

논문 2006-43SP-5-2

# 특징 지도를 이용한 중요 객체 추출

(Extraction of Attentive Objects Using Feature Maps)

박기태\*, 김종혁\*\*, 문영식\*\*\*

(Ki Tae Park, Jong Hyeok Kim, and Young Shik Moon)

## 요약

본 논문에서는 컬러 영상에서 배경의 복잡도와 객체의 위치에 관계없이 영상 내에 존재하는 중요 객체를 자동으로 추출하는 방법을 제안한다. 제안하는 방법은 중요 객체를 추출하기 위해 에지(edge) 정보와 색상(color) 정보를 이용한 특징 지도를 사용한다. 또한, 효과적인 객체 추출을 위해서 참조 지도(reference map)를 제안한다. 참조 지도를 생성하기 위해서는 영상에서 사람의 시각에 두드러지게 구분되는 영역을 표현하는 특징 지도(feature map)를 먼저 생성한다. 그런 다음, 특징 지도들을 효과적으로 결합하여 배경의 영향을 최소화 하면서, 중요 객체가 존재할 확률이 높은 영역들을 포함하는 참조 지도를 생성한다. 특징 지도를 생성하기 위해서는 밝기 차 정보를 나타내는 에지와 YCbCr 컬러와 HSV 컬러 공간에서의 색상 성분을 사용하며, 특징 지도에 대한 생성 방법은 영상 내에서 밝기차이와 색상차이에 의해서 나타나는 경계 부분을 추출하는 방법을 사용한다. 최종적으로 중요 객체가 존재하는 영역을 나타내기 위해서 참조 지도와 특징 지도들을 결합한 결합 지도(combination map)를 생성한다. 결합 지도는 중요 객체의 외곽선 정보만을 표현하기 때문에, 객체 전체를 표현할 수 있는 객체 후보 영역을 추출하는데, 이를 위해서는 객체 후보 영역을 추출하기 위해서 convex hull 알고리즘을 사용한다. Convex hull 알고리즘에 의해서 추출된 영역은 여전히 배경 부분을 포함하고 있으므로, 영상 분할 방법을 적용하여 배경을 제거한 후 영상에서의 중요 객체를 추출한다. 제안한 알고리즘의 성능을 실험적으로 확인한 결과, 평균적으로 84.3%의 정확율과 81.3%의 재현율의 성능을 보였다.

## Abstract

In this paper, we propose a technique for extracting attentive objects in images using feature maps, regardless of the complexity of images and the position of objects. The proposed method uses feature maps with edge and color information in order to extract attentive objects. We also propose a reference map which is created by integrating feature maps. In order to create a reference map, feature maps which represent visually attentive regions in images are constructed. Three feature maps including edge map, CbCr map and H map are utilized. These maps contain the information about boundary regions by the difference of intensity or colors. Then the combination map which represents the meaningful boundary is created by integrating the reference map and feature maps. Since the combination map simply represents the boundary of objects, we extract the candidate object regions including meaningful boundaries from the combination map. In order to extract candidate object regions, we use the convex hull algorithm. By applying a segmentation algorithm to the area of candidate regions to separate object regions and background regions, real object regions are extracted from the candidate object regions. Experiment results show that the proposed method extracts the attentive regions and attentive objects efficiently, with 84.3% precision rate and 81.3% recall rate.

**Keywords :** attentive object, feature map, reference map, combination map, segmentation

\* 학생회원, \*\*\* 평생회원, 한양대학교 컴퓨터공학과  
(Department of Computer Science and Engineering,  
Hanyang University)

\*\* 정회원, (주)코디콤  
(KODICOM Co., LTD)

※ 본 연구는 과학재단 특정기초 연구 (R01-2006-000-10876-0) 지원으로 수행되었음

접수일자: 2006년6월7일, 수정완료일: 2006년8월14일

## I. 서론

최근 컴퓨터 및 통신 기술이 빠르게 발달하면서 사진이나 동영상 같은 멀티미디어 정보 이용에 대한 사용자 의 욕구가 증가하고 있다. 인터넷이나 네트워크 상에서 원하는 데이터를 보다 빠르고 효율적으로 검색하고자

하는 욕구의 증가는 멀티미디어 정보의 증가에 따른 당연한 결과라 할 수 있다. 이에 따라 내용 기반 영상 검색(Content-based Image Retrieval, CBIR) 분야가 활발히 연구되어 왔다<sup>[1-3]</sup>. 대표적인 내용 기반 영상 검색 시스템은 QBIC, Netra, VisualSEEK, Blobworld 등이 있다<sup>[4-7]</sup>.

내용 기반 영상 검색의 목적은 주어진 데이터베이스로부터 색상, 질감, 그리고 형태 등의 특징 값들을 이용하여 사용자가 원하는 영상을 효과적으로 검색할 수 있도록 하는 것이다. 색상, 질감, 그리고 형태 등의 특징들은 하위 레벨 특징(low-level features)에 속하는 것으로 대부분의 검색 시스템에서 사용되고 있다. 하지만 이러한 하위 레벨 특징들은 집, 자동차, 사람 등과 같은 의미를 포함하는 상위 레벨 특징(high-level features)과는 상당히 다르다. 따라서 이 둘을 이용한 검색 결과는 그 만족도에 있어서 큰 차이를 보일 수밖에 없다. 이러한 문제점을 해결하기 위해 몇몇 시스템들은 영상으로부터 의미 있는 영역이나 객체를 추출하려고 노력하고 있다.

