영상 검색에서 의미 기반 영상 분류 기술 동향 [1)]

전재현* 민현석* 김세민* 노용만* 한승완**

최근 인터넷과 같은 정보 공간의 확산으로 영상의 유통 및 소비 시스템이 급격히 변화하면서 영상의 양적 성장이 가속화되고 접근 또한 용이해졌다. 이에 따라 영상 검색에서 영상을 의미적으로 분류하는 것이 도전적이고 중요한 문제가 되었다. 이러한 문제를 해결하기 위한 연구의 한 방향으로 내용기반 영상 분류 기술이 발전되어 왔다. 기존의 내용기반 영상 분류 기술은 영상에 포함된 물체나 배경의 색(color), 모양(shape), 질감(texture)과 같은 영상의 내용으로부터 추출된 특징을 기반으로 한다. 그러나, 영상과 내용기반 특징 사이에 의미 격차(semantic gap)가 발생하여 영상의 고수준 의미를 신뢰성 있게 표현하는데 한계가 있다. 이러한 의미 격차 문제를 해결하기 위한 새로운 연구로서, 내용기반 특징과 고수준 의미 사이의 연결 고리 역할을 수행하는 의미 특징을 이용한 영상 분류 방법이 대두되었다. 본 고에서는 영상 검색에서의 영상 분류 기법을 사용되는 특징에 따라 분류하고, 각 기법에 대한 사례를 알아보고자 한다.

목 차

* KAIST 전지전자, 영상비디오시스템연구실
** ETRI 지식정보보호연구팀/선임연구원

Ⅰ. 서 론

최근 인터넷, IPTV, 온라인 사회적 네트워크(online social network)와 같은 정보 공간의 확산으로 멀티미디어 영상의 유통 및 소비가 급격히 증가하고 있다. 인터넷과 영상 제작 기술의 발달로 영상의 양과 접근성은 높아짐에 따라 영상 검색에서 영상을 분류하는 것이 중요한 문제가 되었다[1],[4].

1990 년대에는 텍스트에 기반한 영상 분류 기술 연구가 주를 이루었다. 그러나 텍스트에 기반한 영상 분류 방법은 영상에 적합한 키워드를 사람이 직접 달아주어야 하기 때문에 영상에 대한 키워드 선정이 주관적이고 키워드를 모든 영상에 연결시키는 고비용이 발생하는 문제점이 있었다[7]. 따라서 새

1) 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT R&D 사업의 일환으로 수행하였음[2009-F-054-01, 유해 멀티미디어 콘텐츠 분석/차단 기술 개발].

로운 연구 방향으로 제시된 것이 영상의 색(color), 모양(shape), 질감(texture) 등의 내용기반 특징을 이용한 영상 분류 방법이다.

내용기반 영상 분류 기술은 영상에 포함된 물체나 배경의 색, 모양, 질감과 같은 내용기반 특징을 분석하여 영상의 의미를 판단하고 영상을 자동 분류하는 것이다. 내용기반 특징을 이용함으로써 사람의 개입 없이 자동으로 영상을 분류하게 되어, 객관적이고 저비용인 영상 분류 및 검색이 가능하게 되었다. 그러나 내용기반 특징은 영상의 모호하고 복잡한 고수준 의미 표현(High-level semantic description)들을 신뢰성 있게 표현하는데 한계가 있다[1]. 이런 문제점은 일반적으로 의미 격차(semantic gap)[1]라고 알려져 있다.

내용기반 연구에 대한 문제점인 의미 격차를 해결하기 위해 내용기반 특징과 고수준 의미 표현 사이의 다리 역할을 하는 의미 특징(semantic feature)을 이용하는 방법이 제안되었다. 의미 특징들을 이용하여 고수준 의미 기반으로 영상을 분류하는 것은 사람이 영상을 인식하여 분류하는 과정과 매우 유사하다. 사람은 영상을 인식할 때 단계적으로 인식한다. 우선 대략적으로 간단한 의미 객체들(semantic objects)을 인식하고, 그 다음에 의미 객체들을 종합해서 영상의 자세한 의미를 파악하게 된다[1]. 이와 같이 고수준 의미 기반으로 영상을 잘 분류하기 위해서는 의미 특징들을 정의하여 사용할 수 있다. 영상 분류에서 주로 사용되는 의미 특징들은 LSCOM (Large-Scale Concept Ontology for Multimedia)에서 제안된 출현빈도(observability), 가능성(feasibility) 등을 사용하여 정의할 수 있다[2],[9].

