Petkovic, "Content-Based Representation and Retrieval of Visual Media : A State-of-the-Art Review," Multimedia Tools and Applications, vol.3, pp.179-202, 1996.
- [11] Robert M. Haralick, K. Shanmugam, Its'hak Dinstein, "Textural Features for Image Classification" IEEE Transactions on Systems, Man, and Cybernetics Vol. SMC-3, No.6 November 1973.
- [12] W. Y. Ma and B. S. Manjunath, "A Pattern Thesaurus for Browsing Large Aerial Photographs," Tech. Rep. ECE TR-96-10, June 1996.
- [13] P. P. Ohanian and Richard C. Dubes, "Performance Evaluation For Four Classes of Textural Features," Pattern Recognition, vol.25, no.8, pp. 819-833, 1992.
- [14] Th. Hermes, Ch. Klauck, J. Krey B, and J. Zhang, "Image Retrieval for Informatin Systems," In Proc. SPIE Vol 2420: Storage and Retrieval for Image and Video Databases, pp.394-405, February 1995.

장 동 식

정보과학회논문지 : 컴퓨팅의 실제  
제 7 권 제 5 호 참조



정 세 환

1972년 11월 21일생. 1997년 동국대학교 산업공학과 졸업. 1999년 고려대학교 산업시스템정보공학 석사. 1999년 ~ 현재 LG 기술원 Innovation Center Digital Vision 그룹 연구원. 관심분야는 컴퓨터 비전, 멀티미디어시스템



유 현 우

1966년 12월 24일생. 1992년 인하대학교 전기공학과 졸업. 1994년 동대학 전기공학 석사. 2001년 고려대학교 산업시스템정보공학 박사. 1994년 ~ 1997년 LG전자 생산기술센터기술개발연구소. 관심분야는 컴퓨터비전, 멀티미디어시스템, 제어이론



손 용 준

1957년 7월 7일생. 1980년 서울공대 자원공학과 졸업. 1983년 동대학 산업공학 석사. 1988년 Texas A&M University 박사과정 수료. 2001년 고려대학교 산업시스템정보공학 박사과정 수료. 관심분야는 컴퓨터비전, 스케줄링, 머신비전