내용기반 영상검색을 위한 객체 및 비객체 영상의 분류 방법

박소정*°, 김성영**, 김민환*
* 부산대학교 컴퓨터공학과
** 금오공과대학교 컴퓨터공학부

Image Classification into Object/Non-object Classes for Content-based Image Retrieval

SoJeong Park*, SungYoung Kim**, MinHwan Kim*
* Dept. of Computer Engineering, Pusan National University
** School of Computer Engineering, Kumoh National Institute of Technology

요 약

본 논문에서는 영상을 자동적으로 객체와 비객체 영상으로 분류하는 방법을 제안한다. 객체 영상은 객체를 포함하는 영상이다. 객체는 영상의 중심 부근에 위치하고 주변 영역과는 상이한 칼라 분포를 가지는 영역들로 정의한다. 영상 분류를 위해 객체의 특징에 기반하여 세 가지 기준을 정의한다. 첫 번째 기준인 중심 영역의 특이성은 중심 영역과 주변 영역간의 칼라 분포의 차이를 통해 계산된다. 두 번째 기준은 영상 내의 특이 픽셀의 분산이다. 특이 픽셀은 영상의 주변영역보다 중심 부근에서 더욱 빈번하게 나타나는 상호 인접한 픽셀들의 칼라 쌍에 의해 정의된다. 마지막 기준은 객체의 핵심 영역 경계에서의 경계 강도이다. 영상을 분류하기 위해서 신경 회로망 학습을 통해서 세 가지 기준들을 통합하도록 한다. 900 개의 영상들에 대해 실험한 결과 84.2%의 분류 정확도를 얻었다.

1. 서론

내용기반 영상검색에서 영상은 시각적인 특징 정보를 요약하여 자동으로 색인된다. 또한 칼라, 질감, 모양, 공간적 분포와 같은 저수준의 특징 정보에 기반하여 검색되고 비교된다. 그러나 사용자의 질의를 이와 같은 저수준의 영상 특징에 기반하여 표현하는 경우, 사용자가 의도했던 실제 내용과는 상당한 의미적 차이(semantic gap)가 있다. 이러한 의미적 차이를 극복하기 위해 초기 입력이나 검색 과정에서 사용자가 직접 개입하여 검색의 방향을 지시하는 결과 내 재검색(relevance feedback)에 대한 연구가 이루어져왔다. 또한 의미적인 특징 정보를 찾기 위한 방법으로 영상으로부터 직접적으로 의미정보를 추출하는 노력들이 있었다. 예를 들면 실내/실외, 도심/경치[1-2] 등과 같은 일반적인 유형으로 영상을 자동 분류하는 방법들이 있다.

반면, 많은 연구자들은 효과적인 내용기반검색 성능의 핵심이 객체 수준에서 영상을 접근하는 능력에 달려 있다고 한다. 왜냐하면 일반적으로 사용자들은 특정한 관심객체를 포함하는 영상을 검색하기를 원하기 때문이다. 그러므로, 내용기반영상검색에서 객체의 포함 여부를 판단하는 과정은 매우 중요한 단계이다. 객체/비객체 영상 분류를 통해 다른

클래스로 분류되는 영상을 제거함으로써 검색 성능을 향상시킬 수 있다. 예를 들면 사용자가 '호랑이' 영상을 검색하고자 할 때, 검색된 카테고리는 단지 객체 영상으로서 한정 지을 수 있다. 또한 객체/비객체 영상의 분류는 객체분류영상에서부터 객체를 추출[3,4]하거나 영상검색 성능을 개선시키기 위해 객체의 유형을 분류하는 객체기반응용의 전처리 단계에 유용하게 사용될 수 있다.

객체영상은 객체가 존재하는 영상이라고 할 수 있다. 그러나 영상에서의 관심객체의 정의는 아직도 해결되지 않은 문제이다[3]. 사용자의 관점에 따라 관심객체의 정의는 주관적이다. 예를 들면, 수평선 위에 불그스름한 떠오르는 해가 있는 영상을 객체라고 간주할 수 있다. 하지만 일반적으로 관심객체는 아니다. 본 논문에서 일출영상은 객체를 포함하기 때문에 객체영상이라고 가정한다.

