

다중 특징을 이용한 인공/자연객체 영상의 자동 분류 방법

구경모*, 박창민**, 김민환*

* 부산대학교 컴퓨터공학과

** 영산대학교 멀티미디어 공학부

Automatic classification of man-made/ natural object image using multiple features

Kyung-Mo Koo*, Chang-Min Park**, Min-Hwan Kim*

* Dept. of Computer Engineering Pusan National University

** School of Multimedia Engineering Youngsan University

Email : kmkool@pusan.ac.kr, cmpark@ysu.ac.kr, mhkim@pusan.ac.kr

요 약

최근 많은 연구에서, 동일한 영상그룹들로부터 추출된 저수준의 특징들을 이용해서 고수준의 정보를 분석한 뒤, 이를 이용해서 영상을 분류하는 방법들을 소개하고 있다. 이러한 연구는 CBIR의 인덱싱에서 저수준의 특징만을 사용할 때 발생하는 의미적인 차이(semantic gap)문제를 해결하여, 검색의 효율을 높일 수 있게 한다. 하지만 이들 연구는 대부분 전경(scenery)영상만을 대상으로 하고 있다. 한편 영상을 객체 단위로 다루는 것은 CBIR의 성능을 크게 향상시킬 수 있는 요인이 된다. 왜냐하면 대부분의 사용자는 관심있는 객체가 포함된 영상을 검색하기 원하기 때문이다. 본 논문에서는 영상의 객체를 인공객체와 자연객체로 분류하는 방법을 제안한다. 인공객체의 경우 자연객체에 비해 상대적으로 직선형태의 에지가 많이 발견되며 객체를 구성하는 패턴이 규칙적이고 방향성을 가진다. 또한 인공객체는 자연객체에 비해 객체영역의 경계가 직선에 의한 단순한 형태로 나타난다. 이러한 특징들을 EDH(edge Direction Histogram)의 에너지, EDAS(Energy Difference of Adjacent Sector)와 가버 필터를 통해 추출하여 분류에 이용한다. 실험을 통하여 각 특징들을 개별적으로 사용해서 76%에서 84% 사이의 분류 정확성을 얻었으며, 제안한 머징 방법을 이용하여 최종적으로 약 90%의 정확성으로 분류하였다.

1. 서론

CBIR(내용기반 검색)은 주어진 영상 데이터베이스에서 사용자가 묘사하는 장면이나 특정 객체가 포함된 모든 영상을 찾는 시스템이다. 이러한 CBIR에서는 영상에서 자동으로 추출된 저수준의 특징(칼라, 질감, 형태 등)들을 통해서 시각적인 정보를 추출하며, 이를 기초로 데이터베이스

에 있는 영상들을 자동으로 인덱싱하여 검색의 효율을 높이고 있다. 하지만 이러한 방법에서는, 사용자가 제공한 질의에서 추출된 특징과 사용자의 의도간에 의미적인 차이(semantic gap)를 가지게 되는 문제점이 있으며 이것은 영상 검색의 성능을 저하 시키는 원인이 된다.

이러한 문제를 해결하기 위해서 최근 전경 영상을 대상

으로 의미적인 정보를 추출하여 분류에 이용하는 연구가 진행되어 오고 있다. Vailaya[1]는 에지특징을 이용하여 도시/풍경영상을 분류하는 방법을 제안하였으며, Szummer[2]는 칼라와 질감의 저수준 특징을 이용해서 실내/실외영상을 분류할 수 있는 고수준의 영상 정보를 추출해 내었다. Oliva[4]는 영상의 파워스펙트럼 정보를 이용하여 인공 영상과 자연 영상을 분류하는 방법을 제안하였다. 이러한 분류를 통한 인덱싱은 고수준 단계에서 불필요한 영상을 걸러 내어 검색의 시간을 줄일 수 있기 때문에 CBIR의 성능을 향상 시키는데 큰 역할을 한다.

위의 연구들은 모두 전경(scenery)영상을 대상으로 한다. 하지만 CBIR에서 영상에 포함된 객체의 역할을 대단히 중요하다. 왜냐하면 사용자는 일반적으로 관심객체가 포함되어 있는 영상을 찾기를 원하기 때문이다. 예를 들어 초원의 배경에 호랑이가 있는 영상이 질의로 주어졌을 경우, 사용자는 호랑이 영역에 관심을 가지며 배경에 관계없이 호랑이가 있는 영상이 검색 결과로 제공 되길 원할 것이다. 따라서 영상을 객체단위로 다루는 것은 내용기반의 성능을 크게 향상 시킬 수 있는 핵심적인 방법이 된다. [5]에서의 영상을 객체/비객체 영상으로 분류함으로써 내용기반 검색의 효율성을 높일 수 있다고 제안한다. 또한 이러한 분류는 객체 추출이나 객체 타입의 분류[6]와 같은 응용에서, 전처리 작업으로 사용될 수 있다. [3]에서는 객체를 30가지 타입으로 정의하고 웨이블릿 계수에 의해 구해진 질감 정보와 신경회로망을 이용하여 객체를 분류한다.

