

VQ를 이용한 영상의 객체 특징 추출과 이를 이용한 내용기반 영상 검색

정세환, 유현우, 장동식

고려대학교 산업공학과

Representative Feature Extraction Of Objects Using VQ and Its Application To Content-Based Image Retrieval

Seh Hwan Jung, Hun Woo Yu, Dong Sik Jang
Department of Industrial Engineering, Korea University

요 약

내용 기반 영상 검색을 위해 본 연구에서는 Vector Quantization을 이용하여 영상을 구성하는 주요 객체들의 특징 추출 방법을 제안한다. 내용 기반 영상 검색 시스템에서 사용되는 영상의 주요 특징들은 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 사용된다. 이러한 특징들 중에서 어떤 특징들을 사용하고 또 어떤 방식으로 결합하느냐에 따라 혹은 영상의 특성을 잘 나타낼 수 있는 주요 특징을 어떻게 추출, 표현하느냐에 따라 검색 성능에 큰 영향을 미친다. 이 중 본 논문에서는 일반적인 색상, 질감 특징 추출방법과 더불어 Vector Quantization 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 크기변화에 무관한 검색을 가능케했다. 연구의 실험 결과 VQ를 이용함으로써 대표특징치 추출시간을 줄일 수 있었고 검색시 색상과 질감 특징의 가중치를 각각 0.5, 0.5로 주는 것이 가장 높은 검색율을 보였으며 제안된 방식에 의해 '사람' 영상의 경우 0.9의 검색율을 보였다.

1. 서론

최근 컴퓨터와 통신기술의 발달과 더불어 멀티미디어 정보 서비스에 대한 요구가 증가하고 있으며, 인터넷상의 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 텍스트기반 검색이 사용되었는데 이 방법은 색인과정에 있어 시간과 비용이 많이 들며 관점의 불일치로 인한 검색 효율성 저하라는 단점을 보인다. 이를 보완하기 위해 멀티미디어 데이터의 내용을 대표할 수 있는 특징을 추출하여, 이를 기반으로 색인과 검출을 수행하는 내용기반 검색방법이 필요하며, 현재 많은 연구가 이루어지고 있다.

내용기반 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 질의와 더불어 영상내의 각 객체들의 위치, 회전 및 크기변화에 무관한 검색을 가능케 할 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 진행중이다. 그런데 EM 알고리즘은 클러스터링 될 클러스터 수를 미리 정해줘야 하며, 임의의 여러 클러스터 수에 대한 클러스터링 결과 중 영상을 구성하는 주요 객체수만큼 가장 적절하게 분류된 클러스터 수를 선정해야하므로 객체의 대표 특징을 추출하는데 비교적 많은 시간이 소요되고 한 영상의 내용을 적절하게 표현하는 대표 특징치를 추출하기가 쉽지않다. 그래서 본 연구에서는 패턴의 분포에 따라 클러스터수를 자

동적으로 결정하는 VQ(vector quantization)알고리즘[4]을 이용하여 클러스터링함으로써 주요 특징치를 빠르게 추출하고 한 영상의 내용을 적절하게 표현하는 대표 특징치를 추출하고자 한다.

2. 영상 특징 추출

영상의 주요 특징들로는 색상, 질감, 형태 및 영상을 구성하고 있는 객체들의 공간적 위치 등이 사용되어 진다. 이 중 본 연구에서 색상(color)과 질감(texture) 특징을 사용했다. 정지영상의 특징 추출과정은 그림 1과 같다.

