

객체영역의 컬러비와 모멘트를 이용한 내용기반 영상검색

김 은 경[†] · 오 준 택^{††} · 김 욱 현^{†††}

요 약

본 논문에서는 영상 내에 존재하는 객체영역의 컬러비와 모멘트를 이용한 영상검색을 제안한다. 객체는 영상의 중심에 위치한다는 가설 하에 미리 정의한 중심영역의 우세컬러를 기반으로 수평-수직 투영을 이용하여 객체영역과 배경영역을 분할함으로써 최적의 공간정보를 획득한다. 또한 영상 내 객체의 회전 및 크기에 불변한 특성을 가지기 위해 컬러비와 모멘트를 특징정보로 이용하며 유사성 측정은 컬러 히스토그램의 구간별 연관성을 고려하기 위해 변형된 히스토그램 인터섹션을 이용한다. 실험결과 제안한 방법이 기존의 영역분할에 의한 방법보다 효율적인 결과를 보였다.

Content-based Image Retrieval using Color Ratio and Moment of Object Region

EunKyong Kim[†] · Juntaek Oh^{††} · Wookhyun Kim^{†††}

ABSTRACT

In this paper, we propose a content-based image retrieval using the color ratio and moment of object region. We acquire an optimal spatial information by the region splitting that utilizes horizontal-vertical projection and dominant color. It is based on hypothesis that an object locates in the center of image. We use color ratio and moment as feature informations. Those are extracted from the splitted regions and have the invariant property for various transformation, and besides, similarity measure utilizes a modified histogram intersection to acquire correlation information between bins in a color histogram. In experimental results, the proposed method shows more flexible and efficient performance than existing methods based on region splitting.

키워드 : 내용기반 영상검색(content-based image retrieval), 객체영역(object region), 수평-수직 투영(horizontal-vertical projection), 우세컬러(dominant color), 컬러비(color ratio), 모멘트(moment)

1. 서 론

컴퓨터 기술의 발달로 디지털 방송 관리, 주문형 비디오, 디지털 라이브러리, 의료 진단 시스템, 홈쇼핑 등의 분야에서 다양한 형태의 멀티미디어 데이터가 활용되고 있으나 정보검색 분야에서 기존의 텍스트 기반 서비스는 멀티미디어 데이터로의 적용에 한계점을 지닌다. 가장 많이 이용하는 멀티미디어 데이터 중 하나인 영상은 컴퓨터 및 네트워크 기술의 개발에 더불어 대량으로 증가하고 있으며 더불어 사용빈도도 점점 증가하고 있는 실정이다. 현재 이를 효율적으로 관리하고 다루기 위해 다양한 목적의 영상 데이터베이스들이 생성되고 있다. 그러나 이러한 영상 데이터베이스에서 사용자가 원하는 영상을 찾기에는 상당한 어려움이 따른다. 따라서 효율적으로 원하는 영상을 검색할 수 있는 방법이 요구되어진다[5, 6, 10].

영상검색은 크게 텍스트기반 영상검색(text-based image retrieval)과 내용기반 영상검색(content-based image retrieval)으로 분류된다. 텍스트기반 영상검색은 주석이나 키워드에 의존하는 검색방법으로 주석 자체의 데이터 양이 작아 저장 및 관리가 용이하고 고속 검색이 가능한 장점이 있다. 그러나 수작업에 의해 영상의 내용을 주석으로 표현해야하는 번거로움이 있고 표현하는 사람의 관점에 따라 영상의 내용이 다를 수 있어 정보의 일관성과 객관성이 보장되지 않는다. 이에 반해 내용기반 영상검색은 영상의 색(color), 형태(shape), 질감(texture) 등의 내용적인 특징이나 속성을 이용하여 영상을 표현하고 이것을 기반으로 검색을 수행한다. 그러므로 영상의 시각적인 특성을 기반으로 데이터베이스를 손쉽게 구축할 수 있고 효율적인 관리 및 검색이 가능하므로 현재 많은 시스템이 내용기반 영상검색을 사용한다[5, 6].

본 논문에서는 영상 내에 존재하는 객체의 효율적인 영역 분할에 의한 영상검색을 제안한다. 먼저, 최적의 공간정보를 가지기 위해 영상 내의 객체는 중심에 위치한다는 가설[4]에

† 종신회원 : 한국인식기술 연구원
 †† 준 회원 : 영남대학교 대학원 컴퓨터공학과
 ††† 정 회원 : 현재 영남대학교 전자정보공학부 교수
 논문접수 : 2002년 2월 14일, 심사완료 : 2002년 7월 24일

근거하여 미리 정의된 중심영역에서 가장 빈번히 나타나는 우세컬러를 기반으로 수평-수직 투영을 이용하여 적응적으로 영상을 객체영역과 배경영역으로 분할한다. 또한 분할된 각 영역에 대해 추출한 컬러비와 모멘트를 특징정보로 이용한다. 컬러비는 어느 정도의 밝기 변화, 잡음의 첨가, 크기 변화 등에 불변하며 모멘트는 주요 객체의 크기 및 회전에 불변하다는 특성을 지닌다. 질의 영상과 영상 데이터베이스 내 영상간의 유사도 측정은 히스토그램의 구간별 연관성을 고려하기 위해 변형된 히스토그램 인터섹션을 이용한다.