본 논문에서는 객체를 인공객체와 자연객체로 분류하는 방법을 제안한다. 이러한 분류는 [3]에서 제안한 것에 비해 더욱 일반적인 분류이다. 또한 일반적인 분류는 세부적인 분류에 비해 더욱더 고수준의 지식을 필요로 한다.

인공객체의 경우 직선의 에지를 많이 가지는 반면 자연객체는 다양한 방향을 가지는 복잡한 곡선의 에지를 포함하는 경향이 있다. 뿐만 아니라 인공객체의 질감은 규칙적이고 단순하며 반복적인 특징을 가지는 반면 자연객체는 복잡하고 불규칙한 특징을 가지고 있다. 이러한 고수준의 정보를 EDH(edge Direction Histogram)이나 가버 필터(gabor filter)를 이용해 추출해서 분류에 사용한다. 또한 본 논문에서는 객체의 형태 정보를 분류에 이용한다. 일반적으로 동일한 객체라 할지라도 시점에 따라서 다른 형태로 보이며, 영상에 투영된 객체만으로는 객체의 완벽한 형태 정보를

추출해내기란 어려운 일이기 때문에 객체의 분류에는 적합하지 않다. 그럼에도 불구하고 영상에서 객체의 형태 정보는 인공/자연객체 분류에 유용하다. 이것은 객체의 정확한 형태 정보를 이용하기 보다는 각 클래스가 가지는 대략적인 형태 정보만을 이용하기 때문이다.

2. 객체영상 분류

2.1 EDH 에너지를 이용한 객체 영상의 분류

에지 정보는 인공객체와 자연객체의 분류에 유용하게 사용될 수 있는 특징이다. 이것은 인공객체의 경우 자연객체에 비해 직선 에지를 많이 가지고 있기 때문이다. 따라서 인공객체의 EDH는 에지 화소가 특정 방향에 대해 높은 빈도를 가지는 반면, 자연객체는 다양한 방향에 대해 고루 분포한다. 이때 객체에서 발견되는 원형에지는 인공 및 자연객체 모두에서 발견되는 특징으로써 분류 정확성을 감소시키는 원인이 된다. 따라서 객체의 원형에지를 제거하는 전처리 작업 후에 EDH를 추출한다. 그리고 에지의 분포가 얼마나 밀집되어 있는지를 측정하기 위해 EDH의 에너지를 계산하여 분류에 이용한다.[8]

2.2 푸리에 스펙트럼의 EDAS(Energy Difference of Adjacent Sector)를 이용한 분류

인공객체의 내부에는 자연객체에 비해 직선의 에지가 많이 분포한다. 이때 객체 내부에 존재하는 직선의 에지는 방향성을 가지는 스펙트럼을 만들어 낸다. 따라서 이러한 특징은 인공/자연객체의 분류에 유용하게 사용될 수 있다.

그림 1은 주파수 공간을 등각도로 N 섹터 개수만큼 분할한 것을 보여준다. 이때 $i(0 \leq i \leq N-1)$ 섹터에서의 에너지를 se_i 라고 정의 하며 인접 섹터의 에너지 차이 EDAS는

$$EDAS = |se_0 - se_1| + \dots + |se_{N-2} - se_{N-1}| + |se_{N-1} - se_0| \quad [1]$$

으로 정의한다.

실제 구현에서는 $16(N=16)$ 개의 방향과 4개의 스케일을 가지는 가버필터뱅크를 이용하며 이때 섹터 에너지 se_i 는 동일한 방향을 가지는 가버 필터에 의해 필터링된 결과의 에너지들의 합이 된다.

그림 2는 인공객체와 자연객체의 EDAS 분포를 보여준다.

직선의 에너지가 많이 분포하는 인공객체의 경우 자연객체에 비해 높은 EDAS 가정을 알 수 있으며, 이것은 인공/자연객체의 분류에 유용함을 보여준다.

분류를 하기 위해서 트레이닝 영상들에 대해서 분류에러를 최소화하는 적절한 분류 임계치를 계산한 후 테스트 영상의 인접색터 에너지 차이와 비교한다.

