

관심 객체 검출에 기반한 객체 및 비객체 영상 분류 기법

김성영*

Object/Non-object Image Classification Based on the Detection of Objects of Interest

Sung-Young Kim*

요약

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 기반을 두고 네 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. 세 번째 기준은 중심 객체의 평균 경계강도이다. 세 번째 기준은 분류 기준들중에서 가장 우수한 분류 성능을 나타내지만 특징값을 추출하기 위해서는 중심 객체를 추출해야 되는 많은 연산을 내포하고 있다. 이에 이와 비슷한 특성을 나타내는 네 번째 기준으로 영상 중심 영역에서의 평균 경계강도를 선택하였다. 네 번째 분류 기준은 세 번째 분류 기준에 비해 분류 성능은 조금 낮지만 빠르게 특징값을 추출할 수 있어 많은 데이터를 빠른 시간 내에 처리해야 되는 대규모 영상 데이터베이스에 적용가능하다. 영상을 분류하기 위해 신경회로망 및 SVM을 사용하여 이들 기준들을 통합하였으며 신경회로망 및 SVM의 분류 성능을 비교하였다.

Abstract

We propose a method that automatically classifies the images into the object and non-object images. An object image is the image with object(s). An object in an image is defined as a set of regions that lie around center of the image and have significant color distribution against the other surround (or background) regions. We define four measures based on the characteristics of an object to classify the images. The center significance is calculated from the difference in color distribution between the center area and its surrounding region. Second measure is the variance of significantly correlated colors in the image plane. Significantly correlated colors are first defined as the colors of two adjacent pixels that appear more frequently around center of an image rather than at the background of the image. Third one is edge strength at the boundary of candidate for the object. By the way, it is computationally expensive to extract third value because central objects are extracted. So, we define fourth measure which is similar with third measure in characteristic. Fourth one can be calculated more fast but show less accuracy than third one. To classify the images we combine each measure by training the neural network and SVM. We compare classification accuracies of these two classifiers.

▶ Keyword : 관심객체(Object of Interest), 중심객체(Central Object), 영상분류(Image Classification), SVM

• 제1저자 : 김성영

• 접수일 : 2006.02.13 심사완료일 : 2006.05.18

* 금오공과대학교 컴퓨터공학부 조교수

※ 본 연구는 금오공과대학교 학술연구비에 의하여 연구된 논문임.

1. 서론

내용기반영상검색(Content-Based Image Retrieval, CBIR)에서는 사용자에게 의해 지정된 특정한 장면이나 객체를 나타내는 모든 영상들을 주어진 데이터베이스로부터 검색한다. 데이터베이스에 포함된 영상들은 영상내의 시각적인 특징 정보를 요약하여 자동으로 색인된다. 영상 검색 단계에서도 영상으로부터 직접 추출된 이런 시각적인 특징 정보에 기반을 두어 검색되고 비교한다. 내용기반영상검색에 주로 사용되는 영상의 시각적인 정보는 주로 색상, 질감, 모양, 공간적 분포와 같은 저수준의 특징 정보가 일반적이다. 이런 저수준 특징 정보들은 쉽게 측정되고 계산될 수 있지만 사용자가 질의 단계에서 의도했던 내용을 검색 질의어에 정확하게 표현하지 못함으로써 검색 질의어와 사용자가 의도했던 실제 내용과는 상당한 의미적 차이(semantic gap)가 발생한다. 이런 의미적 차이에 의해 검색 결과에 상당수의 부정확한 결과 영상들이 포함된다.

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 사용자들이 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 한정될 수 있다. 일반적으로 적절한 영상 분류를 통해 데이터베이스에 포함된 영상들에 대해 인텍싱을 수행함으로써 내용기반영상검색 시스템의 성능을 상당히 향상시킬 수 있다. 이는 질의와 관련 없는 영상들을 사전에 걸러냄으로써 검색

시간과 검색 정확도를 향상시킬 수 있기 때문이다. 방대한 영상 데이터베이스로부터 영상 분류를 통해 의미기반의 인덱스 정보를 생성하기 위해 많은 연구들이 시도되었다[1-4]. 또한 객체/비객체 영상의 분류는 객체분류영상에서부터 객체를 추출[5,6]하거나 영상검색 성능을 개선시키기 위해 객체의 유형을 분류[7]하는 객체기반응용의 전처리 단계로 유용하게 사용될 수 있다. 또한 MPEG-7 관련 분야에서도 활용가능하다[14].