본 논문에서는 영상을 객체와 비객체로 자동적으로 분류하는 방법을 제안한다. 일반적으로 영상에서 객체란 중심 부근에 위치하고 주변과는 상이한 칼라 특징으로 표현된다. 객체 영역은 일정한 가로/세로비를 갖는 경향이 있고 주변영역에 비해 강한 경계를 가지고 있다. 이에 본 논문에서는 객체의 특징에 기반하여 세가지 분류 기준을 정의한다. 첫 번째는 주변영역에 대한 중심영역의 칼라 특이성을 나타낸다. 두 번째는 특이픽셀의 수평 및 수직 분산 중에서 큰 값으로 정의한다. 세 번째는 핵심객체영역[3]의 평균 경계강도로 정의된다. 우리는 이전 연구[7]에서 네 가지의 분류 기준을 순차적으로 적용하여 79.6%의 분류 정확도를 얻었다. 하지만 본 논문에서는 신경회로망을 사용하여 세가지 분류 기준을 효과적으로 조합함으로써 분류 성능을 향상시킬 수 있도록 한다.

2. 객체의 정의

칼라영상에서의 객체는 다음과 같은 조건을 만족하는 경향이 있다.

(1) 영상의 중심부근에 위치한다.

(2) 주변영역과는 상이한 칼라 또는 텍스쳐로 구성된다.

(3) 비교적 크기가 크다.

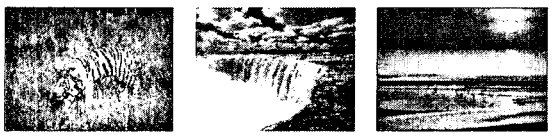
(4) 경계픽셀은 비교적 강한 에지를 가진다.

(5) 모양이 한정되어 있다.

(6) 특정한 의미있는 키워드와 관련이 있다.

(1)~(4)의 조건은 객체를 표현하는데 유용한 정보임이 확인되었다[3]. 추가로 조건 (5)는 시간의 흐름에 따라 모양이 변하는 것은 객체에서 제외한다는 것을 의미한다. 조건 (6)은 쉽게 측정할 수는 없지만 명백하다.

[그림 1(a),(b)]는 영상의 중심 부근에 위치하며 각각 '호랑이' 및 '폭포'라는 키워드를 할당할 수 있는 객체를 나타낸다. 따라서 이 두 영상은 객체 영상으로 분류할 수 있다. 반면 [그림 1(c)]에는 의미 있는 영역도 포함되어 있지 않으므로 비객체 영상이다. 한편 [그림 1(b)]의 폭포는 우리의 관심이 특정한 영역이 아닌 영상의 전체영역이 되므로 관심객체는 아닐 것이다. 사용자가 일반적으로 특정한 관심객체를 포함하는 영상을 찾기를 원하기 때문에 관심객체의 개념은 내용기반검색에 있어서 매우 중요하다. 그러나 저수준의 영상 특징을 가지고 관심객체를 설명하기란 어렵다. 따라서 본 논문에서는 관심 객체가 아니라 객체가 영상 내에 포함되어 있는지 여부를 판단하는데 초점을 맞추었다.



(a) 호랑이 (b) 폭포 (c) 일몰

[그림 1] 객체 영상과 비객체 영상의 예

3. 객체 및 비객체 영상 분류

3.1 관심창 내부의 특이 칼라

특이 칼라[3]은 [수식 1]을 만족하는 칼라의 쌍 (c_i, c_j)으로 정의된다. [수식 1]에서 $C_{DAW}(c_i, c_j)$ 및 $C_{SR}(c_i, c_j)$는 각각 초기 관심창의 코릴로그램(C_{DAW}) 및 주변 영역의 코릴로그램(C_{SR})의 칼라 (c_i, c_j)의 누적 개수를 나타낸다. 관심창은 영상의 중심에 위치하면서 높이와 폭이 영상의 절반크기인 사각형으로 정의한다. 또한, 관심창 외부 영역을 본 논문에서는 주

변영역이라고 정의한다. 상호 밀접하게 연관된 칼라 (c_i, c_j) 에 대한 특이픽셀은 칼라 c_j 를 가지면서 인접한 칼라 c_j 를 가지는 픽셀을 말한다.