현재까지 객체 영역들을 추출과 관련한 연구 방법에는 인간 시각 시스템의 특징에 기반을 두고 시각적인 중요도를 갖는 영역을 구별하는 방법<sup>[8-11]</sup>이 있고, 배경을 제거하거나<sup>[12]</sup>, 객체와 배경을 분리<sup>[13]</sup>하는 방법이 있다. 또한 인접 영역간의 병합 과정을 통해 영상을 두개의 영역으로 표현하거나<sup>[14]</sup>, 질감(texture) 경계를 설정하여 관심영역을 추출하는 방법<sup>[15]</sup> 등이 있다.

일반적으로 의미 있는 영역이 영상에 대한 유용한 정보를 제공하지만 관심 객체(object of interest)는 이보다 높은 수준의 정보를 제공할 수 있다. 하지만 입력 영상으로부터 관심 객체를 구분하는 것은 상당히 어려운 문제이다. 왜냐하면 영상에 대해서 어떤 것을 관심 객체로 결정할 것인가에 대해 명확한 정의를 내릴 수 없기 때문이다. 예를 들어, 나무에 앉아있는 원숭이가 있을 때, 원숭이만 객체로 판단해야 하는지, 아니면 나무를 포함해서 객체로 판단해야 하는지는 사용자에게 따라서 또는 응용분야에 따라서 달라질 것이다. 그리고 명확한 객체가 존재하더라도 완전하게 객체를 하나의 영역으로 추출하는 것은 매우 어렵다는 점이다. 일반적인 영상 분할 방법으로 객체를 포함하는 영상을 분할하면 객체 내부의 다양한 특징으로 인해 객체는 필연적으로 다수의 영역으로 분리되어 나타나게 된다.

따라서, 객체를 추출하는 과정은 영상과 객체 특징의 다양성을 효과적으로 반영해야 하므로 결코 쉬운 문제

가 아니다. 하지만 영상에 포함된 관심 객체를 추출할 수 있다면, 영상의 핵심 정보에 대한 효과적인 표현이 가능하게 되어 다양한 응용 분야에 유용하게 활용할 수 있다. 추출된 관심 객체는 객체 기반의 영상 검색, MPEG4 혹은 MPEG7의 비디오 객체 요소(video object components) 추출을 통한 비디오 객체 평면(video object plane, VOP)의 구성, 다양한 크기의 디스플레이 장치에 동일한 크기의 영상 전송 그리고 객체 및 배경에 상이한 압축률을 적용함으로써 보다 효과적인 영상 압축이 가능할 것이다.

## II. 제안하는 방법

본 논문은 영상에서 배경의 복잡도나 객체의 위치에 관계없이 중요 객체를 자동으로 추출하는 방법을 제안한다. 그림 1은 본 논문에서 제안한 알고리즘의 전반적인 순서도를 보여주고 있다. 제안된 알고리즘은 다음과 같이 5단계로 구성되어 있다. 첫 번째 단계에서는 입력 영상에 대해서 에지와 색상 정보를 이용하여 특징 지도(feature map)들을 생성한다. 두 번째 단계에서는 특징 지도들을 서로 비교하여 공통적으로 나타나는 부분을 추출함으로써 중심 객체가 존재할 수 있는 기준 영상인 참조 지도(reference map)를 생성한다. 세 번째 단계에서는 참조 지도를 기준으로 하여 특징 지도들로부터 제거된 객체의 정보들을 추가함으로써 중요 객체를 표현할 수 있는 결합 지도(combination map)를 구성한다. 네 번째 단계에서는 convex hull 알고리즘을 통해 다각형의 객체 후보 영역을 결정한다. 마지막 단계에서는 추출된 객체 후보 영역에 대해서 배경을 완전히 제거하기 위해서 객체 후보 영역에 대해 영상 분할을 적용하여 최종적으로 영상 내에서의 중요 객체를 추출한다.

### 1. 특징 지도(Feature Maps) 생성

특징 지도의 사용은 Itti에 의해 처음으로 제안되었다<sup>[16]</sup>. 하지만 Itti의 방법이 영상 내에서 다른 영역과 현저하게 차이가 나는 곳에 시선이 무의식적으로 집중되는 인간의 시각 특성을 반영하여 중요도가 높은 부분을 구해내는 것과 달리 제안하는 방법에서는 영상내의 밝기 차이와 색상 차이에 의해서 나타나는 경계 부분을 구해내는 것에 중점을 둔다. 왜냐하면 기존의 방법은 영상 내에 중요 객체보다 배경이 더 밝거나 복잡해질수록 객체를 제대로 추출해 주지 못하기 때문이다. 따라서, 본 논문에서는 특징 지도는 Itti의 방법으로 생성하