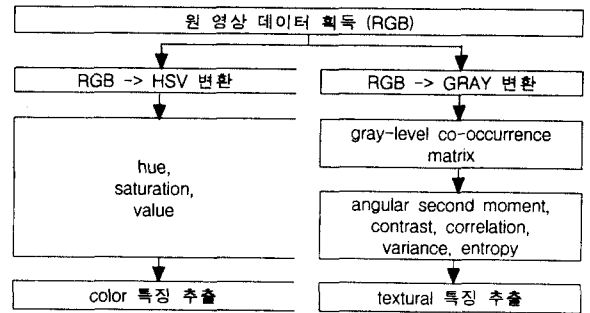


그림1. 색상 및 질감 특징 추출 과정

2.1 색상 특징 추출

색상 특징으로는 RGB모델을 HSV모델로 변환하여 얻은 색도(hue),채도(saturation),명도(value)를 이용했다[1,2]. 색도 18개, 채도,명도는 각각 3개씩으로하며 그레이(gray) 5개를 추가하여 색상 특징치를 167개로 양자화(quantization)하여 사용함으로써 색상특징치를 단순화하였다.

2.1 질감 특징 추출

질감 특징으로는 gray-level co-occurrence matrix를 이용하여 얻을 수 있는 여러 개의 질감 특징중 angular second moment, contrast, correlation, variance, entropy의 5개를 이용했다[3]. 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 질감특징치를 추출하여 평균함으로써 회전엔 무관한 질감 특징치를 얻었다. 각 질감 특징치를 계산하는 식은 식(1)-식(5)과 같으며 각 질감 특징치는 0-1사이 값으로 정규화하여 사용하였다.

$$Angular\ Second\ Moment = \sum_i \sum_j \{p(i, j)\}^2 \quad (1)$$

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

$$Correlation = \frac{\sum_i \sum_j (ij)p(i, j) - \mu_x \mu_y}{\sigma_x \sigma_y} \quad (3)$$

$$Variance = \sum_i \sum_j (i - \mu)^2 p(i, j) \quad (4)$$

$$Entropy = -\sum_i \sum_j p(i, j) \log(p(i, j)) \quad (5)$$

여기서 p(i,j) : gray-level co-occurrence matrix의 (i,j)번째 엔트리
 px(i) : p(i,j)의 행을 합하여 얻은 주변 확률 matrix에서의 i번째 엔트리
 py(j) : p(i,j)의 열을 합하여 얻은 주변 확률 matrix에서의 j번째 엔트리
 Ng : gray level

$\mu_x, \mu_y, \sigma_x, \sigma_y$: px와 py의 평균과 표준편차

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

색상특징은 화소단위 특징인 반면, 질감특징은 지역단위 특징이다. 따라서 VQ를 이용하기 위해서는 특정영역의 질감 특징치를 영역안의 모든 화소에 부여하는 등의 방법을 통하여 질감특징도 화소단위 특징치를 갖도록 변환한다.

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

3. VQ를 이용한 객체의 대표 특징치 추출

한 화소당 3가지 색상 특징치와 5가지 질감 특징치, 즉, 8차원 벡터의 특징치를 추출하였는데 이것은 객체의 대표 특징치 추출을 위한 데이터로 사용된다. 즉, 그림 2와 같이 한 영상의 각 화소가 갖는 특징치 벡터가 입력데이터가 되고 이 데이터는 VQ[4]에 의해서 클러스터링되어 최종적으로 영상을 구성하는 각 객체들의 대표 특징치가 얻어진다. 이는 영상 검색에 활용된다.

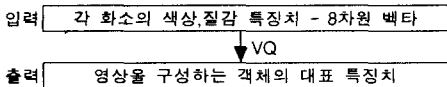


그림2. VQ이용시 입력데이터와 출력값

VQ는 분류될 데이터의 클러스터링 개수를 미리 정하여 주지 않아도 적절한 클러스터링 개수만큼 데이터를 자동적으로 분류한다. 이런 특징을 본 연구에 사용함으로써 EM[5,6]을 비롯한 다른 알고리즘에 비해 영상을 구성하는 주요 객체의 수만큼 비교적 유연하게 대표 특징을 추출할 수 있었고 특징 추출시간을 줄일 수 있었다. VQ에 의한 데이터의 클러스터링 절차는 그림 3과 같으며 클러스터링 절차중 사용할 임계값을 본 연구에서는 0.4로 하였다. 거리계산식은 식(6)과 같으며 입력벡터와 가장 가까운 클러스터를 선정하는 식은 식(7)과 같고 클러스터의 중심값을 계산하는 식은 식(8)과 같다.