객체영상은 객체가 존재하는 영상이라고 할 수 있다. 사용자는 일반적으로 특정한 관심객체를 포함하는 영상을 찾기를 원하기 때문에 관심객체의 개념은 내용기반검색에 있어서 매우 중요하다. 그러나 영상에서의 관심객체의 정의는 아직도 해결되지 않은 문제이다[6]. 사용자의 관점에 따라 관심객체의 정의는 주관적이다. 본 논문에서는 중심 객체라는 개념을 정의하여 객체 포함 여부를 판단하는데 기준으로 사용한다. 중심 객체는 피실험자에 의해 선택된 객체에 대한 특징을 통계적으로 조사하여 표현하였다. 중심 객체는 영상 검색의 대상이 되는 관심 객체와 비교적 일치하는 것을 확인할 수 있다[5]. 중심 객체는 II장에서 보다 자세하게 정의한다.

본 논문에서는 영상을 객체와 비객체로 자동적으로 분류하는 방법을 제안한다. 일반적으로 영상에서 객체란 중심 부근에 위치하고 주변과는 상이한 칼라 특징으로 표현된다. 객체 영역은 일정한 가로대 세로비를 갖는 경향이 있고 주변영역에 비해 강한 경계를 가지고 있다. 본 논문에서는 이런 객체의 특징에 기반을 두고 네 가지 분류 기준을 정의하였다. 첫 번째는 주변영역에 대한 중심영역의 칼라 특이성을 나타낸다. 두 번째는 특이픽셀의 수평 및 수직 분산 중에서 큰 값으로 정의한다. 세 번째는 중심 객체의 평균 경계강도로 정의된다. 네 번째는 영상 중심 영역의 평균 경계강도를 나타낸다. 세 번째와 네 번째 분류 기준은 모두 경계강도를 사용하고 있으며 비슷한 성격은 나타낸다. 중심 객체의 평균 경계강도가 가장 우수한 분류 성능을 나타내고 있지만 이 분류 기준은 중심 객체의 추출이 완료되어야 측정할 수 있으므로 많은 계산량이 필요하다[5]. 또한 객체/비객체 영상의 분류는 중심 객체의 추출을 위해 사전에 영상을 분류하는 전처리 과정으로 활용되어야함에도 불구하고 영상 분류를 위해 중심 객체를 추출하는 과정을 요구하고 있다[12]. 반면 중심 영역의 평균 경계 강도는 중심 객체의 평균 경계 강도에 비해서는 분류 정확도가 조금 낮기는 하지만 간단한 연산으로 빠르게 계산될 수 있는 장점이 있다. 이에 본 논문에서는 분류 목적에 따라 세 번째 및 네 번째

기준을 선택하여 사용할 수 있도록 한다. 예를 들어, 많은 데이터를 적정 시간 내에 처리해야 되는 대규모 내용기반영상검색 시스템에서는 빠른 처리 시간이 우선적으로 요구되는 반면 규모가 작은 영상 데이터베이스의 경우에는 처리 속도에 비해 높은 정확도를 필요로 할 수 있다. 물론 네 가지 분류 기준을 모두 사용함으로써 가장 높은 분류 성능을 얻을 수 있다. 그리고 본 논문에서는 이들 분류 기준을 조합하여 분류 정확도를 계산하기 위해 식별자로서 신경회로망 및 support vector machine(SVM)을 사용하였고 이들 두 가지 식별자의 성능을 비교하였다. SVM은 구조적인 위험을 최소화시키는 방법으로 수행되며 일반적으로 신경회로망에 비해 높은 분류 성능을 나타내는 것으로 알려져 있다[10].