본 고에서는 영상 검색에서의 영상 분류 기법과 사례들을 기술한다. 본 고의 구성은 다음과 같다. Ⅱ장에서는 영상 분류 기술을 기반 특징에 따라 세분하고 각각에 대해 상세히 서술하며, Ⅲ장에서는 내용기반 특징과 의미 특징을 이용한 영상 분류 사례를 살펴본다. 마지막으로 Ⅳ장에서는 영상 분류 기술의 향후 연구 방향을 기술한다.

Ⅱ. 영상 분류 기술

영상 검색에서 영상을 분류하는 방법은 사용되는 특징에 따라 크게 내용기반 특징을 이용한 연구와 의미 특징을 이용한 연구로 나눌 수 있다.

1. 내용기반 특징을 이용한 영상 분류 기법

영상 분류 기법에 주로 사용되는 내용기반 특징(content-based feature)으로는 크게 영상의 색, 모양, 질감 등이 있다. 각각의 특징이 영상 분류에 어떻게 사용되는지 살펴보면 다음과 같다.

가. 색 특징

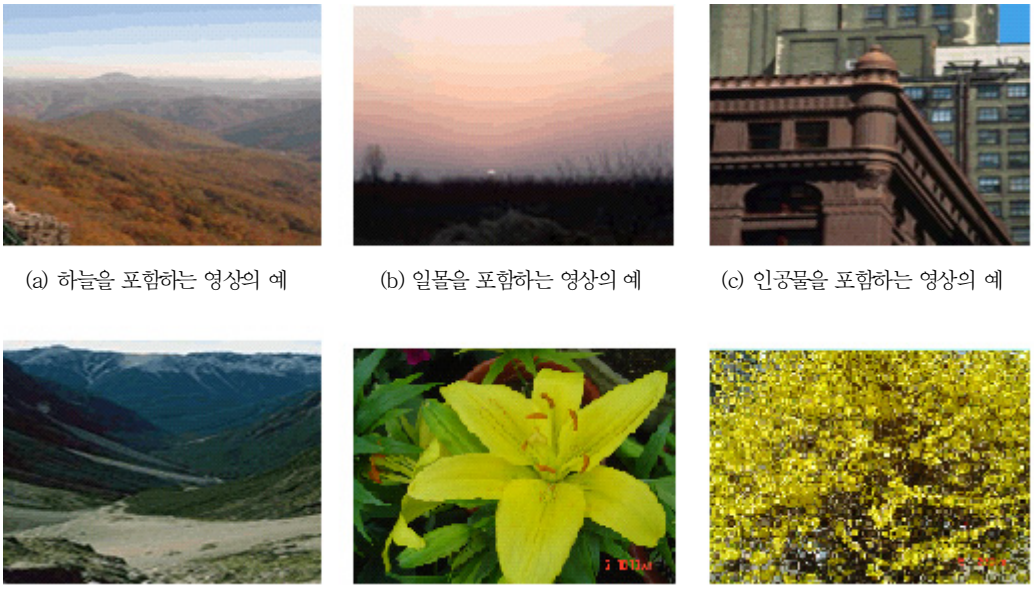
영상을 표현하기 위해 RGB, HSV, LUV, YCbCr 등과 같은 다양한 색 공간이 존재하고, 다양한 색 공간에서 분류 목적에 따라 색 히스토그램(color histogram)[3]-[5], 지배적인 색(dominant color)[4],[8], 평균 색(average colors)[6], 색 분산(color variance)[6],[8] 등의 색 특징(Color feature)들을 추출하여 영상 분류에 이용할 수 있다.