$$\frac{C_{DAW}(c_i,c_j)-C_{BR}(c_i,c_j)}{C_{DAW}(c_i,c_j)} \geq 0.1 \qquad (1)$$

3.2 분류 기준

3.2.1 중심 영역의 특이성

중심 특이성은 [수식 2]와 같이 관심창이 주변 영역에 대해 갖는 칼라분포의 특이성으로 나타난다. *CS* 는 영상의 중심 부근의 특이 픽셀들의 밀도에 비례하여 증가한다. [그림 2(a)]와 같이 관심객체인 호랑이의 위치에서는 특이 픽셀들이 조밀하게 나타나며 또한 큰 중심 특이성으로 나타난다. 반면 [그림 2(b)]의 비객체 영상은 중심 특이성이 낮게 나타난다. 하지만 [그림 2(c)]와 같이 배경 칼라가 점진적으로 변화하는 비객체 영상에서는 중심 특이성이 높게 나타남으로써 분류 성능을 저하시킬 수 있다.

$$CS = \frac{\sum_i \sum_j \max\left(C_{DAW}(c_i,c_j)-C_{BR}(c_i,c_j),0\right)}{\sum_i \sum_j C_{DAW}(c_i,c_j)} \qquad (2)$$

(a) CS=0.466 (b) CS=0.002 (c) CS=548

[그림 2] 영상들에 대한 특이 픽셀 분포

3.2.2 특이 픽셀의 최대 분산

비객체 영상에서의 특이 픽셀은 객체 영상보다 공간적으로 넓게 분포하는 경향이 있다. 그러므로 특이 픽셀의 수평 또는 수직 방향의 분산은 객체영상과 비객체 영상을 구별하는 기준으로 사용 가능하다. 두 번째 기준은 [수식 3]과 같이 수평방향의 분산 V_x 과 수직방향의 분산 V_y 중 최대값으로 정의된다. P_x 와 P_y 는 특이 픽셀의 x 및 y 좌표이고, W 와 H 는 영상의 폭과 높이를 나타낸다. 그리고 N 은 영상내의 특이 픽셀의 총 개수를 나타낸다.

$$V_x(SP) = \frac{\sum\left((P_x - m_x)^2 / W\right)}{N},\ V_y(SP) = \frac{\sum\left((P_y - m_y)^2 / H\right)}{N} \qquad (3)$$
$$V(SP) = Max(V_x(SP), V_y(SP))$$

3.2.3 핵심 객체 영역의 경계강도

세 번째 분류 기준은 핵심 객체 영역[3]의 경계에서의 경계강도이다. 중심 객체 추출 방법[3]은 객체가 존재하는 경우에는 좋은 성능을 나타내지만 객체가 존재하지 않는 경우에도 의미 없는 영역을 객체로 추출한다. 하지만 실제 객체의 경계에서는 높은 경계 강도가 나타나는 반면 상대적으로 비객체 영상에서는 약하게 나타난다. 핵심 객체 영역의 경계강도는 [수식 4]에 의해 정의된다. 여기서 ∇f_i 는 i번째 경계 픽셀의 경계강도를 나타내고 N 은 경계 픽셀의 총 개수를 나타낸다.

$$E(CO) = \frac{\sum_i \nabla f_i}{N} \qquad (4)$$

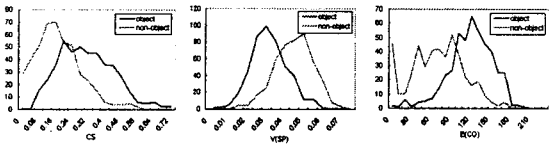
3.3 신경회로망에 의한 분류

설명된 분류 기준들은 그 자체로서 유용하지만 분류 정확도는 세가지 방법의 조합에 의해 더욱 개선될 수 있다. 우리는 이전[7]에 분류 기준들을 순차적으로 적용함으로써 영상 분류를 수행하였는데, 각각의 기준들에 적당한 가중치를 할당하는 것이 어려워 분류 정확도가 떨어졌다. 이에 본 논문에서는 분류 기준들을 조합하는 방법으로 오류역전파 신경회로망을 적용하였다.