II. 객체의 정의

영상의 중심 영역은 경계 주변에 비해 객체에 대한 중요하고 많은 정보를 제공한다. 일반적으로 사람들은 그림을 그리거나 사진을 찍을 때 관심이 있고 흥미로운 대상을 영상의 중심에 위치시키는 경향이 있다. 이와 같이 영상 내에 위치하는 객체들은 영상의 중심 부근에 위치하는 확률이 높다. 객체를 배경과 분리하기 위한 기존의 여러 방법에서도 영상 중심의 중요성을 강조하고 있다. 내용 기반 영상 검색을 위해 객체를 배경으로부터 분리하기 위해 시도했던 방법 [8]에서는 배경을 분리하기 위해 영상의 중심 영역 및 경계 인접 영역에 대한 위치 정보를 사용하였다. 또한 복잡한 배경으로부터 객체를 추출하기 위한 시도[5]에서는 실험적으로 영상 내부에 존재하는 객체들의 위치 분포를 확인하였다. 이 실험 결과에서는 영상에 포함된 의미 있는 객체는 영상의 중심 부근에 일정한 크기로 나타나는 것으로 주장하고 있다. 여기에서는 영상의 관심 객체를 다음과 같은 조건을 만족하는 것으로 정의하였다[12].

- (1) 영상의 중심부근에 위치한다.
- (2) 주변영역과 상이한 칼라 또는 텍스처 특징이 있다.
- (3) 비교적 크기가 크다.
- (4) 경계픽셀은 비교적 강한 에지를 가진다.

사용자가 일반적으로 특정한 관심객체를 포함하는 영상을 찾기를 원하기 때문에 관심객체의 개념은 내용기반검색에 있어서 매우 중요하다. 중심 객체 추출 결과[5]에서 위의 조건에 부합하는 중심 객체가 관심 객체와 대체적으로 일치하는 것을 확인할 수 있다.

III. 객체 및 비객체 영상 분류

3.1 관심창 내부의 특이 칼라

특이 칼라는 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. 관심창(Default Attention Window, DAW)은 영상의 중심에 위치하면서 높이와 폭이 영상의 절반크기인 사각형으로 정의한다. 또한, 관심창 외부 영역을 본 논문에서는 주변영역(Surround Region, SR)이라고 정의한다. 상호 밀접하게 연관된 칼라 (c_i, c_j) 에 대한 특이픽셀은 칼라 c_i 를 가지면서 인접한 칼라 c_j 를 가지는 픽셀을 의미한다.

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

3.2 분류 기준

3.2.1 중심 영역의 특이성

중심 특이성(Center Significance, CS)은 수식 (2)와 같이 관심창이 주변 영역에 대해 갖는 칼라분포의 특이성으로 나타난다. CS는 영상의 중심 부근의 특이 픽셀들의 밀도에 비례하여 증가한다. 그림 1(a)와 같이 관심객체인 갈매기의 위치에서는 특이 픽셀들이 조밀하게 나타나며 또한 큰 중심 특이성으로 나타난다. 반면 그림 1(b)의 비객체 영상은 중심 특이성이 낮게 나타난다. 하지만 그림 1(c)와 같이 배경 칼라가 점진적으로 변화하는 비객체 영상에서는 중심 특이성이 높게 나타남으로써 분류 성능을 저하시킬 수 있다.

$$CS = \frac{\sum_i \sum_j \max(C_{DAW}(c_i, c_j) - C_{SR}(c_i, c_j), 0)}{\sum_i \sum_j C_{DAW}(c_i, c_j)} \dots\dots\dots (2)$$

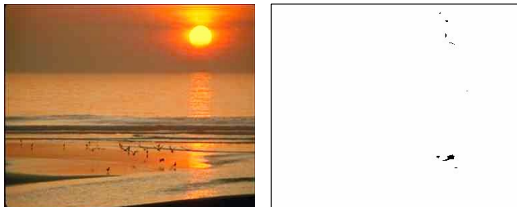
$$V_x(SP) = \sum ((P_x - m_x)^2 / W) / N$$

$$V_y(SP) = \sum ((P_y - m_y)^2 / H) / N$$

$$V(SP) = \max (V_x(SP), V_y(SP)) \dots\dots\dots (3)$$



(a) CS=0.486



(b) CS=0.002



(c) CS=548

그림 1. 영상들에 대한 특이 픽셀 분포 및 중심 영역의 특이성 값
Fig. 1. Distribution of significant pixels and value for center significance

3.2.3 중심 객체의 평균 경계강도

중심 객체의 평균 경계강도는 중심 객체로 추출된 영역에서의 평균 경계강도로 정의된다. 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. 중심 객체의 경계강도는 수식 (4)에 의해 정의된다. 여기서 ∇f_i 는 i 번째 경계 픽셀의 경계강도를 나타내고 N 은 경계 픽셀의 총 개수를 나타낸다.