영상이 포함하는 고수준 의미를 근사적으로 표현하기 위해 색 특징을 다양하게 이용할 수 있다. 예를 들어, (그림 1) (a)의 하늘(Sky)이나 (b)의 일몰(sunset)을 포함하는 영상들을 분류하기 위해서 색 히스토그램, 지배적인 색, 평균 색 등을 이용할 수 있다. 하늘은 주로 푸른색을 띄기 때문에 푸른색이 영상에서 많은 비중을 차지하고, 일몰의 경우는 영상에서 노을이 많은 부분을 차지함으로 붉은색이나 노랑색이 많은 비중을 차지한다.

나. 모양 특징

영상에서 모양을 추출하는 방법들로는 경계 방향 분포(edge direction distribution)[3],[4], 고차 자기 상관 경계(High-Order Autocorrelation edge)[3] 등이 있다.

모양 특징(Shape feature)들은 영상이 포함하는 고수준 의미가 사람이 만든 인공물이냐, 아



(a) 하늘을 포함하는 영상의 예 (b) 일몰을 포함하는 영상의 예 (c) 인공물을 포함하는 영상의 예

(d) 자연풍경을 포함하는 영상의 예 (e) 꽃을 포함하는 영상의 예 (f) 꽃나무를 포함하는 영상의 예

(그림 1) 고수준 의미들을 포함하고 있는 영상들의 예[MPEG-7 VCE2 dataset]

니면 자연풍경이냐를 구분하기 위해 이용될 수 있다. (그림 1) (c)의 사람이 만든 인공물 같은 경우는 일반적으로 수직적이거나 수평적인 경계(edge)를 많이 가진다. 그러나 (그림 1) (d)의 자연풍경의 경우는 무작위적인 경계 방향성을 가진다.

다. 질감 특징

영상으로부터 질감은 다양한 방법으로 얻어질 수 있다. 질감 특징(Texture feature) 추출의 대표적인 방법을 살펴보면, 가보 필터(Gabor filter)를 이용한 질감 특징[6], 타무라 질감 특징(Tamura texture feature)[6],[8], 웨이블릿 질감 특징(wavelet texture feature)[7],[8], MSAR 질감 특징[4] 등이 있다.

영상으로부터 추출된 질감 특징은 색이나 모양 특징으로 결정될 수 없는 영상 내에 존재하는 반복적인 패턴을 발견하는데 이용된다. (그림 1) (e)의 꽃과 (f)의 나무는 유사한 색상을 나타내고 지배적인 색 분포 또한 비슷하여, 색상 정보로는 분류하기가 어렵다. 반면에, 이미지 상에서 질감이 부드러운 꽃과 빈 공간 때문에 질감이 거친 꽃나무의 특징을 이용하면 쉽게 분류할 수 있다.

2. 의미 특징을 이용한 영상 분류 기법

영상에 내포되어 있는 모호하고 복잡한 특성으로 인해 내용기반 특징만을 활용하여 영상의 고수준 의미를 표현하는데 한계가 존재한다. 이와 같은 표현의 한계를 일반적으로 의미 격차(semantic gap)[10]라고 한다. 영상과 특징 사이에 나타나는 의미 격차를 줄이고 영상의 분류 성능을 높이기 위한 방법으로 제안된 것이 의미 특징을 이용한 영상의 의미 분류 방법이다.

영상의 고수준 의미를 잘 표현하기 위해 간단하고 기본적인 의미 특징(semantic feature)들을 정의하여 사용할 수 있다. 영상 분류에서 주로 사용되는 의미 특징들은 LSCOM 에서 제안된 출현빈도(observability), 가능성(feasibility) 등을 사용하여 정의할 수 있다[9].