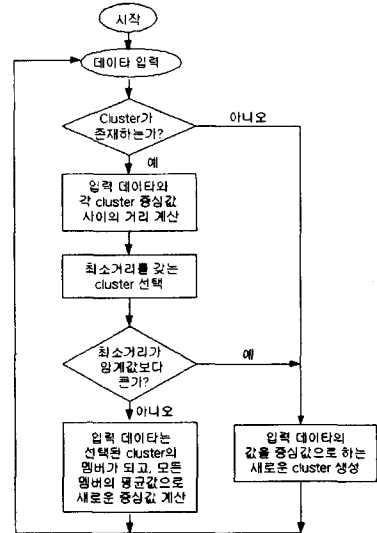


그림3. VQ에 의한 데이터 클러스터링 절차

$$d = \|X^{(b)} - C_k\| = \left[\sum_{i=1}^N (x_i^{(b)} - c_{ki})^2 \right]^{\frac{1}{2}} \quad (6)$$

$$\|X^{(b)} - C_k\| < \|X^{(b)} - C_j\| = \begin{cases} j=1, \dots, M \\ j \neq k \end{cases} \quad (7)$$

$$C_k = \frac{1}{N_k} \sum_{x \in S_k} X \quad (8)$$

여기서 $X^{(b)}$: p번째 입력벡터, N: 벡터 dimension

C_j : j번째 cluster의 중심값

C_k : 입력데이터와 가장 가까운 cluster의 중심값

N_k : 클러스터 포함 멤버(데이터)수, S_k : k번째 클러스터 집합

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 4와 같다.

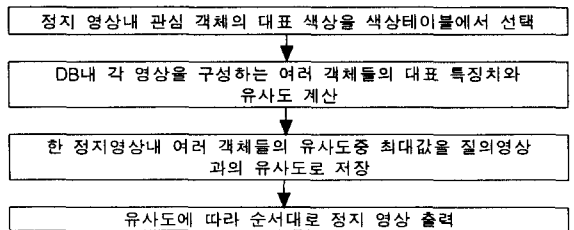


그림4. 영상 검색 과정

본 연구에서는 그림 5와 같이 각 영상마다 정지 영상내 객체들의 대표 특징치중 주요 대표 객체 4개를 선정하여 색 상태이블에 표현하였으며, 각 정지 영상내 대표 색상중 검색을 원하는 객체의 대표 색상을 선택함으로써 검색을 시작한다. 대표 색상이 선정되면 각 정지영상마다 계산해 놓은 객체들의 대표 특징치들과 유사도 계산을 한다. 유사도 계산을 위해서는 각 특징치들간의 거리계산이 필요한데 본 논문에서는 식(9)의 유클리드 거리 계산법이 이용되었으며 유사도는 식(10)에 의해서 두 객체 특징치간의 유사도를 0-1사이의 값으로 표현했다.

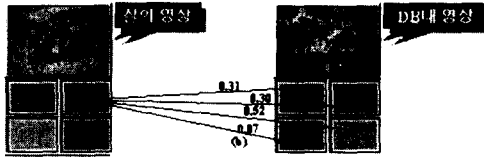


그림5. 영상 검색 과정. (a)질의 영상의 관심 객체 특징치를 색상으로 표현, DB내 저장 특징치는 0.4233:0.2366:0.5722:0.0019:0.0054:0.5864:0.0156:0.5762. (b) DB내 영상의 각 객체들과의 유사도 비교, 3번째 객체가 0.87로 최대값을 나타냄.

$$d_{qp} = w_9 [w_1 (H_Q - H_D)^2 + w_2 (S_Q - S_D)^2 + w_3 (V_Q - V_D)^2]^{1/2} + w_{10} [w_4 (ASM_Q - ASM_D)^2 + w_5 (CONT_Q - CONT_D)^2 + w_6 (COR_Q - COR_D)^2 + w_7 (VAR_Q - VAR_D)^2 + w_8 (ETRP_Q - ETRP_D)^2]^{1/2} \quad (9)$$