$$E(CO) = \left(\sum_i \nabla f_i \right) / N \dots\dots\dots (4)$$

그런데 중심 객체의 평균 경계강도를 계산하기 위해서는 먼저 중심 객체(central object, CO)에 대한 추출이 선행되어야 한다. 중심 객체 추출[5]은 그림 2와 같이 JSEG 알고리즘[13]에 의한 영상 분할, 특이 픽셀 추출, 추출된 특이 픽셀 정보를 이용한 DAW 위치의 재설정 과정, 핵심객체영역(core object region, COR) 선택, 배경 영역 분석, 핵심객체영역 확장 등의 여러 단계를 거쳐 수행된다. 따라서 다른 분류 기준에 비해 상당한 연산량을 요구하게 되며 객체/비객체 영상 분류 작업이 객체 추출을 위한 전처리 단계가 아니라 객체/비객체를 분류하기 위한 목적으로 전략하게 된다. 이에 본 논문에서는 중심 객체의 평균 경계강도를 대신하여 조금은 낮은 분류 성능을 나타내지만 상당히 빠르게 계산될 수 있는 중심 영역의 평균 경계강도를 사용하도록 한다.

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 영상 내의 특이 픽셀의 총 개수를 나타낸다.

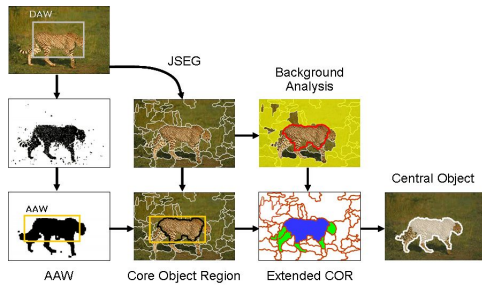


그림 2 중심 객체 추출 과정
Fig. 2. Process on Extracting Central Objects

3.2.4 중심 영역의 평균 경계강도

중심 영역의 평균 경계강도는 DAW 내부의 평균 경계 강도로 계산된다. DAW는 중심 객체의 분포 확률이 가장 높은 영역으로 설정되므로 DAW 내부의 평균 경계 강도는 객체가 존재할 경우에는 객체 내부의 평균 경계 강도를 나타낸다. 객체가 존재할 경우 객체의 경계 및 객체 내부의 질감 변화에 의해 강한 경계 강도를 나타낸다. 따라서 객체가 존재하는 경우에는 객체가 존재하지 않은 경우에 비해 높은 경계강도를 나타낸다.

3.3 객체/비객체 영상 분류

설명된 분류 기준들은 그 자체로서 유용하지만 분류 정확도는 이들 기준의 조합에 의해 더욱 개선될 수 있다. 분류 기준들을 조합하는 식별자(classifier)로서 본 논문에서는 support vector machine (SVM)[10,11] 및 신경회로망(neural network)을 사용하여 성능을 비교하였다. SVM은 두 개의 클래스를 분류하는 일반적인 방법으로서 입력 패턴을 고차원의 특징 공간으로 매핑할 수 있는 커널을 사용한다는 점과 대역적으로 최적의 식별을 할 수 있다는 특징이 있다.

3.3.1 신경회로망

우선 본 논문에서는 분류 기준들을 조합하는 방법으로 오류역전파 알고리즘(back-propagation algorithm)을 적용하였다.

3.3.2 Support Vector Machine

3.3.2.1 최적의 패턴 구분 초평면

$(x_i, y_i) \ 1 \leq i \leq N$ 를 식별자의 학습(training)을 위한 샘플 데이터라고 하면 d 차원의 각 샘플들은 $y_i \in \{-1, 1\}$ 의 레이블을 갖는 클래스에 속하게 된다. 따라서 SVM에서의 목

적은 동일한 레이블을 갖는 샘플들이 초평면(hyper-plane)의 같은 편에 위치하고 다른 레이블을 갖는 샘플들은 초평면의 반대편에 위치하도록 하는 것이다. 이것은 수식 (5)를 만족하는 w 및 b 를 찾는 것이다. 만약에 수식 (5)를 만족하는 초평면이 존재한다면 입력 샘플의 집합은 선형적으로 분리가능(linearly separable)하다고 할 수 있다.