예를 들어, LSCOM 에서 제안된 출현빈도와 가능성의 특성들에 기반하여 도시(city)의 한 장면을 나타내는 영상들의 의미 특징들을 정의할 수 있다. 도시를 나타내는 영상들을 관찰해 보면, (그림 2)와 같이 자동차, 포장도로, 보도, 건축물, 창문 등과 같은 의미 특징들이 높은 출현빈도를 보이므로 의미 특징의 후보들로 선택할 수 있다. 그리고 후보로 선택된 의미 특징들을 가능성이라는 특성을 기반으로 분석해 보면, 색, 모양, 질감 등의 내용기반 특징을 활용하여 자동추출이 가능하므로 가능성 특성을 만족한다. 의미 특징 하나가 정의되는 예를 들어 보면, 포장도로의 도시 영상들에서 많은 출현 빈도를 보이므로 의미 특징의 후보로 선택될 수 있다. 후보로 선

(그림 2) 도시를 표현하는 영상들의 예[MPEG-7 VCE2 dataset]

택된 포장도로는 회색의 아스팔트에 하얀색의 경계선이 그어져 있으며, 아스팔트의 질감이 매끄럽지 않으므로 색 히스토그램과 질감 특징을 추출하여 포장도로인지 판별이 가능하므로 가능성 특성 또한 만족하게 된다. 두 가지 모든 특성을 만족하므로 포장도로는 의미 특징으로 정의된다. 포장도로 이외에도 자동차, 보도, 건축물, 빌딩, 창문 등의 의미 특징들 색, 모양, 질감 특징들을 추출하여 의미 특징 여부를 판별할 수 있고 높은 출현빈도를 보이므로 도시라는 고수준 의미를 지원하는 의미 특징들로 정의된다.

높은 출현 빈도를 보이며 내용기반 특징을 활용하여 자동추출이 가능한 의미 특징들은 영상이 포함하는 고수준 의미를 지원하게 된다. 예를 들면, (그림 2)와 같이 도시라는 고수준 의미를 포함하는 영상들은 자동차, 포장도로, 건축물, 창문, 보도 등이 나타나게 된다. 그러나 (그림 1)의 (a), (b)와 같이 자연풍경을 포함하는 영상들은 도시라는 고수준 의미를 지원하는 자동차, 포장도로, 건축물, 창문, 보도 등이 나타나지 않게 된다. 이와 같이 영상에서 도시를 지원하는 의미 특징들이 나타나는지 여부를 통해 영상이 도시를 표현하는 영상인지 아닌지를 판별할 수 있다.

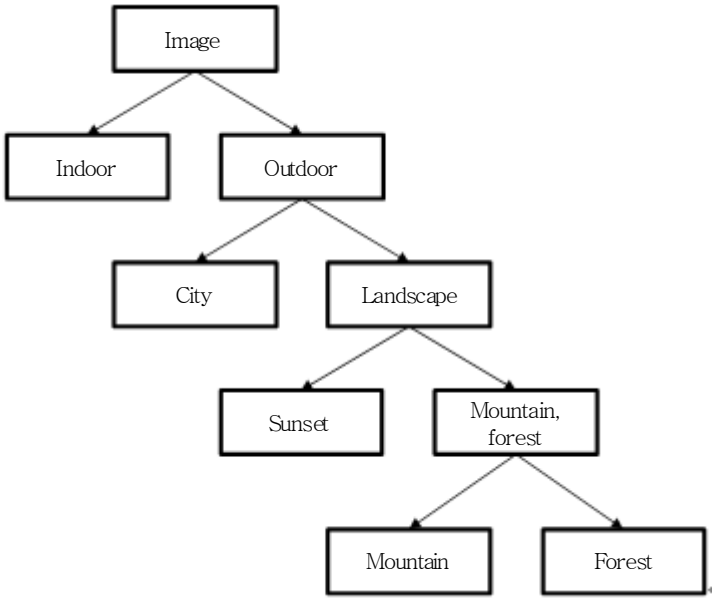
III. 영상 분류 사례

1. 내용기반 특징을 이용한 영상 분류 사례

[4]에서는 영상들을 (그림 3)과 같이 8 가지의 고수준 의미들로 분류하고, 내용기반 특징들을 이용하여 영상을 자동 분류하였다.