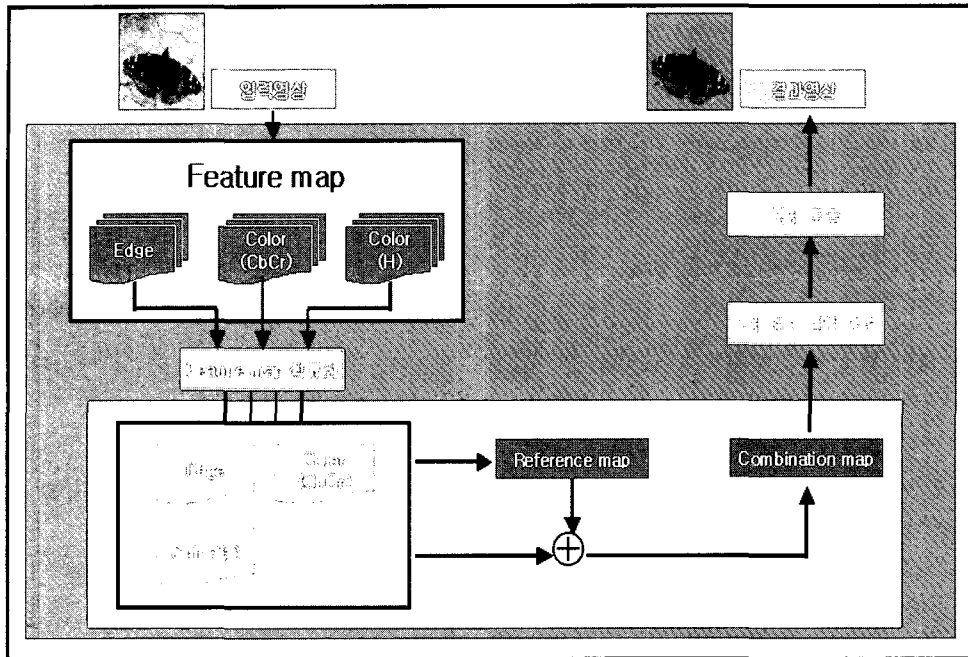


그림 1. 제안한 알고리즘의 순서도  
Fig. 1. Flowchart of the proposed algorithm.

지만, 밝기(Intensity), 색상(Color), 질감(Texture) 정보를 사용하지 않고 에지와 두 개의 색상 정보를 이용하여 특징 지도를 생성한다. 그림 2는 기존 방법과 제안하는 방법의 차이점을 보여 주고 있다. 그림 2에서 Itti의 방법에 의해 생성된 밝기(Intensity) 지도와 방향성(Orientation) 지도를 보면 객체와 배경을 구분하기 힘든 반면에, 제안한 방법에 의해 생성된 특징 지도들은 객체 정보를 공통적으로 잘 표현하고 있는 것을 확인할 수 있다.

가. 에지 지도(Edge Map)

본 논문에서는 밝기 차이 값을 갖는 에지 정보를 이용하여 에지 지도를 생성한다. 영상의 에지 정보를 표현하기 위해서 소벨(Sobel) 알고리즘을 사용한다. 에지 지도를 생성하기 위해서 원본 영상을 1/2, 1/3, 1/4로 축소한 후 각 축소 영상에 대해 소벨 알고리즘을 이용하여 에지영상을 생성한다. 그런 다음, 축소된 영상에 대하여 그림 3과 같이 필터를 적용시켜 필터의 중심 화소와 주변 화소간의 밝기 차이의 평균값을 구한다. 식 (1)은 그림 3에 대한 수식을 표현하고 있다.

$$F_s(x, y) = \frac{1}{24} \left( \sum_{i=1}^{24} |I_s(x, y) - N_i| \right) \quad (1)$$

식 (1)에서 S는 영상의 축소비율(1/2, 1/3, 1/4)을 나

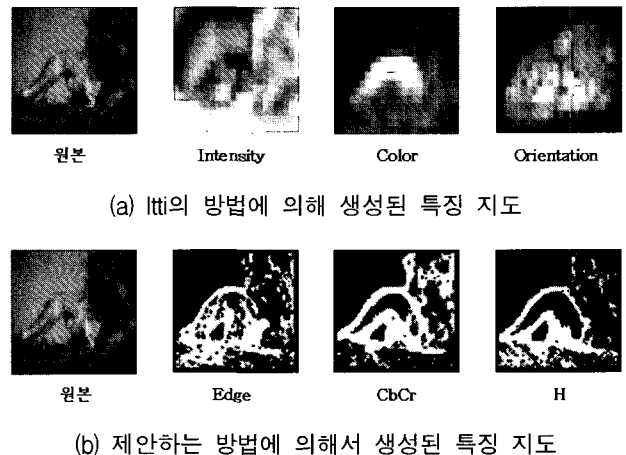


그림 2. 특징 지도 비교  
Fig. 2. Comparison of feature maps : (a) feature maps created by Itti's method; (b) feature maps created by the proposed method.

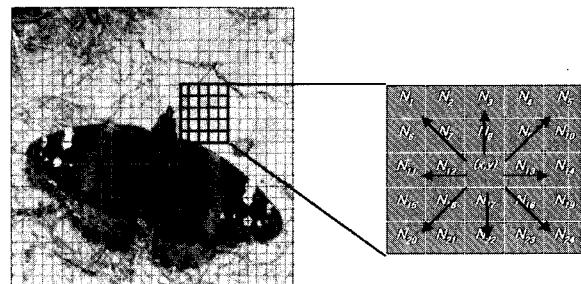


그림 3. 중심 화소와 주변 화소간의 밝기차 계산  
Fig. 3. Calculation of intensity difference between center pixel and neighborhood pixels.