2. 관련 연구

내용기반 검색방법에서 가장 중요한 것은 각 영상이나 객체를 기술하는데 있어 사용하는 특징 집합을 어떻게 선택하느냐가 중요한 관건이라 할 수 있다. 또한 물체의 회전, 이동, 크기 변화에 강건해야 하며 기억공간의 절약과 빠른 처리를 위해 영상의 파라미터의 개수가 적어야 한다. 현재 이러한 조건들을 만족시킬 수 있는 다양한 내용기반 검색에 대한 연구가 활발히 진행 중에 있다.

현재 가장 많이 이용하는 영상의 특징정보는 컬러정보로써 히스토그램을 이용하여 추출한다. 이것은 유사한 영상들은 유사한 컬러 분포를 갖는다는 가설 하에 Swain에 의해 제안되었다[1]. 이 방법은 계산이 쉽고 효율적이며 영상의 회전과 작은 이동에도 민감하지 않다는 장점을 가진다. 그러나 전혀 다른 영상이라도 유사한 히스토그램 분포를 가지는 경우 유사한 영상으로 판단하는 단점을 가진다. 현재 히스토그램 방법의 단점을 보완하기 위해 공간정보를 이용하는 많은 연구들이 제안되었다[2, 3, 10, 13].

컬러정보와 함께 공간정보를 이용하는 방법으로, Stricker [2, 3]는 영상 내의 물체는 주로 중앙에 있다는 가설 하에 영상을 부분적으로 겹쳐지는 5개의 퍼지 영역으로 분할함으로써 공간정보를 획득하며 분할된 영역으로부터 추출한 모멘트 정보를 기반으로 L1-norm을 이용하여 유사도를 측정하였다. Striker의 방법은 퍼지 영역으로부터 특징값들을 추출하므로 이동과 회전에 민감하지 않은 특성을 가지나 분할하는 경계 영역에 객체가 존재할 경우 객체의 특징을 제대로 고려하지 못한다는 단점을 가진다.

Dawn[7]에 의해 제안된 CBA(a new content based comparison algorithm) 방법은 히스토그램 방법과 CCV(color coherence vector) 방법에 공간정보를 위한 영역분할을 추가한 방법이다. CCV는 각 화소의 컬러값에 대한 연속성의 여부를 기반으로 한 색상정보의 표현방법으로 색상정보의 지역적 특성을 고려하여 같은 색상 분포를 가지는 다른 영상들을 구별할 수 있는 장점을 가지나 많은 처리시간을 요구한다. CBA 방법은 화소의 컬러값에 대해 평균값 처리를 적용한 후, 영상의 컬러공간을 양자화하였다. 그리고 전체 화소의 수를 기준으로 11개의 면적(area) 구간으로 분할한 후, 영상간에 같은

색상정보를 가지는 영역들의 면적이 동일한 경우 같은 영역으로 분류한다. 분류된 영역의 개수와 각 영역의 색상정보에 대해 빈도수의 거리 측정값을 이용하여 유사도를 측정하였다.

Hsu[4]에 의해 제안된 방법은 영상의 전반적인 느낌을 표현하기 위해 공간지식과 함께 통합된 컬러 정보 기술을 제안하였다. 영상을 구분하는 주요한 객체가 중심에 위치한다는 가설에 따라 영상을 일정한 크기의 중심영역과 바깥영역으로 나누어서 각 영역에 대해 컬러집합을 생성한 후, 히스토그램을 이용하여 대표 컬러를 선택한다. 또한 대표컬러에 따른 공간정보를 얻기 위해 이벤트 차폐 방법과 최대 엔트로피 분리 방법을 사용하였다.

이러한 방법들에 사용된 컬러정보는 영상 내에서의 이동, 회전, 크기 변화에 불변하다는 장점을 가지나 잡음의 첨가 등과 같이 색상 히스토그램을 변화시키는 요인에 대해서는 민감하며 영상 내의 객체에 대해서 정확한 특징을 추출하기 어렵다는 단점을 지닌다.

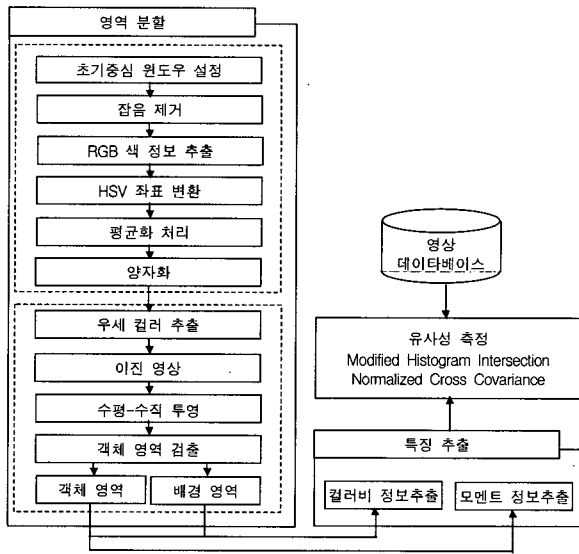
Jain[5, 6]에 의해 제안된 방법은 컬러정보와 함께 모양정보를 결합한 방법으로 Swain이 제안한 방법을 약간 변형시킨 색상 히스토그램 인터섹션 방법과 방향 히스토그램 인터섹션 방법을 이용하였다. 또한 검색 시간을 단축시키기 위해 클러스터링 방법과 branch and bound-based matching 방법을 이용하였다. 모양정보를 결합함으로써 검색의 효율은 높일 수 있었지만 컬러 히스토그램 자체를 변화시키는 요인에 대해서는 민감한 반응을 보였다.