(그림 3)을 보면, 영상들은 크게 실내(indoor) 영상과 실외(outdoor) 영상으로 나뉘어진다. 그리고 실외 영상은 도시 영상과 풍경(landscape) 영상으로 나누어지며, 풍경 영상은 다시 일몰 영상과 산(mountain)/숲(forest) 영상으로 나누어진다. 그리고 마지막으로, 산/숲 영상들은 산 영상과 숲 영상으로 나누어진다.

일반적으로 실외 영상은 공간적으로 균일한 색 분포를 가진다. 예를 들어, 하늘은 위에 존재

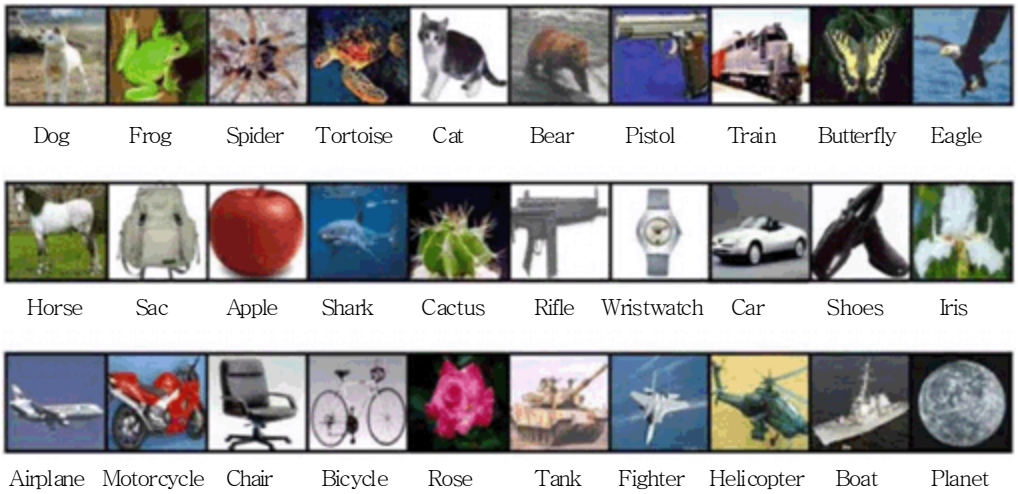


(그림 3) 영상의 의미 분류 예[4]

하며 전형적으로 푸른색이다. 반면에, 실내 영상은 실외 영상보다 다양한 색 분포와 균일한 조명 분포를 특징으로 갖는다. 그래서 공간적 색과 명암도(intensity)의 분포에 따라 실내와 실외 영상을 구분할 수 있다. 반면에, 모양 특징은 실내와 실외 영상을 구분하는데 유용하지 않다. 왜냐하면 실내와 실외에 유사한 모양을 가지는 물체들이 존재할 수 있기 때문이다. 따라서 실내와 실외 영상을 구분하기 위한 내용기반 특징으로써 공간적 색 분포와 명암도 분포를 이용하는 것이 좋다.

도시 영상들은 사람들이 만든 물체들이 존재하기 때문에 주로 강한 수직 및 수평적 경계를 갖는다. 반면에 풍경 영상들은 경계의 방향성이 무작위적으로 분포한다. 이러한 특징으로부터 도시 영상과 풍경 영상은 경계의 분포를 이용해서 분류할 수 있다. 그러나 색 특징의 경우, 사람이 만든 물체들도 역시 임의의 색을 가질 수 있기 때문에 도시 영상과 풍경 영상을 구분하는데 적당하지 않다. 그러므로 도시 영상과 풍경 영상을 구분하기 위한 일반적인 특징으로써 경계의 분포를 자주 이용한다.