내고,  $I_s$ 는 축소된 에지 영상,  $F_s$ 는 축소비율에 해당하는 특징 지도를 의미하고, 값은  $[0, 255]$ 의 값을 가진다. 본 논문에서는 에지 지도를 생성하기 위해서 원 영상 크기의 1/2배, 1/3배, 1/4배로 축소하여 사용하기 때문에 모두 3장의 서로 다른 크기를 가진 에지 지도를 생성한다. 그런 다음, 서로 다른 크기를 가진 세 장의 에지 지도를 결합하여 에지 지도를 대표할 수 있는 한 장의 새로운 대표 에지 지도를 생성한다. 대표 에지 지도를 생성하는 방법은 먼저 크기가 서로 다른 영상들이므로, 각 영상의 크기를 1/4배의 영상의 크기로 맞추는 정규화 과정을 수행한다. 그런 다음, 세 장의 에지 지도를 결합하여 평균 영상을 구함으로써 에지 지도를 생성한다. 대표 에지 지도를 생성하는 방법은 식 (2)와 같다.

$$RFM_{Edge} = \frac{1}{3} \sum_{s=2}^4 N \left( F_{\frac{1}{s}}^{Edge} \right) \quad (2)$$

식 (2)에서  $F_{\frac{1}{s}}^{Edge}$ 는 식(1)을 통해서 생성된 각각의 축소 영상에 대한 에지 지도를 나타내고,  $N(\cdot)$ 은 괄호안의 에지 지도의 크기를 원영상의 1/4 크기로 정규화한다는 것을 의미한다.  $RFM$ (Representative Feature Map)은 정규화 과정을 통해 생성된 하나의 대표 에지 지도를 의미한다. 그림 4는 원 영상으로부터 에지 지도를 생성하기 위한 과정을 나타내고 있다.

#### 나. 색상 지도(Color Map)

본 논문에서는 색상 지도를 생성하기 위해서 YCbCr 색상 모델에서 Cb와 Cr, 그리고 HSV 색상 모델에서 H 채널을 사용하여 색상 지도를 생성한다. 기존 방법에서는 빨간색(Red), 녹색(Green), 파란색(Blue)의 RGB 색상 모델에 노란색(Yellow)을 추가로 정의하여 색상지도를 생성하였다<sup>[15]</sup>. 하지만, 이 방법은 RGB 색상 모델이 각 채널간 상관관계가 크고, 서로 다른 색상간의 거리가 인간이 인지하는 거리와 대응되지 않는다는 단점이 있다.

##### (1) CbCr 색상 지도

본 논문에서는 YCbCr 색상 모델을 사용하여 색상 지도를 생성한다. YCbCr 색상 모델은 색상 정보로부터 광도를 분리하는 또 하나의 색상 모델이다. 여기서 밝기 성분인 Y 성분은 사용하지 않고 두 가지 색상 성분 Cb, Cr을 이용한다. Cb와 Cr에 대한 각각의 색상 지도는 에지 지도를 생성할 때와 같은 방법으로 생성한다.

Cb, Cr 채널에 대해서 각각 세 장씩의 지도들을 정규화시킨 후에 한 장씩의 평균영상을 생성한 후, Cb, Cr 각각의 평균영상을 다시 합쳐서 하나의 CbCr 색상 지도로 만든다. 이를 수식으로 정리하면 식 (3)과 같다.

$$RFM_{CbCr} = \frac{1}{2} \left( \frac{1}{3} \sum_{s=2}^4 N \left( F_{\frac{1}{s}}^{Cb} \right) + \frac{1}{3} \sum_{s=2}^4 N \left( F_{\frac{1}{s}}^{Cr} \right) \right) \quad (3)$$

##### (2) H 색상 지도

YCbCr 컬러 모델은 Cb, Cr 성분을 이용해서 색을 표현한다. 본 논문에서는 CbCr 맵의 다양한 색상을 표현해 주지 못한다는 단점을 보완하기 위해 인간의 색상 인식 작용과 유사하면서, 충분한 색상정보를 표현할 수 있는 HVS 색상 모델의 H(hue) 성분도 이용한다. 앞에서 설명한 에지 지도 및 Cb,Cr 색상 지도들과 동일한 과정으로 구성되며 식(4)의 방법으로 하나의 H 색상 지도를 생성한다.

$$RFM_H = \frac{1}{3} \sum_{s=2}^4 N \left( F_{\frac{1}{s}}^H \right) \quad (4)$$

그림 5는 다양한 영상들에 대해서 에지 지도, CbCr 색상 지도, H 색상 지도를 생성한 예제 들을 보여 주고 있다. 그리고 이 예제들을 통해서 Itti가 제안한 방법처럼 생성된 특징 지도들을 결합한 후 평균 영상을 이용하여 객체 정보를 추출한다면, 배경의 영향을 많이 받을 수 있다는 것을 알 수 있다. 그러므로 본 논문에서는 배경의 영향을 최대한 줄이면서 객체를 표현할 수 있는 참조 지도에 대한 생성 방법을 제안한다.