김봉기[13]는 고정된 5개의 분할영역에 대해 색상의 평균, 표준편차, 왜도 정보와 향상된 불변 모멘트를 이용한 2단계 영상검색기법을 제안하였다. 향상된 불변 모멘트를 이용하여 효율성 및 검색 속도를 향상시킬 수 있었으나 고정된 영역 분할에 의해 객체에 대한 정확한 특징은 추출할 수 없었다.

본 논문에서는 영상 내에 존재하는 객체의 크기에 따라 적응적으로 객체영역과 배경영역을 분할함으로써 최적의 공간정보를 얻을 수 있다. 또한 잡음 제거와 함께 컬러비와 모멘트정보를 특징정보로 이용함으로써 히스토그램을 변화시키는 잡음의 첨가나 빛의 변화에 대해 둔감하며 객체의 회전, 크기 변화 등에 불변하는 특징을 지닌다.

3. 영역분할에 따른 내용기반 영상검색

본 논문에서 제안하는 영상검색시스템의 전체 구성은 (그림 1)과 같이 영역분할, 특징추출, 유사성 측정으로 구성된다. 영역분할 부분은 미리 정의한 중심영역에서 추출한 우세컬러를 기반으로 수평-수직 투영(horizontal-vertical projection)을 이용하여 영상 내의 객체영역과 배경영역을 분할한다. 특징추출 부분은 분할된 객체영역과 배경영역에 대해 컬러비와 모멘트를 추출하며 유사도 측정 부분은 히스토그램 인터섹션(histogram intersection)[1]과 normalized cross-covariance [14] 등을 이용하여 질의영상과 미리 구축한 영상 데이터베이스에 저장된 영상들간의 유사도를 계산한다.



(그림 1) 제안한 영상검색시스템

3.1 영역분할

인간이 영상을 볼 때는 배경영역보다 객체영역의 변화에 더 민감하며 객체영역은 일반적으로 영상의 중심에 위치한다는 가설에 근거하여 영상의 중심으로부터 일정 범위를 가지는 초기 영역을 설정한 후, 초기 영역 내에서 가장 많은 동일한 컬러 값을 가지는 화소들이 이루는 영역을 객체영역으로 결정한다. 본 논문에서는 영상 내에 존재하는 잡음을 제거하기 위해 median filter를 적용한 후, 인간의 시각 능력과 유사한 특성을 갖는 HSI(hue, saturation, intensity) 컬러공간의 hue 정보만을 이용한다. 식 (1)은 RGB(red, green, blue) 컬러공간으로부터 HSI 컬러공간으로의 변환식을 나타낸다. H, S, I는 화소의 hue, saturation, intensity에 대한 값을 나타내며 R, G, B는 red, green, blue에 대한 값을 나타낸다.

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

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

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

자연계에 존재하는 컬러들은 무수히 많지만 미세한 변화의 경우 사람의 눈이 인식하지 못할 뿐 아니라 처리과정도 복잡하기 때문에 컬러공간을 양자화할 필요가 있다. 따라서 HSI 컬러공간의 hue 정보에 대해 유사한 색상끼리 일정한 크기의 집합으로 나누어 히스토그램 구간을 결정하는 균등양자화 방법을 이용하여 18개의 구간으로 균집화한다.

일반적으로 영상을 식별하는 데 있어 가장 중요한 객체는 영상의 중심에 위치한다는 가설을 근거로 초기에 설정한 중심영역에서 양자화된 hue 정보의 히스토그램($H_{cw}[i]$)을 추출한 후, 가장 큰 값을 갖는 구간을 주도적인 객체영역의 컬러, 즉 우세컬러(D_c)[9]로 결정한다. 이러한 중심영역에서의

우세컬러 결정은 식 (2)과 같다.

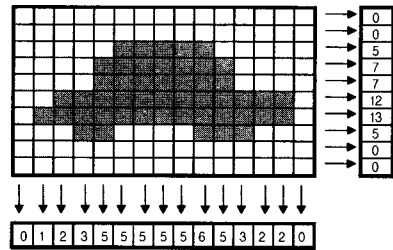
$$D_c = \max(H_{cw}[i]) \quad 0 \leq i \leq 17 \quad (2)$$

영상 내에 뚜렷한 객체가 존재하지 않더라도 우세컬러에 의해 검출된 영역을 객체영역으로 설정한다. 영상의 전체영역에 대해 우세컬러에 따른 이진영상을 생성하면 영상 내의 객체뿐만 아니라 유사한 값을 가지는 다른 영역들도 검출될 수 있으므로 객체의 크기에 따른 정확한 영역의 검출을 위해 수평-수직 투영 방법을 이용한다. (그림 2)는 이진영상에 대해 수평-수직 투영을 적용한 결과를 보여주며 수평축과 수직축에 대한 투영은 식 (3)과 식 (4)와 같다. $P_h(i)$ 와 $P_v(j)$ 는 수평축과 수직축에 대한 화소값 $f(i, j)$ 의 누적 화소 수이며 i, j 는 x 축과 y 축에 대한 좌표를 나타낸다.