풍경 영상을 일몰 영상과 산/숲 영상 그룹으로 나누는데 있어서, 다음과 같은 특징을 활용할 수 있다. 일몰 영상들은 대개 노란색이나 붉은색과 같은 채도가 높은 색을 가진다. 반면에 산 영상들은 푸른색의 하늘을 배경으로 가지는 경향이 있다. 그리고 숲 영상들은 녹색이 많이 분포한다. 그러므로 영상에서의 전체적인 색 분포와 색의 포화도를 이용하여 일몰, 산, 그리고 숲 영상

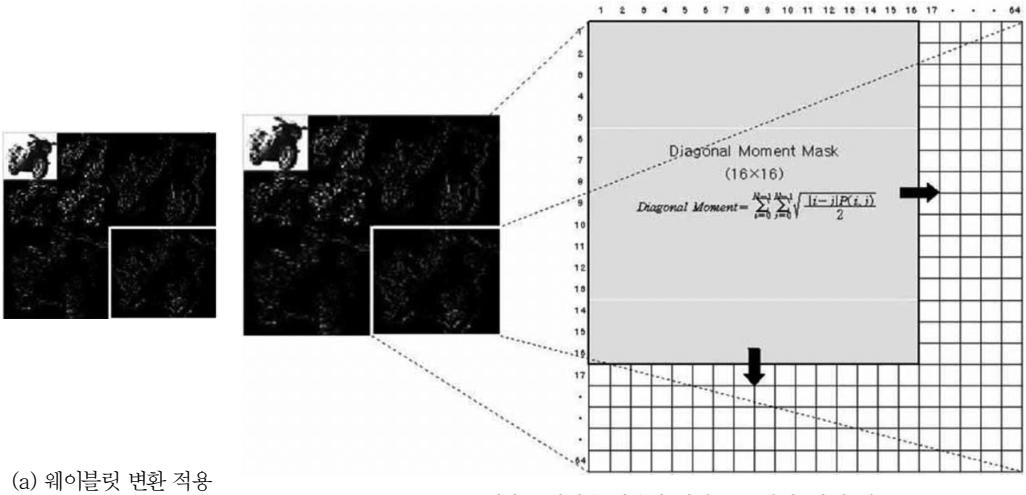


(그림 4) 영상의 의미 분류 예[7]

들을 구분할 수 있다.

[7]은 내용기반 특징을 사용하여 영상을 (그림 4)와 같은 30 가지 고수준 의미로 분류하는 문제를 다루고 있다. (그림 4)와 같은 다양한 영상의 경우, 단순히 색과 질감 특징만을 사용하여 높은 분류 성능을 기대하기는 어렵다.

[7]에서는 이러한 환경에서 분류의 성능을 높이기 위해 웨이블릿 변환(wavelet transform) 이미지로부터 추출한 모양 기반 질감 특징을 이용한다. 예를 들어, 오토바이 영상으로부터 질감



(a) 웨이블릿 변환 적용 후의 결과 영상 예

(b) 웨이블릿 변환을 이용한 질감 특징 추출 과정 예

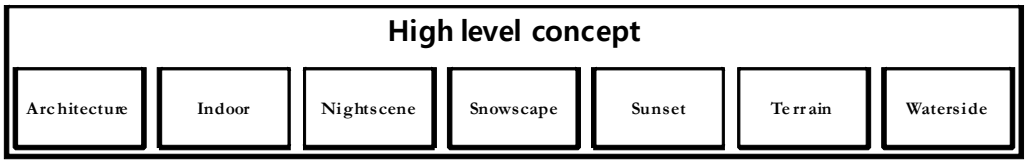
(그림 5) 웨이블릿 변환 기반 질감 특징 사용 예[7]

특징을 구하는 과정은 다음과 같다. 먼저, 오토바이 영상에 웨이블릿 변환을 적용하면 (그림 5) (a)와 같은 결과를 얻게 되고, (그림 5) (a)의 하얀색 테두리로 표시된 대각 고주파 대역(diagonal high-frequency band) 부분으로부터 질감 특징이 추출된다. 대각 고주파 대역의 크기가 64×64 일 때, 64×64 크기의 대역(band)에서 (그림 5) (b)와 같이 16×16 크기의 윈도를 오른쪽과 아래쪽으로 8 화소(pixel)씩 옮겨 가면서 총 49 번의 대각 모멘트(diagonal moment)를 구할 수 있는데, 위와 같은 과정으로부터 얻어진 49 개의 대각 모멘트 값들이 최종적인 질감 특징 값들로 이용되어 영상들을 분류한다.