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) > 0, \quad i=1, \dots, N \dots\dots\dots (5)$$

수식 (6)을 만족하도록 수식 (5)의 w 및 b 의 크기를 재조정하는 것이 가능하다.

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1, \quad i=1, \dots, N \dots\dots\dots (6)$$

이와 같은 초평면들 중에서 가장 가까이 위치한 샘플들과의 거리가 최대가 되는 초평면을 최적의 패턴 구분 초평면(optimal separating hyper-plane, OSH)이라고 한다. 따라서 OSH를 찾는 것은 수식 (2)의 조건하에서 $\|\mathbf{w}\|^2$ 를 최소화시키는 것과 동일하다. 이것은 초평면과 가장 근접한 샘플 사이의 거리가 $1/\|\mathbf{w}\|$ 이기 때문이다.

$\|\mathbf{w}\|^2$ 는 볼록 함수(convex)이므로 수식 (2)의 조건하에 이것을 최소화시키는 것은 라그랑주 상수 (Lagrange multipliers)를 사용함으로써 라그랑주 상수 $\alpha \geq 0$ 이고 $\sum_{i=1}^N y_i \alpha_i = 0$ 인 조건하에서 수식 (7)을 최대화시키는 것과 동일하다.

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j \dots\dots\dots (7)$$

수식 (7)을 최대화시키는 해가 구해졌다면 \mathbf{w} 와 b 는 각각 수식 (8) 및 수식 (9)가 된다. \mathbf{x}_A 및 \mathbf{x}_B 는 각각 클래스 A, B에 속하는 서포트 벡터(support vector)이다. 서포트 벡터는 수식 (6)의 등호가 성립하는 샘플들을 의미한다.

$$\mathbf{w} = \sum_{i=1}^N \alpha_i y_i \mathbf{x}_i \dots\dots\dots (8)$$

$$b = \frac{1}{2} (\mathbf{w}^T \mathbf{x}_A + \mathbf{w}^T \mathbf{x}_B) \dots\dots\dots (9)$$

3.3.2.2 비선형 SVM

입력 샘플의 집합이 선형적으로 분리가능하지 못한 경우 수식 (10)을 만족하는 변수 $\xi_i \geq 0$ 를 사용하여 수식 (6)을 만족하지 못하는 샘플들이 존재하는 것도 가능하게 한다.

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq 1 - \xi_i, \quad i=1, \dots, N \dots\dots\dots (10)$$

따라서 일반화된 OSH는 수식 (10)과 $\xi_i \geq 1$ 를 만족하며 수식 (11)을 최소화시키는 해로써 간주할 수 있다.

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi_i \dots\dots\dots (11)$$

입력 샘플들이 고차원의 특징 공간으로 사상하여 선형 SVM을 적용할 수 있다. 이때 특징 공간에 입력 샘플들을 직접 사상하는 과정을 생략하고 단지 사상된 샘플들 간의 유사도만을 계산하기 위해 커널함수를 도입함으로써 수식 (7)은 수식 (12)와 같이 변경된다.

$$W(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \dots\dots\dots (12)$$

일반적으로 활용되는 커널함수는 다음과 같다. 이러한 커널 사상함수는 특징공간을 무한 고차원 공간으로 사상할 수 있다.

$$K(\mathbf{x}, \mathbf{x}_j) = (\mathbf{x}^T \mathbf{x}_j + 1)^p \dots\dots\dots (13)$$

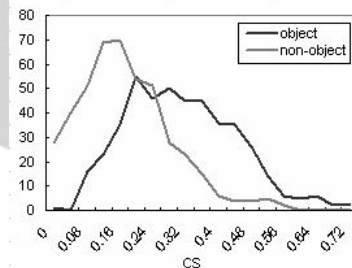
$$K(\mathbf{x}, \mathbf{x}_j) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x} - \mathbf{x}_j\|\right) \dots\dots\dots (14)$$

$$K(\mathbf{x}, \mathbf{x}_j) = \tanh(\beta_0 \mathbf{x}^T \mathbf{x}_j + \beta_1) \dots\dots\dots (15)$$

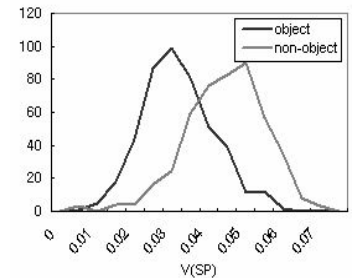
수식 (13) ~ 수식 (15)는 각각 polynomial learning machine, radial-basis function network 및 two-layer perceptron에 해당하는 커널 함수이다.