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

$$S_{qp} = e^{-\frac{d_{qp}}{2}} \quad (10)$$

5. 실험 및 결과

5.1 실험환경

실험을 위해 펜티엄 II333, RAM 64M 사양의 PC를 이용했으며, 실험 데이터로 192x128 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 맞맞게 cluster의 개수를 자동적으로 결정짓지 못하기 때문에 각 영상마다 2~10개의 cluster로 분류하는데 걸리는 시간을 계산하였다. EM이 VQ에 비해 대략 1.4배정도의 시간이 소요됨을 알 수 있었다. 소요되는 시간은 아래의 그림6 과 표 1과 같다.

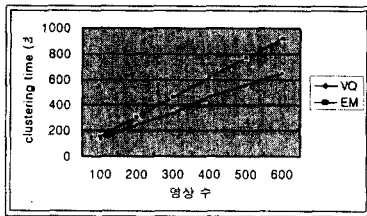


그림 6. clustering 소요시간 비교

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

표 1. clustering 소요시간 비교

두 번째 실험으로 색상과 질감의 가중치를 변화하여 질의를 해보았다. 아래 표 2와 같이 색상과 질감의 가중치를 동일하게 0.5로 주는 것이 가장 좋은 결과를 나타냈다.

(테스트 영상:'말'관련 영상, C:색상, T:질감)

검출 영상수	검출된 관련 영상 수					
	C:1.0	T:0.0	C:0.0	T:1.0	C:0.5	T:0.5
5	4		4		5	4
10		8		4	9	9
15			12	9	14	12
20				10	16	14

표 2. 색상과 질감 특징의 각 가중치에 대한 검출결과

마지막 실험으로 총 600개의 영상을 DB에 저장해 놓고 그 중 질의로 사용할 영상(장미, 말, 계곡, 일몰, 곰, 사람 영상)을 각각 20개씩 포함시킨후 각 질의 영상에 대해 검색해 보았다. DB에 저장하여 사용하는 데이터에 따라 검출율은 다소 차이가 있었으나 본 실험에서는 대체로 높은 검출율을 얻을 수 있었다. 600개중 20개 영상을 검출한 결과, 표 3과같이 '사람' 영상의 경우 관련 영상이 18개가 검출되어 0.9의 높은 검출율을 보였으며, '일몰' 영상의 경우 0.65로 다소 낮은 검출율을 보였다.

검출 영상 수	검출된 관련 영상 수					검출율						
	장미	말	계곡	일몰	곰	사람	장미	말	계곡	일몰	곰	사람
5	5	5	5	5	5	1	1	1	1	1	1	1
10	9	9	10	8	10	0.9	0.9	1	0.8	1	0.9	0.9
15	13	14	15	11	13	0.87	0.93	1	0.73	0.87	0.93	0.93
20	16	16	16	13	17	0.8	0.8	0.8	0.65	0.85	0.9	0.9

표 3. 색상과 질감 특징의 각 가중치에 대한 검출결과

통상 precision, recall를 이용하여 검색성능을 판단하는데 본 연구에서는 DB에 저장된 관련영상과 검출영상 수를 20개씩으로 동일하게 함으로써 precision과 recall에 의한 결과 수치를 동시에 나타내는 검출율을 이용했다. 계산식은 식 (11)과 같다.

$$\text{검출율} = \frac{\text{관련 영상 수}}{\text{총 검출 영상 수}} \quad (11)$$

6. 결론

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

7. 참고 문헌

[1] Ioannis Pitas, Digital, Image Processing Algorithms, Prentice Hall, England Cliffs, NJ, 1993
 [2] Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Addison-Wesley Publishing Company, 1993
 [3] 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
 [4] Abhijit. S.Pandy, Pattern Recognition With Neural Networks in C++, IEEE Press, 1995
 [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