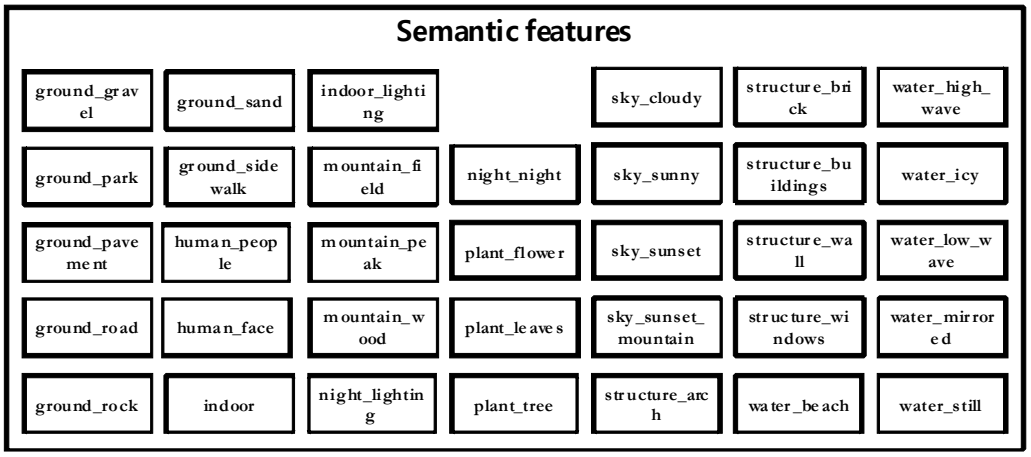
2. 의미 특징을 이용한 영상 분류 사례

영상들을 (그림 6) (a)와 같이 7 가지의 고수준 의미들로 분류하기 위해 [1]은 의미 특징들을 이용하여 영상을 분류하였다.

[1]에서는 입력된 영상들로부터 색, 모양, 질감 등의 내용기반 특징들을 추출하여, 추출된 내용기반 특징들로부터 (그림 6) (b)와 같이 정의된 의미 특징들을 자동 추출하게 된다. 이렇게 추



(a) 영상의 고수준 의미 분류 예[1]



(b) 영상의 고수준 의미를 지원하는 의미 특징들의 예[1]

(그림 6) 의미 특징을 이용한 영상 분류

출된 의미 특징들은 영상의 7 가지 고수준 의미들을 표현하기 위해 사용된다. 예를 들면, 건축물(architecture)의 고수준 의미를 표현하는 영상들을 관찰해 보면, 지면_자갈(ground_gravel), 지면_포장도로(ground_pavement), 지면_도로(ground_road), 지면_보도(ground_sidewalk), 구조물_아치(structure_arch), 구조물_벽돌(structure_brick), 구조물_빌딩(structure_buildings), 구조물_벽(structure_wall), 구조물_창문(structure_windows) 등이 높은 출현빈도를 보이므로 건축물을 지원하는 의미 특징들 후보로 선택할 수 있다. 이렇게 후보로 선택된 의미 특징들은 각각 색, 모양, 질감 등의 특징이 다르므로 구분이 가능하고 자동추출이 가능하다. 지면_자갈과 지면_포장도로의 경우 둘 다 회색이 많이 분포하여 색으로는 구분이 불가능하다. 그러나 지면_자갈이 둥근 모양을 가지고, 지면_포장도로는 수직적인 모양을 가지므로 모양 특징으로 구분이 가능하다. 이와 같이 색, 모양, 질감 등의 특징으로 구분이 가능하고 자동추출이 가능하여 후보로 선택된 의미 특징들은 의미 특징들로 정의된다.