IV. 실험결과 및 토의

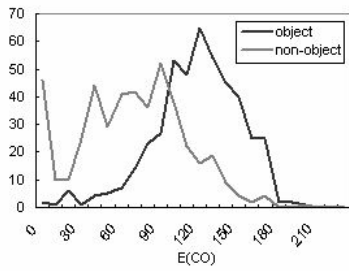
코렐 Gallery photo-CD에서 무작위로 선정된 객체영상 425개와 비객체영상 472개, 총 897개의 영상을 대상으로 영상 분류 실험을 수행하였다. 우선 각 분류 기준들의 구별 능력을 평가하였다. 그림 3(a)-(d)는 객체 및 비객체 영상들에 대한 중심 영역의 특이성(CS), 특이 픽셀의 최대 분산(V(SP)), 중심 객체의 평균 경계강도(E(CO)) 및 중심 영역의 평균 경계강도(ES)의 분포 곡선을 나타낸다. 그림 3에서 보듯이 객체 및 비객체 영상에 해당되는 각각의 곡선이 분명하게 분리되지 않고 있는 것을 알 수 있다. 따라서 각각의 기준들이 그 자체만으로는 좋은 성능을 제공하지는 못한다.



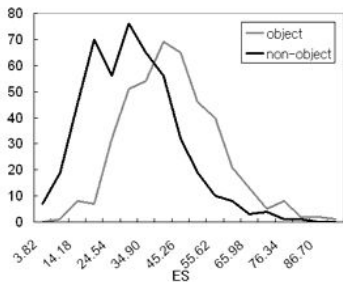
(a) 중심 영역의 특이성 분포



(b) 특이 픽셀의 최대 분산



(c) 중심 객체의 평균 경계강도



(d) 중심 영역의 평균 경계강도

그림 3. 분류 기준에 따른 측정치의 분포
Fig. 3. The distribution of values for input images when each measure is applied

SVM 및 신경회로망을 사용하여 분류 정확도를 계산하는 과정에서 주어진 데이터가 작아서 학습 데이터에 편향되는 경향을 방지하기 위하여 SVM은 3-폴드 크로스 밸리데이션을 사용하여 평가하였다. n-폴드 크로스 밸리데이션(n-fold cross-validation)[6]은 주어진 데이터를 n개의 동일한 크기의 폴드(fold)로 분할하고 이 중 하나의 폴드를 평가 데이터(test data), 나머지 n-1개의 폴드를 학습 데이터(training data)로 나누어 평가한다. 여기서 n개의 폴드로 분할하는 작업 또한 무작위로 n번 수행한 후 평균적인 성능을 측정한다. SVM을 수행하기 위한 C 값은 실험적으로 다양한 값에 대해 적용하여 최적의 것을 선택하였다. 또한 커널 함수는 polynomial function (수식 13)과 radial-basis function(RBF) (수식 14)을 사용하였으며 polynomial function의 사용을 위한 차수(p) 및 RBF의 사용을 위한 시그마(σ)는 역시 실험적으로 최적의 값을 선택하였다. 신경회로망의 경우에는 숨은 레이어의 개수를 역시 실험적으로 최적의 값을 선택하였다. 표 1은 중심 영역의 특이성(CS), 특이 픽셀의 최대 분산(V(SP)), 그리고 중심 객체의 평균 경계강도(E(CO))를 분류 기준으로 사용하고 SVM 및 신경회로망

을 식별자로 사용했을 경우의 분류 정확도를 나타낸다. 신경회로망의 경우 85.396%의 정확도를 나타냈으며 SVM은 85.842%의 분류 정확도를 나타내었다. 표 2는 중심 영역의 특이성(CS), 특이 픽셀의 최대 분산(V(SP)), 그리고 중심 영역의 평균 경계강도(ES)를 분류 기준으로 사용하고 SVM 및 신경회로망을 식별자로 사용했을 경우의 객체와 비객체 영상의 평가 결과를 보여준다. 신경회로망의 경우 81.271%의 정확도를 나타냈으며 SVM은 81.605%의 분류 정확도를 나타내었다.