#### 2. 참조 지도(Reference Maps) 생성

입력영상의 배경이 복잡해질수록 객체를 정확하게 표현할 수 없기 때문에 배경의 영향을 줄이기 위해서 참조 지도(Reference map, RM)를 제안한다. 참조 지도는 배경의 영향을 최소화하면서 객체가 존재할 확률이 가장 큰 영역을 나타내는 영상이다. 참조 지도를 생성하는 방법은 에지 지도를 기준으로 두 개의 색상 지도를 각각 결합하여 공통적인 부분을 나타낸다. 에지 지도는 객체 영역 부분을 잘 표현해 주고, 색상 지도들과는 다른 밝기 차 정보를 가지고 있다. 그리고, 객체와 배경의 경계 부분을 명확하게 표현해 주기 때문에 에지 지도를 기준으로 정한다. 참조 지도를 생성하는 방법은 수식 (5)와 같다.

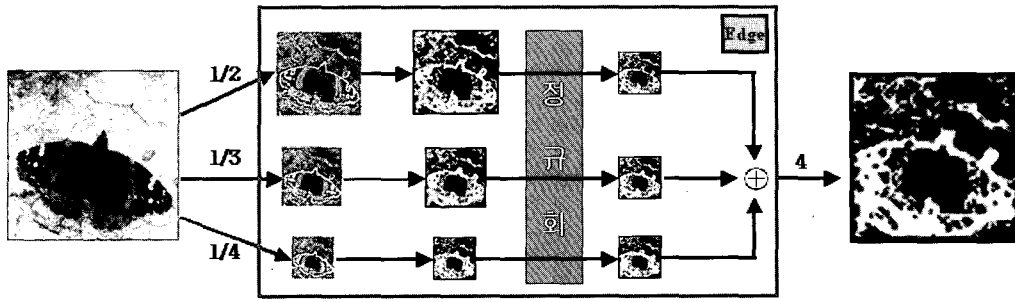


그림 4. 에지 지도 생성 과정  
Fig. 4. Procedure for creating edge map.

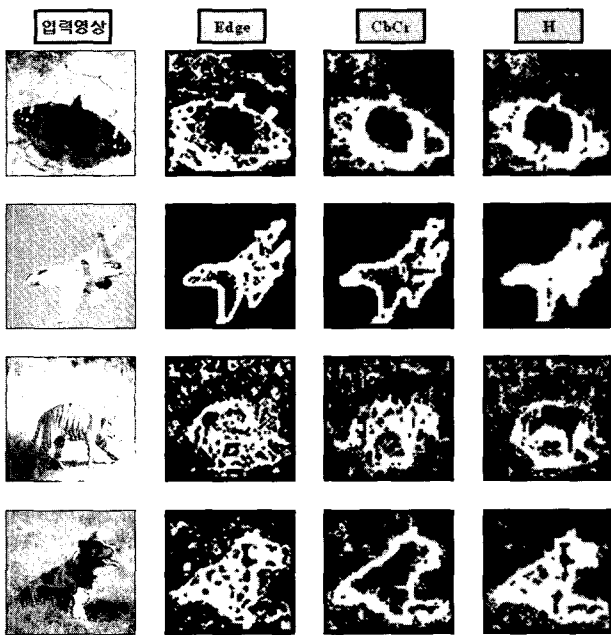


그림 5. 다양한 영상에 대한 특징 지도  
Fig. 5. Feature maps created from various images.

$$CRFM_{E\_CbCr} = \frac{1}{2} (RFM_{Edge} \oplus RFM_{CbCr})$$

$$CRFM_{E\_H} = \frac{1}{2} (RFM_{Edge} \oplus RFM_H)$$

$$RM = \frac{1}{2} (CRFM_{E\_CbCr} \oplus CRFM_{E\_H}) \quad (5)$$

식 (5)에서  $\oplus$  는 두 특징 지도들의 공통적인 부분들의 밝기 값을 더하여 결합시켜주는 연산자이고,  $CRFM$  (Combined Representative Feature Map)은 두 특징 지도들의 공통적인 부분들을 더해서 만들어진 중간 단계의 결합 지도이다. 두 개의  $CRFM$  을 결합해서 최종적으로 참조 지도를 생성한다. 그림 6은 참조 지도를 생성하기 위한 과정을 나타내고 있다. 그림 6에서 보듯이 참조 지

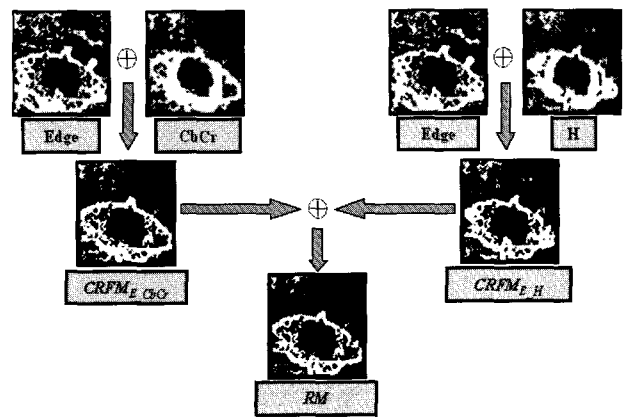


그림 6. 참조 지도 생성과정  
Fig. 6. Procedure for creating reference map.

도의 결과 영상을 보면 중요 객체 영역을 표현하면서 배경 영역은 상대적으로 제거되었음을 확인할 수 있다.