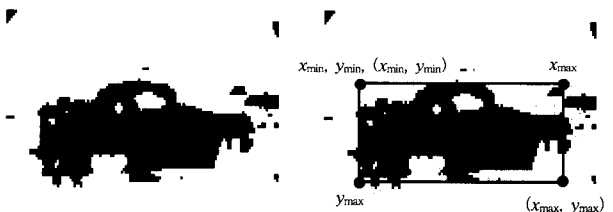
$$P_h(i) = \sum_{j=0}^n f(i, j) \quad (3)$$

$$P_v(j) = \sum_{i=0}^n f(i, j) \quad (4)$$

수평-수직 투영을 적용한 후, 영역의 크기가 미리 정의한 임계치보다 작은 경우 잡음으로 인식하여 제거하고 임계값 이상의 결과값을 가지는 영역을 객체영역으로 검출한다. 객체영역에 대해 수평 투영에 의한 시작점 및 끝점을 x_{min}, x_{max} 라고 하고 수직 투영에 의한 시작점 및 끝점을 y_{min}, y_{max} 라고 할 때 객체영역의 시작점 및 끝점은 $(x_{min}, y_{min}), (x_{max}, y_{max})$ 와 같이 표현할 수 있다. (그림 3)은 자동차에 관련된 이진영상에 대해 이러한 좌표점을 기준으로 분할한 영상을 보여준다.



(그림 2) 수평-수직 투영



(그림 3) 수평-수직 투영을 이용한 영역분할

3.2 특징 추출

컬러정보를 추출하기 위해 분할한 객체영역과 배경영역에서의 컬러비를 이용한다. 컬러비는 영상 내에 존재하는 객체에 대해 크기, 회전의 변화가 가미되어도 결정된 객체영역의 크기에 따른 컬러들의 비율을 구함으로써 일정한 값을 가지

는 특징을 지닌다.

객체영역과 배경영역에 대한 컬러비는 식 (5)와 식 (6)과 같다.

$$O_{ratio}[i] = \frac{H_o[i]}{ObjectSize} \quad \sum_{i=0}^{17} O_{ratio}[i] = 1.0 \quad (5)$$

$$B_{ratio}[i] = \frac{H_b[i]}{BackgroundSize} \quad \sum_{i=0}^{17} B_{ratio}[i] = 1.0 \quad (6)$$

식 (5)에서 $O_{ratio}[i]$ 와 $B_{ratio}[i]$ 는 객체영역과 배경영역에 대한 히스토그램 구간별 컬러비를 나타내며 $H_o[i]$ 과 $H_b[i]$ 는 객체영역과 배경영역의 hue 정보를 기반으로 추출한 히스토그램의 i 번째 컬러구간의 값을 나타낸다. $ObjectSize$ 와 $BackgroundSize$ 는 객체영역과 배경영역에 대한 크기를 나타내며 배경영역은 영상의 전체크기에서 객체영역을 제외한 영역이다.

컬러정보와 함께 불변한 객체영역에 대한 모양정보를 추출하기 위해 불변 모멘트(invariant moment)를 이용한다. 이러한 모멘트는 패턴인식분야에서 물체를 기술하는 데 있어 우수한 특성을 가진다[2, 6, 13].

2차원 연속함수 $f(x, y)$ 에 대해 차수 $(p+q)$ 를 가지는 모멘트와 디지털 영상에서 중앙모멘트는 식 (7)과 식 (8)과 같이 정의된다.

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (7)$$

$$\mu_{pq} = \sum_x \sum_y (x-x')^p (y-y')^q f(x, y) \quad x' = \frac{m_{10}}{m_{00}}, y' = \frac{m_{01}}{m_{00}} \quad (8)$$

정규화된 중앙모멘트 η_{pq} 는 식 (9)과 같이 정의된다.

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^\gamma} \quad \gamma = \frac{p+q}{2} + 1, (p+q) = 2, 3, 4, \dots \quad (9)$$

불변 모멘트 집합 ϕ_i 는 정규화된 중앙 모멘트로부터 식 (10)을 이용하여 쉽게 유도할 수 있으며 이동, 회전, 크기 변화 및 거울 영상에 대해서 불변하다는 특징을 가진다.

$$\begin{aligned} \phi_1 &= \eta_{20} + \eta_{02} \\ \phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2 \quad \phi_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2 \\ \phi_4 &= (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2 \\ \phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12}) [(3\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ \phi_6 &= (\eta_{20} - \eta_{02}) [(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\ &\quad + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\ \phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12}) [(\eta_{30} + \eta_{12})^2 - 3(\eta_{21} + \eta_{03})^2] \\ &\quad + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03}) [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \end{aligned} \quad (10)$$

3.3 유사성 측정

질의영상에서 추출한 각 영역별 특징들과 영상 데이터베이스 내 영상들에 대한 특징간에 유사성을 비교하여 질의영상에 가장 유사한 영상을 검색한다. 본 논문에서는 히스토그램 인터섹션[1]과 normalized cross-covariance[14]를 이용하여 유사성을 측정한다.