표 1. CS, V(SP), E(CO)를 조합 적용한 객체 및 비객체 영상의 분류 평가 결과

Table 1. Evaluation of the classification result based on CS, V(SP), E(CO)

	신경회로망		SVM	
	객체	비객체	객체	비객체
Precision	0.883	0.832	0.878	0.843
Recall	0.798	0.905	0.814	0.898
F-measure	0.838	0.867	0.845	0.870

표 2. CS, V(SP), ES를 조합 적용한 객체 및 비객체 영상의 분류 평가 결과

Table 2. Evaluation of the classification result based on CS, V(SP), ES

	신경회로망		SVM	
	객체	비객체	객체	비객체
Precision	0.829	0.800	0.813	0.819
Recall	0.762	0.868	0.795	0.835
F-measure	0.794	0.828	0.804	0.827

표 3은 네 가지 분류 기준을 모두 조합하여 분류한 경우의 정확도이다. 네 가지 분류 기준을 모두 사용했을 때 보다 정확도가 향상되는 것을 볼 수 있다. 표 4 및 표 5는 각 분류 기준에 따른 분류 정확도를 나타낸다. 각 분류 기준별 정확도는 그렇게 높지 않지만 네 가지 분류 기준을 SVM을 사용하여 통합함으로써 86.622%의 정확도를 나타내는 것은 각각의 방법들이 가치가 있다는 것을 보여준다. 신경회로망을 사용하였을 경우 86.068%의 정확도를 나타내었다.

표 3. 네 가지 분류 기준을 모두 조합 적용한 객체 및 비객체 영상의 분류 평가 결과

Table 3. Evaluation of the classification result based on CS, V(SP), E(CO) and ES

	신경회로망		SVM	
	객체	비객체	객체	비객체
Precision	0.895	0.836	0.888	0.849
Recall	0.800	0.915	0.821	0.907
F-measure	0.845	0.874	0.853	0.877

중심 객체의 평균 경계강도를 계산하기 위해서는 영상분할, 핵심객체 영역 선택, 배경 분석 및 핵심객체 영역 확장 등의 많은 단계를 처리해야 된다. 영상 분할을 위한 평균 소요 시간이 7.53초이며 나머지 중심 객체 추출 단계의 소요 시간이 평균 5.41초로 계산되었다. 따라서 중심 객체의 평균 경계강도를 계산하기 위해서는 한 장의 영상에 최소 12초 이상의 처리 시간이 소요된다. 이에 비해 중심 영역의 평균 경계강도를 계산하기 위해서는 1초미만이 소요되므로 계산 시간에 많은 차이를 나타낸다. 따라서 중심 영역의 평균 경계강도가 중심 객체의 평균 경계강도에 비해 낮은 분류 성능을 나타내지만 빠른 처리 시간이 요구되는 응용 분야에서는 효과적으로 사용가능하다.

표 4. CS 및 V(SP)에 따른 객체 및 비객체 영상의 분류 평가 결과
Table 4. Evaluation of the classification result for CS and V(SP), respectively

		CS		V(SP)	
		신경 회로망	SVM	신경 회로망	SVM
객체	Precision	0.723	0.741	0.754	0.764
	Recall	0.671	0.614	0.751	0.739
	F-measure	0.696	0.672	0.752	0.751
비객체	Precision	0.722	0.689	0.766	0.772
	Recall	0.769	0.807	0.780	0.794
	F-measure	0.745	0.749	0.778	0.783

표 5. E(CO) 및 ES에 따른 객체 및 비객체 영상의 분류 평가 결과
Table 5. Evaluation of the classification result for E(CO) and ES, respectively

		E(CO)		ES	
		신경 회로망	SVM	신경 회로망	SVM
객체	Precision	0.750	0.753	0.667	0.679
	Recall	0.795	0.831	0.642	0.638
	F-measure	0.776	0.790	0.655	0.658
비객체	Precision	0.807	0.832	0.689	0.691
	Recall	0.771	0.754	0.712	0.729
	F-measure	0.789	0.791	0.700	0.709