### 3. 결합 지도(Combination Maps) 생성

결합 지도는 참조 지도 영상을 기준으로 각 특징 지도들과 비교하여 참조 지도를 생성할 때 제거된 중요 객체영역을 복원한 영상이다. 참조 지도는 수식 (6)과 수식 (7)과 같이 생성할 수 있으며, 참조 지도와 각각의 특징 지도들을 비교해서 공통적으로 존재하는 부분의 값은 그대로 유지해주고, 공통적으로 존재하지 않은 부분의 값은 상대적으로 낮추어 줌으로써 배경 영역을 효과적으로 제거하면서 참조 지도에 나타나지 않았던 중요 객체영역을 표현할 수 있다.

$$RFM(x, y) = \begin{cases} RFM(x, y) \\ \text{if } RFM(x, y) \times RM(x, y) > 0 \\ w \times RFM(x, y) \text{ otherwise} \end{cases} \quad (6)$$

수식 (6)에서  $RFM$ 은 참조 지도에 기반한 수정된 특징 지도를 나타내고,  $w$ 는  $[0, 1]$  사이의 값을 갖는 가중치를

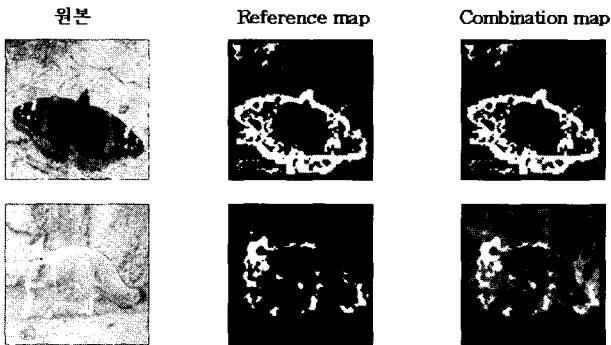


그림 7. 예제 영상에 대한 참조 지도와 결합 지도  
Fig. 7. Reference map and combination map created from example images.

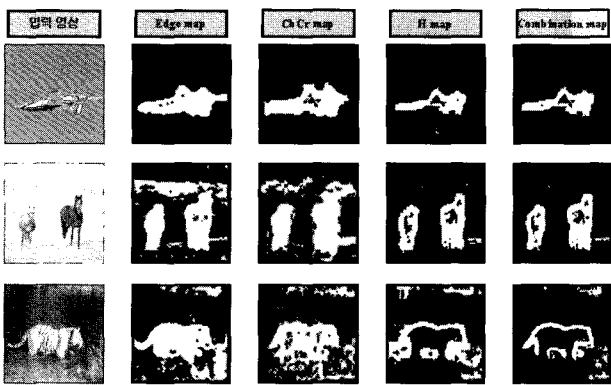


그림 8. 다양한 영상에 대한 특징 지도와 결합 지도  
Fig. 8. Feature maps and combination map created from various images.

의미한다. 에지 지도 및 두 종류의 색상 지도들에 대해서 각각 수정된 특징 지도들을 생성한 후, 수식 (7)을 이용하여 최종적으로 결합 지도를 생성한다. 그림 7은 원본 영상에 대한 참조 지도와 결합 지도를 보여주고 있다.

$$CM = \frac{1}{3}(RFM_{Edge} + RFM_{CbCr} + RFM_H) \quad (7)$$

그림 7에서 보이는 바와 같이, 배경과 객체의 색이 뚜렷하게 구분될 경우 참조 지도와 결합 지도의 차이가 거의 없지만 배경과 객체의 색이 비슷할 경우 참조 지도에서 나타나지 않던 객체 부분이 결합 지도에서 나타나는 것을 확인 할 수 있다. 그림 7은 다양한 영상에 대한 특징 지도들과 결합 지도들을 보여주고 있다.

#### 4. 객체 후보 영역 추출 및 객체 추출

최종적으로 결합 지도가 완성이 되면 영상 내의 중요 객체를 추출한다. 그러나 결합 지도는 중요 객체 전체를 표현하고 있는 것이 아니고, 객체의 경계선 영역을 표현하고 있는 것을 알 수 있다. 그러므로 본 논문에서는 중요

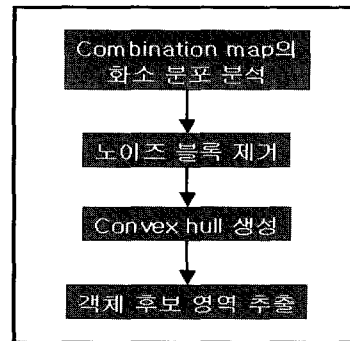


그림 9. 객체 후보 영역 추출과정  
Fig. 9. Extraction of candidate object regions.

객체를 추출하기 위해서 먼저 결합 지도에 나타난 객체의 경계선을 포함하는 객체 후보 영역을 추출한다. 그림 9는 결합 지도로부터 객체 후보 영역을 추출하는 과정을 나타내고 있다.

#### 가. 화소 분포 분석 및 노이즈 제거

본 논문에서는 효과적인 중요 객체 영역을 추출하기 위해 참조 지도를 생성한 후 중요 객체 영역이 아닌 영역들을 제거한다. 이를 위해, 먼저 참조 지도를 8 x 8 크기의 서브 블록으로 나뉜다. 그리고 서브 블록 내의 화소수를 조사하여 객체에 해당하는 화소 수가 65% 이상이면 해당 블록을 객체 영역 블록으로 정의하고, 그렇지 않은 경우에는 배경 영역으로 정의한다. 이러한 조건에 의해서 참조 지도를 서브 블록 단위로 객체 영역과 배경 영역으로 분할할 수 있다. 그런 다음, 각 블록에 대해 레이블링 알고리즘을 적용하여 최종적으로 노이즈 블록을 제거한다. 본 논문에서는 많은 실험을 통해서 블록들의 위치에 상관없이 블록 수가 6개 이하인 레이블을 갖는 블록들을 노이즈 블록으로 간주하여 객체 영역에서 배경으로 분리한다. 그림 10은 서브 블록들에 대해서 레이블링을 통한 노이즈를 제거한 결과를 보여주고 있다.