질의영상과 영상 데이터베이스에서 추출한 특징들의 히스토그램 H, D 가 N 차원의 벡터로 구성될 때 각각 $(h_0, h_1, h_2, \dots, h_{N-1}), (d_0, d_1, d_2, \dots, d_{N-1})$ 로 표현할 수 있으며 h_i, d_i 는 히스토그램 H, D 의 i 번째 구간에서의 값을 의미한다. 히스토그램 인터섹션에 의한 유사도 HI_S 는 식 (11)과 같이 정의된다.

$$HI_S = \frac{\sum_{i=0}^{N-1} \min(h_i, d_i)}{\sum_{i=0}^{N-1} (h_i)} \quad (11)$$

HI_S 는 유사성을 나타내고 $0 < HI_S \leq 1$ 의 범위를 가지며 유사한 영상일수록 HI_S 는 1에 가까운 값이 되고 유사성이 떨어지는 영상일수록 HI_S 는 0에 가까운 값을 가진다. 그러나 HI_S 는 히스토그램 구간간의 연관성을 고려하지 않는다. 예를 들어, 컬러정보를 기반으로 추출한 히스토그램 H 와 D 가 각각 (1,1,1,1,5)와 (5,1,1,1,1)로 구성된 경우와 (1,1,1,1,5)와 (1,1,1,5,1)로 구성된 경우 HI_S 는 모두 5/9로 동일한 유사도를 가지지만 시각적으로 느끼기에는 (1,1,1,1,5)와 (1,1,1,5,1)로 구성된 것이 더 유사하다. 그러므로 컬러정보에 의한 유사도 측정은 식 (12)와 같이 변형된 히스토그램 인터섹션을 이용한다.

$$HI_S = \frac{\sum_{i=0}^{N-1} \min(h_i, d_i)}{\sum_{i=0}^{N-1} (h_i)} \times H_C \quad (12)$$

$$H_C = \frac{(N - M_a \times |h_{\max} - d_{\max}|)(N - M_b \times |h_{\min} - d_{\min}|)}{N^2}$$

H_C 는 히스토그램 H 와 D 에서 최대값과 최소값을 가지는 구간간의 연관성을 고려한 값으로 $h_{\max}, h_{\min}, d_{\max}, d_{\min}$ 은 각 히스토그램에서 최대값, 최소값을 가지는 구간의 인덱스를 나타내며 M_a 와 M_b 는 연관성에 대한 가중치로 상수이다. H_C 는 비교하는 히스토그램 H 와 D 에서 최대값과 최소값을 가지는 구간간의 인덱스가 근접할수록 1에 가까운 값을 가진다.

Normalized cross-covariance[14]는 벡터들간의 연관성에 따른 유사도 측정방법으로 식 (13)과 같이 정의된다.

$$NCC_S = \frac{H^m \cdot D^m}{\|H^m\| \|D^m\|} \quad (13)$$

$$H^m = H - \text{mean}(H), D^m = D - \text{mean}(D)$$

$\text{mean}(H)$ 와 $\text{mean}(D)$ 는 히스토그램 H 와 D 의 평균값을 나타내며 \cdot 와 $\| \quad \|$ 는 내적과 크기값을 의미한다. 유사도 NCC_S 는 H^m 과 D^m 이 유사할수록 1에 가까운 값을 가진다.

컬러정보를 기반으로 질의영상의 객체영역과 배경영역에서 측정된 유사도를 O_S, B_S 이라 하고 모양정보를 기반으로 객체영역에서 측정된 유사도를 M_S 라고 할 때 전체 유사도 G_S 는 식 (14)과 같이 정의한다.

$$G_S = \varphi_c(\omega_o O_S + \omega_b B_S) + \varphi_m M_S \tag{14}$$

$$\omega_o + \omega_b = 1, \varphi_c + \varphi_m = 1$$

O_S 와 B_S 는 객체영역과 배경영역에 대해 컬러정보를 기반으로 측정된 유사도이며 M_S 는 객체영역에 대해 모양정보를 기반으로 측정된 유사도를 나타낸다. ω_o 와 ω_b 는 O_S 와 B_S 에 대한 가중치를 나타내며 φ_c 와 φ_m 는 컬러정보와 모양정보를 기반으로 측정된 유사도의 가중치를 나타낸다. ω_o 가 ω_b 보다 클 경우, 영상 내의 객체영역을 중심으로 더 유사한 영상을 검색하며 반대의 경우, 배경영역을 중심으로 더 유사한 영상을 검색한다.

4. 실험결과 및 분석

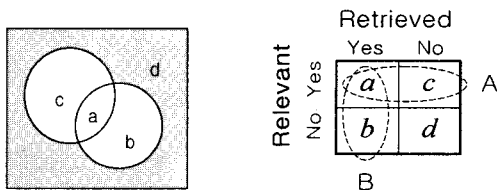
본 논문은 구축한 영상 데이터베이스에서 질의영상을 뽑아 유사한 영상들을 검색하는 query by example 방식을 이용한다. 실험환경은 Pentium 450Mhz의 Window 2000 환경에서 자바 언어로 구현하였으며 실험영상은 영상의 중심에 객체가 존재하는 1750개의 다양한 컬러영상들로, 256×256 크기로 정규화한 후 이용하였다. 유사도 측정 함수는 기본적으로 히스토그램 인터섹션 방법을 이용한다.

초기의 중심영역은 영상의 중심으로부터 영상의 전체 크기의 1/4만큼 떨어진 영역으로 결정하였으며 수평-수직 투영에서 수평, 수직축에 대한 누적 화소의 수가 20이하인 영역은 잡음으로 처리하였다. 컬러정보를 기반으로 측정된 객체영역과 배경영역의 유사도에 대한 가중치는 0.6, 0.4로 설정하였으며 컬러정보와 모양정보를 기반으로 측정된 유사도에 대한 가중치는 0.5, 0.5로 설정하였다. 또한 검출한 객체영역이 영상 크기의 1/10보다 작을 경우 초기에 설정한 영역을 객체영역으로 하였다.