4. 실험 결과 및 토의

코렐 Gallery photo-CD에서 무작위로 선정한 객체 영상 450개와 비객체영상 450개, 총 900개의 영상을 대상으로 영상 분류 실험을 수행하였다. 우선 각 분류 기준들의 구별능력을 평가하였다. [그림 3(a)-(c)]는 객체 및 비객체 영상들에 대한 중심 영역의 특이성, 특이 픽셀의 최대 분산, 핵심 객체 영역의 경계강도 분포 곡선을 나타낸다. 그림 보듯이 객체 및 비객체 영상에 해당되는 각각의 곡선이 분명하게 분리되지 않고 있는 것을 알 수 있다. 따라서 각각

의 기준들이 그 자체만으로는 좋은 성능을 제공하지는 못한다.



(a) CS (b) V(SP) (c) E(CO)

[그림 3] 세가지 기준에 대한 측정치의 분포

주어진 데이터가 작아서 학습 데이터에 편향되는 경향을 방지하기 위하여 오류역전파 신경회로망은 6-폴드 크로스 밸리데이션[6] (6-fold cross-validation)을 사용하여 평가하였다. n-폴드 크로스 밸리데이션 (n-fold cross-validation)는 주어진 데이터는 n개의 동일한 크기의 폴드(fold)로 분할하고 각각의 폴드를 평가데이타, 나머지 n-1개의 폴드를 학습 데이터로 나누어 평가한다. 여기서 n개의 폴드로 분할하는 작업 또한 무작위로 n번 수행한 후 평균적인 성능을 측정한다.

[표 1] 객체 및 비객체 영상의 분류 평가 결과

	객체	비객체
Precision	0.90	0.80
Recall	0.78	0.90
F-measure	0.83	0.85

[표 2] 각 방법들을 독립적으로 적용한 분류 평가

		CS	V(SP)	E(CO)	조합
객체	F-measure	0.70	0.75	0.78	0.83
비객체	F-measure	0.71	0.76	0.76	0.85
정확도		70.2%	75.3%	77.1%	84.2%

[표 1]은 신경회로망 학습기법을 6-폴드 크로스 밸리데이션으로 평가했을 때의 객체와 비객체 영상의 평가 결과를 보여준다. [표 2]는 앞서 말한 분류방법 세가지를 조합하지 않았을 때와 조합했을 때 평가 결과이다. 각 방법들에 대한 분류 기준치를 설정하는 것이 어려웠으나, 세가지 방법을 조합하여 84.2%의 정확도를 보였다. 이것은 각각의 방법들이 가치가 있다는 것을 보여준다.

5. 결론

본 논문에서는 84.2%의 정확도를 가지는 객체 및 비객체 영상 분류 방법을 제안하였다. 영상 분류를 위해 세 가지 분류 기준(중심 영역의 칼라 분포 특이성, 특이 픽셀의 최대 분산, 핵심 객체 영역의 경계강도)을 기반으로 신경회로망 학습기법을 이용하였다. 제안된 방법은 영상검색과 영상 색인의 성능을 개선하는데 유용하게 사용될 수 있다.

[참고문헌]

[1] Vailaya, A., Jain, A.K., and Zhang, H.J., "On Image Classification: City images vs. landscape", Pattern Recognition. ,1998, pp.1921-1936

[2] Vailaya, A., Figueiredo, M.A.T., Jain, A.K., and Zhang, H.J., "Image Classification for Content-Based Indexing", IEEE Trans. on Image Processing, 2001, pp.117-130

[3] Kim, S.Y., Park, S.Y., and Kim, M.H., "Central Object Extraction for Object-Based Image Retrieval", Int'l Conf. on Image and Video Retrieval (CIVR) , 2003, pp.39-49

[4] Serra, J.R. and Subirana, J.B., "Texture Frame Curves and Regions of Attention Using Adaptive Non-cartesian Networks", Pattern Recognition , 1999, pp.503-515

[5] 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

[6] Witten, I.H., Frank, E., "Data Mining", Academic Press , 2000

[7] 박소연,김성영,김민환, "칼라분포특성을 이용한 객체 및 비객체 영상의 분류 방법", 한국멀티미디어회. 추계학술발표대회논문집, 2003, pp.318-321