그림 10. 서브 블록 영상 및 노이즈 제거 영상  
Fig. 10. Image divided into sub blocks and image eliminating noises.

#### 나. 객체 후보 영역 추출

노이즈 블록이 제거된 영상에서 중요 객체를 포함하는

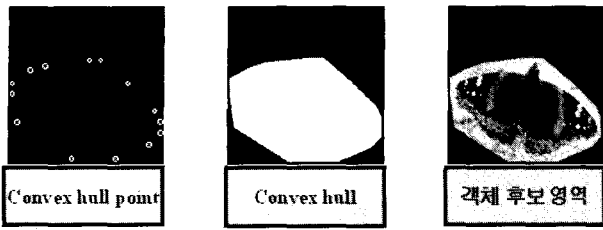


그림 11. Convex hull 알고리즘을 이용한 객체 후보 영역 추출

Fig. 11. Candidate object region extracted by convex hull algorithm.



그림 12. 다양한 영상에 대한 객체 후보 영역 추출

Fig. 12. Extraction of candidate object regions from various images.

다각형 영역을 구한다. 본 논문에서는 Andrew가 제안한 monotone chain 알고리즘<sup>[17]</sup>을 사용하여 convex hull로 구성된 영역을 객체 후보 영역으로 나타낸다. 그림 11은 객체 후보 영역을 검출하기 위한 과정을 보여주고 있다. 그림 12는 다양한 영상에 대한 객체 후보 영역 추출 결과를 보여주고 있다.

다. 중요 객체 추출

그림 12에서 객체 후보 영역들을 보면 여전히 객체 영역 뿐만 아니라 배경 영역을 포함하고 있는 것을 확인할 수 있다. 따라서 본 논문에서는 영상 분할 알고리즘을 이용하여 영상 내의 중요 객체 영역만을 추출하기 위한 방법을 제안한다. 영상 분할 알고리즘은 속도와 성능이 좋은 것으로 알려진 Wang의 영상 분할 기법을 사용한다<sup>[18]</sup>. 중요 객체 추출 방법은 객체 후보 영역에서 convex hull 경계선에 인접한 영역들을 배경 영역으로 가정하고 객체 후보 영역에서 제거한다. 이러한 방법은 객체 후보 영역이 객체를 포함하는 최소의 영역이므로 적용이 가능하다. 그림 13은 실제 영상에 대한 객체 추출 결과를 보여주고 있다.



그림 13. 중요 객체 추출 결과

Fig. 13. Extraction of attentive object.

II. 제안하는 방법

본 논문에서는 영상 검색 또는 객체 추출과 관련한 연구에서 주로 사용되는 Corel DB 영상을 사용하였다. 10개의 클래스를 사용하였고, 사용된 영상의 수는 160장이다. 그림 14는 본 논문에서 실험에 사용된 영상들 중에서 대표적인 영상들을 보여주고 있다.

1. 객체 후보 영역 추출 실험 결과

제안된 방법의 객체 후보 영역 추출 결과를 기존 방법들과 비교하였다. 기존방법 1은 Itti가 제안한 방법<sup>[12]</sup>에 의해서 구해진 객체 후보 영역이고 기존방법 2는 Hu가 제

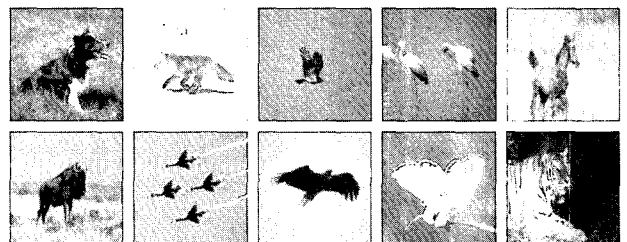


그림 14. 실험에서 사용된 클래스의 대표 영상

Fig. 14. Representative images used in experiments.

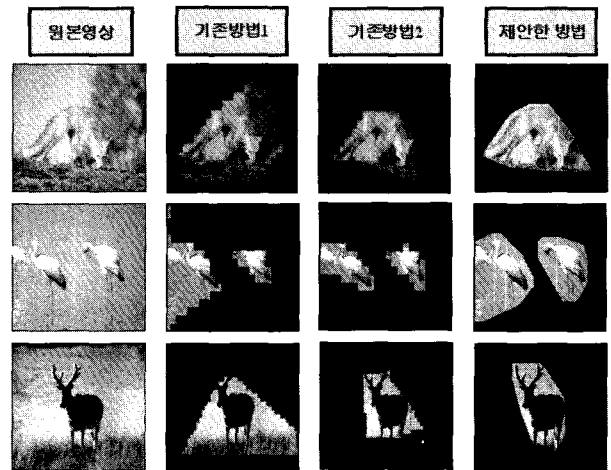


그림 15. 객체 후보 영역 추출 결과 비교

Fig. 15. Comparison of extracting candidate object regions.