본 논문에서는 유사성에 대한 성능평가 척도로서 recall과 precision을 이용한다[8, 11, 13]. Recall은 영상 데이터베이스 내에서 질의 영상과 관련된 영상 중에서 검색된 영상의 비율을 의미하고 precision은 검색된 영상 중에서 질의 영상과 관련된 영상의 비율을 의미한다. A를 질의영상과 관련된 영상들의 집합, B를 검색된 영상들의 집합이라 하고 a, b, c, d는 (그림 4)와 같이 정의할 때 recall과 precision은 식 (15)와 같이 정의할 수 있다.

$$Recall = P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{a}{a+c} \tag{15}$$

$$Precision = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{a}{a+b}$$



(그림 4) precision과 recall을 이용한 성능평가
(a + b + c + d = 영상 데이터베이스, a + c = 질의영상과 관련된 영상들, a + b = 질의영상에 의해 검색된 영상들)

<표 1>은 기존에 제안된 방법과 본 논문이 제안하는 방법에 대한 효율성을 보여준다. <표 1>의 결과를 종합하면 본 논문이 제안한 방법이 (a) 질의영상의 전체크기에 대해 히스토그램을 이용한 검색방법이나 (b) Striker의 방법, (c) Jain의 방법보다 우수하다는 것을 알 수 있다. (그림 5)는 영역분할에 따른 기존의 방법과 제안한 방법에 의해 검색된 영상들을 보여준다. (그림 5) (a)는 질의영상을 나타내며 (그림 5) (b)는 질의영상의 전체크기에 대해 히스토그램을 이용하여 검색한 영상들을 보여준다. (그림 5) (c)는 영상의 객체가 중심에 위치한다는 가정 하에 영상을 일정한 크기의 고정된 중심영역과 바깥영역으로 분할한 후, 각 영역에 대해 히스토그램을 이용하여 검색한 영상들을 보여준다. (그림 5) (d)는 본 논문이 제안하는 방법에 의해 검색한 영상들을 보여준다. (그림 5) (b)는 공간정보를 고려하지 않아 유사하지 않은 영상도 검색하였으며 (그림 5) (c)는 (그림 5) (b)보다 효율적인 결과를 보이지만 중심영역을 고정함으로써 객체의 위치나 크기가 변화할 경우에는 유사성이 현저히 떨어지는 결과를 보인다. 그러나 본 논문이 제안한 방법은 영상 내의 존재하는 객체의 크기에 따라 적응적으로 객체영역과 배경영역을 분할함으로써 객체영역에 대해 정확한 컬러비와 모멘트를 추출할 수 있어 객체에 변화가 가미된 유사한 영상들에 대해 강건한 검색이 가능하다. (그림 6)은 다양한 질의영상들에 대해 검색한 영상들로 (a), (c), (e), (g), (i), (k), (m), (o), (q), (s)는 질의영상을 나타내며 (b), (d), (f), (h), (j), (l), (n), (p), (r), (t)는 검색한 영상들을 보여준다. 영상들은 유사도 순서에 따라 좌측에서 우측으로 6위까지 나열하였다. (그림 6)에서는 질의영상에 대해 잡음을 추가하거나 빛의 밝기 변화, 객체의 크기 및 회전 등 다양한 변형이 가미된 유사한 영상들도 강건하게 검색하였음을 알 수 있다. 또한 변형된 히스토그램 인터섹션에 의해 질의영상과 유사한 컬러정보를 가지는 영상들을 많이 검색하였으며 배경영역보다 객체영역에서 측정된 유사도에 가중치를 더 할당함으로써 객체의 특징정보에 민감함을 알 수 있다. 그리고 hue 정보만으로는 RGB 컬러공간의 영상 내에 존재하는 객체들에 대한 세밀한 정보는 표현할 수 없기 때문에 전체적으로 모양정보보다 컬러정보에 민감하였다. 특히, (그림 6) (i), (s)와 같이 객체의 크기가 상대적으로 작은 경우 컬러정보에 많이 민감함을 알 수 있다. <표 2>는 유사도 측정 함수로 이용한 히스토그램 인터섹션과 normalized cross-covariance간의 영상검색 성능 비교를 보여주며 히스토그램 인터섹션이 normalized cross-covariance 방법보다 좀 더 우수하였다. (그림 7)과 (그림 8)은 같은 질의영상에 대해 히스토그램 인터섹션과 normalized cross-covariance 방법을 이용한 영상검색을 보여준다. (그림 7)의 (a), (c), (e)과 (그림 8)의 (a), (c), (e)는 질의영상을 나타내고 (그림 7)의 (b), (d), (f)와 (그림 8)의 (b), (d), (f)는 각 질의영상에 따른 검색된 영상들을 보여준다. 각 질의영상에서 검출된 객체영역은 파란색의 하늘, 빨간색의 꽃, 짙은 노란색의 사막이고 객체영역을 제외한 다른 부분은 배경영역이다. (그림 9)는 제안한 방법에서 낮은 효율성을 갖는 질의영상들로 (a)는 검출한 객체영역이 너무 작아 초기에 설정한 중심영역을 객체영역

으로 설정된 경우이며 (b)와 (c)는 뚜렷한 객체가 존재하지 않거나 배경과의 컬러정보가 너무 유사하여 분할을 수행하지 못하고 질의영상 자체를 하나의 객체영역으로 설정된 경우로 객체에 대한 정확한 특징을 추출할 수 없었다.