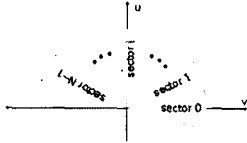


그림 1 퓨리에 스펙트럼에서 섹터 분할

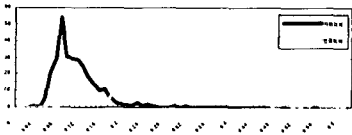


그림 2 인공객체와 자연객체의 EDAS 분포

2.3. 질감(Texture) 특징을 이용한 분류

객체의 질감은 인공객체와 자연객체에서 큰 차이를 보인다. 예를 들어 인공객체에서 나타나는 질감은 대개 규칙적이고 반복적인 형태로 나타나는 반면 자연객체는 그렇지 못하다. 이러한 특징은 객체를 분류하는데 유용하게 사용될 수 있다.

본 논문에서는 질감 특징을 추출하기 위해 6개의 스케일과 4개의 방향을 가지는 가버필터뱅크를 이용해서 24차원의 특성 벡터를 추출한다. 하지만 가버 필터는 임의의 형태의 특징 영역만을 대상으로 적용하기에는 어려움이 따른다. 이 문제를 해결하기 위해서 먼저 영상에서 비객체 영역을 RGB(0,0,0)값으로 할당하여 비객체 영역의 질감 정보를 배제시킨다. 그리고 객체의 경계를 블러링함으로써 객체의 경계가 질감 특성에 미치는 영향을 감소시킨다. 그 후 가버필터를 이용해서 객체영역의 질감 정보를 추출한다.

분류를 하기 위한 가장 간단한 방법으로 k-NN 분류방법이 있다. 하지만 이 방법은 트레이닝 영상의 수에 비례하여 시간 복잡도가 증가하므로 효율적이지 못하다. 따라서 분류 효율성을 위해, 먼저 트레이닝 영상에 대해 k-means 알고리즘을 이용해 감독 군집화를 한다. 이때 동일한 클러스터에 속하는 영상 중에 높은 비율을 가지는 클래스를 그 클러스터의 클래스로 정의 하며, 이 때 중심 벡터는 클러

스터의 특징을 대표하게 된다.

객체를 분류하기 위해서는 객체의 질감 특징을 추출한 후 감독 군집화 과정에서 생성된 중심벡터 중 가장 가까운 벡터의 속성으로 분류한다. 중심벡터의 수를 결정하는 k값은 실험에 의해 적절히 결정한다.

2.4. 형태 특성을 이용한 분류

영상에 투영되는 동일한 객체의 형태는 시점에 따라 다양하게 나타난다. 따라서 다양한 시점을 가질 수 있는 사진 영상의 경우, 객체의 형태 정보를 분류 특징으로 사용하는데 어려움이 따른다. 하지만 인공객체영역의 경계를 이루는 부분은 자연객체에 비해 직선의 형태를 띄며 형태가 단순하다. 이것은 객체의 대략적인 형태 정보가 인공객체와 자연객체를 분류할 수 있는 유용한 정보가 될 수 있음을 의미한다.

본 논문에서는 객체의 형태정보를 추출하기 위해 먼저 영상을 객체 영역과 비객체 영역의 이진 영상으로 만든다. 그 후 6개의 방향과 4개 스케일을 가지는 가버필터뱅크를 적용하여 24차원의 특성 벡터를 생성한다.

분류 방법으로는 질감을 이용한 분류와 마찬가지로 k-means 알고리즘을 이용해 감독 군집화를 하여 생성된 각 클러스터의 중심 벡터를 이용한다.

3. 다중 특징을 이용한 분류 방법

본 논문에서는 다수의 분류 방법에 대한 결과들을 머징하여 방법을 제안한다. 일반적으로 분류 방법이 3개 이상 존재할 때, 각 방법에서의 분류 결과를 토대로 다수결의 원칙을 사용하여 분류를 할 수 있다. 하지만 분류된 결과들이 같은 비율로 상충될 경우 다수결에 의한 분류가 모호해진다. 이 문제를 해결하기 위해 각각의 분류 방법의 성능을 기초로 가중 평균하는 방법을 이용한다.

$$p_{manmade} = \sum_i (a_i \cdot m_i + (1 - m_i) \cdot (1 - a_i)) \quad [2]$$

$p_{manmade}$ 은 자연객체로 분류될 확률이다. $a_i (0 \leq a_i \leq 1)$ 는 i 번째 방법의 분류 정확성을 의미한다. m_i 는 i 번째 방법으로 분류했을 때 인공객체로 분류되었을 경우 1, 그렇지 않을 경우 0을 가진다. $p_{manmade}$ 가 0.5보다 클 경우 인공객체로 분류하며, 그렇지 않을 경우에는 자연 객체로 최종 분류한다.