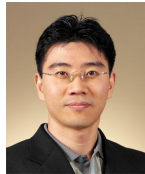
V. 결론

본 논문에서는 네 가지 분류 기준(중심 영역의 칼라 분포 특이성(CS), 특이 픽셀의 최대 분산(V(SP)), 중심 객체의 평균 경계강도(E(CO)), 중심 영역의 평균 경계강도(ES))에 의해 객체 및 비객체 영상을 분류하는 방법을 제안하였다. 영상 분류를 위해 사용된 기준 중에서 중심 객체의 평균 경계강도는 특징값의 계산에 많은 연산량과 계산 시간을 요구하므로 비슷한 특징을 가지며 분류 성능은 조금 떨어지지만 빠른 처리가 가능한 중심 영역의 평균 경계강도를 선택적으로 사용할 수 있도록 하였다. 네 가지 분류 기준에 대한 세 가지 경우의 조합(CS+V(SP)+E(CO), CS+V(SP)+ES, CS+V(SP)+E(CO)+ES)에 대해 신경회로망과 SVM을 적용하여 분류 성능을 비교한 결과 SVM이 신경회로망에 비해 조금 높은 성능을 나타내는 것을 확인하였다. CS+V(SP)+E(CO)의 조합이 CS+V(SP)+ES에 비해 조금 높은 성능을 나타내었지만 CS+V(SP)+ES 조합을 사용하면 중심 객체 추출을 위한 전처리 단계에서 빠르게 영상 분류가 가능한 것을 확인하였다. 네 가지 분류 기준을 SVM으로 분류한 결과 86.622%의 정확도를 얻었다. 제안된 방법은 영상검색과 영상 색인의 성능을 개선하는데 유용하게 사용될 수 있다.

참고문헌

- [1] A. Vailaya, A.K. Jain, and H.J. Zhang, "On Image Classification: City images vs. landscape," *Pattern Recognition*, pp.1921-1936, 1998
- [2] M. Szummer, and W. Picard, "Indoor-outdoor image classification," *IEEE Int'l Workshop Content-Based Access Image Video Databases*, pp. 42-51, 1998
- [3] A. Vailaya, M.A.T. Figueiredo, A.K. Jain, and H.J. Zhang, "Image Classification for Content-Based Indexing," *IEEE Trans. on Image Processing*, pp.117-130, 2001
- [4] A. Oliva, A. Torralba, A.G. Dugue, and J. Herault, "Global Semantic Classification of Scenes using Power Spectrum Templates," *CIR'99*, Newcastle UK, 1999
- [5] S.Y. Kim, S.Y. Park, and M.H. Kim, "Central Object Extraction for Object-Based Image Retrieval," *Int'l Conf. on Image and Video Retrieval (CIVR)*, pp.39-49, 2003
- [6] J.R. Serra, and J.B. Subirana, "Texture Frame Curves and Regions of Attention Using Adaptive Non-cartesian Networks," *Pattern Recognition*, pp.503-515, 1999
- [7] S.B. Park, J.W. Lee, and S.K. Kim, "Content-based image classification using a neural network. *Pattern, Recognition Letter*, pp.287-300, 2004
- [8] Y. Lu and H. Guo, "Background Removal in Image indexing and Retrieval," *International Conference on Image Analysis and Processing*, pp. 933-938, 1999
- [9] S. Haykin, "Neural Network - A Comprehensive Foundation," Prentice Hall, 1999
- [10] I.H. Witten, E. Frank, "Data Mining," Academic Press, 2000
- [11] O. Chapelle, P. Haffner, and V. Vapnik, "SVMs for histogram-based image classification," *IEEE Transactions on Neural Networks*, 10(5), 1999
- [12] S. Kim, S. Park, and M. Kim, "Image Classification into Object/Non-object Classes," *Int'l Conf. on Image and Video Retrieval (CIVR)*, pp. 393-400, 2004
- [13] Y. Deng, B.S. Manjunath, and H. Shin, "Color Image Segmentation," *IEEE Conference on CVPR*, pp. 446-451, 1999
- [14] 주경수, "MPEG-7 기반의 영상정보 검색시스템 설계 및 구현," *한국컴퓨터정보학회*, 10권 1호, 2005

저자 소개



김성영

2003년 부산대학교 컴퓨터공학과
공학박사

2004년 3월 ~ 현재 금오공과대학교
컴퓨터공학부 조교수

<관심 분야> 칼라 영상 분할, 내용
기반영상검색, 영상 분류