<표 1> 성능 평가

실험방법	성능평가	Recall	Precision
(a) 전체영역		0.52	0.41
(b) Striker의 방법		0.76	0.69
(c) Jain의 방법		0.78	0.73
(d) 제안한 방법		0.91	0.87

<표 2> 유사도 측정 함수간의 성능 비교

실험방법	성능평가	Recall	Precision
(a) 변경된 히스토그램 인터섹션을 이용한 방법		0.91	0.87
(b) Normalized cross covariance를 이용한 방법		0.87	0.84



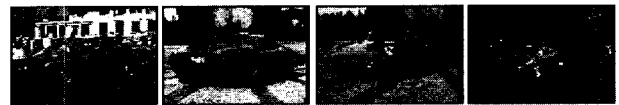
(a) 질의영상



(b) 전체영역을 기반으로 한 검색영상

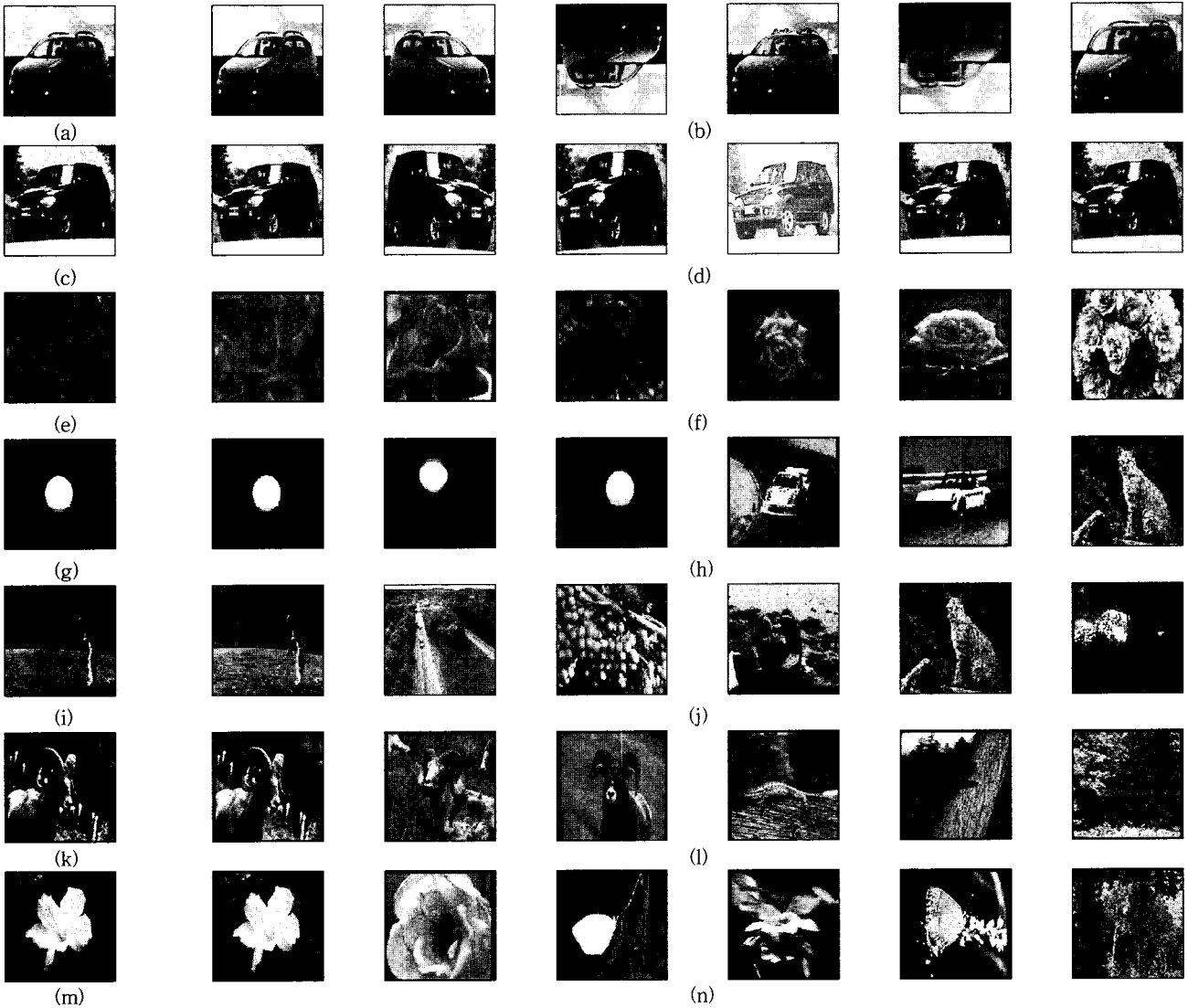


(c) 고정된 영역분할을 기반으로 한 검색영상



(d) 제안한 방법에 따른 검색영상

(그림 5) 영역분할에 따른 기존의 방법과 제안한 방법간의 비교



5. 결 론

본 논문은 우세컬러에 따른 영역분할에 의한 새로운 내용기반 영상검색방법을 제안한다. 기존의 영역분할에 따른 내용기반 영상검색은 공간정보를 추출하기 위해 영상 내의 객체를 고려하지 않고 영상을 일정한 영역으로 분할함으로써 객체에 대한 정확한 특징을 추출할 수 없다. 그러나 제안한 방법은 주요 객체가 영상의 중심에 위치한다는 일반적인 가설에 근거하여 초기에 설정한 중심영역으로부터 추출한 우세컬러를 기반으로 수평-수직 투영을 이용하여 영상 내에 존재하는 객체영역을 적응적으로 검출함으로써 분할한 객체영역에 대해 정확한 특징을 추출할 수 있다. 그리고 영상검색을 위한 특징정보로 분할한 각 영역에 대해 컬러정보인 컬러비와 모양정보인 모멘트를 이용함으로써 히스토그램의 분포를 변화시키는 요인이나 영상 내에 존재하는 객체에 대해 회전, 크기 등의 변화가 가미된 유사한 영상이라도 검색이 가능하다. 또한 영상의 객체영역과 배경영역에서 측정된 유사도에 가중치를 할당함으로써 영역별 중요도를 결정할 수 있다.

향후과제로는 영상 내의 객체를 뚜렷이 구분할 수 없는 영상이나 객체와 배경간의 컬러정보가 유사한 경우 잘못된 객체영역을 검출하거나 질의영상 자체를 객체영역으로 인식하므로 효율이 낮은 영상검색을 유발하였다. 그러므로 시각적으로 중요한 요소인 영상 내의 객체영역을 정확히 검출할 수 있는 영역분할에 관한 연구가 기대된다.

참 고 문 헌

[1] M. J. Swain and D. H. Ballard, "Color Indexing," International Journal of Computer Vision, Vol.7, No.1, pp.11-31, 1991.
 [2] M. Stricker and M. Orengo, "Similarity of Color Image," Proc. of SPIE, Vol.2420, pp.381-392, 1995.
 [3] M. Striker and A. Dimai, "Color Indexing with Weak Spatial Constraints," Proc. of SPIE, Vol.2670, pp.29-40, 1996.
 [4] W. Hsu, T. S. Chua and H. K. Pung, "An Integrated Color Spatial Approach to Content Based Image Retrieval," Multimedia 95, pp.305-313, 1995.
 [5] A. K. Jain and A. Vailaya, "Image Retrieval using Color and Shape," Pattern Recognition, Vol.29, No.8, pp.1233-1244, 1996.