영상들을 출현빈도와 가능성의 특성들을 기반으로 관찰하여, 미리 정의된 의미 특징들을 추출한다. 내용기반 특징 대신 추출된 의미 특징들을 사용하여 영상이 어떠한 고수준 의미를 포함하는지 분류할 수 있다. 이와 같이 의미 특징을 사용하여 영상을 분류함으로써, 내용기반 특징을 사용하였을 경우 발생하는 의미 격차를 줄이게 되고, 영상의 모호하고 복잡한 내용을 보다 더 정확하고 자세히 표현 할 수 있어 영상 분류 성능을 높일 수 있다.

IV. 향후 연구 방향

본 고에서는 영상 검색에서의 영상 분류 방법을 기반이 되는 특징에 따라 내용기반 특징을 이용하는 방법과 의미 특징을 이용하는 방법으로 나누어 살펴보았다. 내용기반 특징을 이용하는 방법은 텍스트 기반의 영상 분류 방법에서 발생하는 주관적이고 고비용의 단점을 해결한다. 그러나 여전히 영상과 내용기반 특징 사이에 의미 격차가 존재하는 문제점을 가지고 있다. 이러한 의미 격차라는 문제점을 해결하기 위해 여러 논문에서 영상들을 관찰하여 영상이 포함하는 고수준 의미를 지원하는 의미 특징들을 정의하고, 정의된 의미 특징들을 이용하여 영상의 분류 성능을 향상시키는 방법들이 연구되고 있다. 향후 영상 검색에서의 영상 분류 성능을 더욱 높이기 위해서는, 지금까지 제안된 방법들에서 사용된 특징을 최적화하기 위한 특징 선택(feature selection) 기법에 대한 연구와 새로운 특징을 개발하는 연구 등이 필요하다.

<참 고 문 헌>

[1] S. G. Yang, S. K. Kim, and Y. M. Ro, "Semantic home photo categorization," IEEE Tran. On Circuits and System for Video Technology, vol.17, No.3, Mar. 2007, pp.324-335.

[2] M. Boutell, A. Choudhury, J. Luo, and C. M. Brown, "Using semantic features for scene classification: how good do they need to be?," ICME 2006, pp.785-788.

[3] Z. Aghbari, and A. Makinouchi, "Semantic approach to image database classification and retrieval," NII journal, No.7, 2003, pp.1-8.

[4] A. Vailaya, M. A. T. Figueiredo, A. K. Jain, and H. J. Zhang, "Image classification for content-based indexing," IEEE Tran. On Image processing, Vol.10, No.1, Jan. 2001, pp.117-130.

[5] D. Han, W. Li, and Z. Li, "Semantic image classification using statistical local spatial relations model," Multimedia Tools and Applications, Vol.39, No.2, Sep. 2008, pp.169-188.

[6] Y. Gao, and J. Fan, "Semantic image classification with hierarchical feature subset selection," MIR'05, Nov. 2005, pp.135-142.

[7] S. B. Park, J. W. Lee, and S. K. Kim, "Content-based image classification using a neural network," Pattern Recognition Letters, Vol.25, 2004, pp.287-300.

[8] J. Fan, Y. Gao, and H. Luo, "Multi-level annotation of natural scenes using dominant image components and semantic concepts," MM'04, Oct. 2004, pp.540-547.

[9] M. Naphade, J. R. Smith, J.Tesic, S. F. Chang, W. Hsu, L. Kennedy, A. Hauptmann, and J. Curtis, "Large-scale concept ontology for multimedia," IEEE Multimedia, Vol.13, No.3, Jul.-Sep.2006, pp. 86-91.

[10] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," IEEE Tran. On Pattern Analysis and Machine Intelligence, Vol.22, No.12, Dec. 2000, pp.1349-1380.

* 본 내용은 필자의 주관적인 의견이며 NIPA의 공식적인 입장이 아님을 밝힙니다.