제안한 방법은 정확성이 더 높은 분류 방법을 더 신뢰함으로써 보다 다수결에 의한 모호한 결과를 없애고 좀 더 정확한 분류 결과를 얻을 수 있게 한다.

3. 실험 및 토의

본 논문에서 실험을 위하여 코렐 포토 영상에서 객체가 있는 영상 중에 임의로 인공객체영상과 자연객체영상을 각각 300개씩 선택하였다. 이때 해프톤 영상과 손상된 영상은 실험에서 제외 시켰다.

그림 3은 질감 특징에 대해 군집화 알고리즘으로 생성된 클러스터의 중심 벡터를 이용해서 분류했을 때의 차이를 보여 주고 있다. 그림에서 볼 수 있듯이 소수의 중심 벡터를 이용한 분류가 k-NN을 이용한 분류에 비해 정확성 면에서 큰 차이를 보이고 있지 않는다. 반면 시간 복잡도 면에서는 군집화 알고리즘이 높은 성능을 보임을 예상할 수 있다.

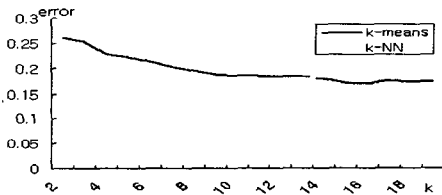


그림 3 클러스터 개수에 변화에 따른 분류 정확성

각 분류 방법을 평가하기 위해서 600개의 테스트 영상에 대해 임의로 400개를 트레이닝 영상으로 선택하고 나머지 200개의 영상에 대해서 분류 정확성을 계산하며, 이를 100회 반복하여 분류 정확성을 평균한다. 질감 특징을 사용한 경우 클러스터 개수를 12로 정했으며, 형태 특징을 이용한 경우에는 15로 결정 했다. (이것은 k-NN 분류를 통해 얻을 수 있는 정확성을 기준으로 함)

표 1 각각의 특징을 이용한 분류 결과

feature	energy of EDH	shape	EDAS	Texture	머징 결과
accuracy(%)	82.245	76.78	85.08	83.15	89.203

표 1은 각 분류 방법의 성능과 제안된 머징 방법을 이용해서 최종 분류한 결과를 나타낸다.

4. 결론 및 향후 연구

본 논문에서는 객체영상을 인공객체영상과 자연객체영상

으로 자동 분류하는 방법을 제안하였다. 이것은 객체 분류에서 가장 상위 분류에 속하며, 이 연구의 결과로 영상데이터베이스의 인덱싱에서 큰 성능 향상을 얻을 것으로 기대 된다.

제안된 방법 중 트레이닝 영상을 군집화를 할때, 초기 시드에 따라 군집의 형태가 다르게 나타난다. 이는 분류 정확성에 영향을 미치며, 이를 해결 하기 위해 최적화 군집을 위한 초기 시드 선택 방법이 적용되어야 할 것이다.

[참고 문헌]

- [1] Vailaya, A., Jain, A.K., and Zhang, H.J., "On image classification: city images vs landscapes," Pattern Recognition., 1998, pp. 1921-1936
- [2] M. Szummer, R.W. Picard, "Indoor-outdoor image classification," IEEE Int'l Workshop Content-Based Access Image Video Databases, 1989, pp. 42-51
- [3] Park, S.B., Lee, J.W., and Kim, S.K. "Content-Based image classification using a neural network," Pattern Recognition Letter, 2004, pp. 287-300
- [4] Oliva, a., Torralba, A., Dugue, A.G., and Hérault, J. "Global Semantic Classification of Scenes using Power Spectrum Templates," CIR'99. Newcastle. UK., 2003
- [5] Kim, S.Y, Park, S.Y., and Kim, H.H, "Image Classification into Object/Non-object Classes," Int'l Conf. on Image and Video Retrieval(CIVR), to be appear
- [6] Kim, S.Y, Park, S.Y., and Kim, H.H, "Central Object Extraction for Object-based Image Retrieval," Int'l Conf. on Image and Video Retrieval(CIVR), 2003, pp.39-49
- [7] Yossi Rubner, "Perceptual metrics for image database navigation," Computer Science Department Stanford University, 1999
- [8] 구경모, 박창민, 김민환, "인공객체영상 및 자연객체영상에 대한 자동분류," 한국멀티미디어학회 춘계학술발표대회논문집, 2003, pp. 330-333