 [6] A. K. Jain and A. Vailaya, "Shape-Based Retrieval : A Case Study with Trademark Image Databases," Pattern Recognition, Vol.31, No.9, pp.1369-1390, 1998.
 [7] Dawn, Xiaodong and Song, "CBA-A New Content Based Comparison Algorithm," <http://www.cs.cornell.edu/info/people/skyxd/paper1.html>.
 [8] John R. Smith and Shih-Fu Chang, "Tools and Techniques for Color Image Retrieval," Proc. of SPIE, Vol.2670, pp.426-

437, 1996.
 [9] K. C. Ravishankar, B. G. Prasad, S. k. Gupta and K. K. Biswas, "Dominant Color Region Based Indexing for CBIR," Proc. of ICIAP, pp.887-892, 1999.
 [10] B. Atalay, C. Celik, I. Haritaoglu and K. Kilic, "Integrating Color and Spatial Information for the Image Indexing," <http://citeseer.nj.nec.com/atalay97integrating.html>.
 [11] H. Yamamoto, H. Iwasa, N. Yokoya and H. Takemura, "Content-Based Similarity Retrieval of Images Based on Spatial Color Distributions," Proc. of ICIAP, pp.951-956, 1999.
 [12] T. Gevers and V. K. Kojcovski, "Image Segmentation by Directed Region Subdivision," Proc. of IAPR, Vol.1, pp.342-346, 1994.
 [13] 김봉기, 오해석, "색상과 모양정보를 이용한 2단계 영상검색 기법", 한국멀티미디어학회논문지, 제1권 제1호, pp.173-182, 1998.
 [14] S. Ravela and R. Manmatha, "On computing global similarity in images," WACV98, pp.82-87, 1998.



김 은 경

e-mail : ekkim@cse.yu.ac.kr
 2000년 영남대학교 컴퓨터공학과 졸업 (학사)
 2002년 영남대학교 대학원 컴퓨터공학과 졸업(공학석사)
 2002년~현재 한국인식기술 연구원
 관심분야 : 영상처리, 시각정보처리, 패턴 인식



오 준 택

e-mail : ojt@cse.yu.ac.kr
 1999년 영남대학교 컴퓨터공학과 졸업 (학사)
 2001년 영남대학교 대학원 컴퓨터공학과 졸업(공학석사)
 2001년~현재 영남대학교 대학원 컴퓨터공학과 박사과정
 관심분야 : 문자인식, 영상처리, 시각정보처리



김 욱 현

e-mail : whkim@yu.ac.kr
 1981년 경북대학교 전자공학과 졸업(학사)
 1983년 경북대학교 대학원 컴퓨터공학과 (공학석사)
 1993년 일본 쓰쿠바대학 공학연구과 (공학박사)

1983년~1993년 한국전자통신연구원 선임연구원
 1994년~현재 영남대학교 전자정보공학부 부교수
 관심분야 : 시각정보처리, 패턴인식, 영상처리