안한 방법<sup>[19]</sup>에 의해서 구해진 객체 후보 영역이다. 결과를 정량적으로 비교할 수 있는 척도가 제안되지 않아서 객관적인 비교를 하기에는 어려움이 있지만, 그림 15에서 볼 수 있듯이 제안한 방법에 의한 객체 후보 영역에 대한 추출 결과가 전반적으로 우수함을 확인할 수 있다. 이에 반해, 기존 방법들은 객체와 배경이 인접한 영역부분에서는 객체의 일부분을 제대로 검출해 내지 못하는 것을 볼 수 있다. 이와 같은 결과는 객체의 에지 정보를 고려하지 않음으로써 나타날 수 있다.

2. 객체 추출 실험 결과

중요 객체 추출에 대한 결과를 검증하기 위해서 그래픽 툴을 이용하여 영상 내의 객체 영역을 정확하게 추출하여 기준 영상(ground truth)을 생성하고, 제안한 알고리즘을 이용하여 비교 영상을 생성하였다. 성능평가를 위한 척도는 정확율과 재현율을 사용하였다. 정확율(precision)을 측정하기 위해 기준 영상과 비교해서 제안한 방법에 의해 추출된 객체 영역이 얼마나 정확한지를 측정하는 것과 재현율(recall)을 계산하기 위해 제안한 방법에 의해 추출된 객체 영역이 얼마나 실제 객체 영역을 포함하고 있는가에 대한 비율을 계산하여 평가하는 방법이다. 식(8)와 식(9)은 두 가지 평가 방법을 나타내고 있다.

$$precision = \frac{N(S_M \cap S_A)}{N(S_M)} \times 100 \quad (8)$$

$$recall = \frac{N(S_M \cap S_A)}{N(S_A)} \times 100 \quad (9)$$

위 식에서,  $S_M$ 은 실제 객체 영역에 존재하는 화소들의 위치 값들을 나타내는 집합이고,  $S_A$ 는 제안한 방법에 의해서 추출된 객체 영역에 존재하는 화소들의 위치 값들을 나타내는 집합이다. 그리고  $N(S_M \cap S_A)$ 는 실제 객체 영역과 제안한 방법에 의해 추출된 객체 영역이 일치하는 화소의 수를 의미하고,  $N(S_M)$ 은 실제 객체 영역에 존재하는 객체의 화소 수를 나타내고,  $N(S_A)$ 는 알고리즘에 의해서 추출된 객체 영역의 픽셀 수를 나타낸다. 그림 16은 원본 영상에서 중요 객체를 추출한 결과를 보여주고 있다. 그리고 표 1은 실험에서 사용한 10개의 클래스 영상들에 대한 중요 객체 추출 결과를 보여주고 있다. 실험 결과 평균적으로 84.3%의 정확율과 81.3%의 재현율의 성능을 보였다.

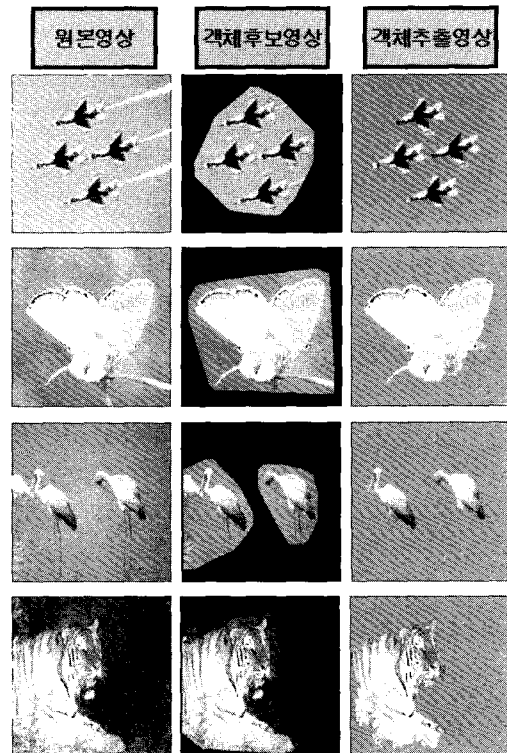


그림 16. 객체 추출 결과 영상  
Fig. 16. Result images of object extraction.

표 1. 객체 추출 실험 결과  
Table 1. Experimental results of object extraction.

단위 : %

클래스	정확율	재현율
개	73.7	87.9
여우	80.1	80.8
매	87.2	81.2
말	88.9	67.8
영양	80.3	76.2
비행기	86.2	96.1
독수리	87.5	87.1
나비	91.1	86.3
호랑이	80	74.1
물새	87.6	75.7
평균	84.4	81.3

그림 17과 그림 18은 정확율과 재현율에 대한 결과를 보여주고 있다.

실험 결과를 통해 제안한 방법은 배경이 단순하거나 복잡한 영상일지라도 효과적으로 중요 객체를 추출할 수 있었고, 또한 객체의 위치에 상관없이 영상 내에 하나 이상의 중요 객체를 포함하고 있더라도 자동적으로 모든 객체를 추출할 수 있었다. 하지만 다음의 두 가지 점에 대해서 유의할 필요가 있다. 첫 번째는 영상에서 실제 객체 영역(ground truth)을 추출하는 것이 주관적



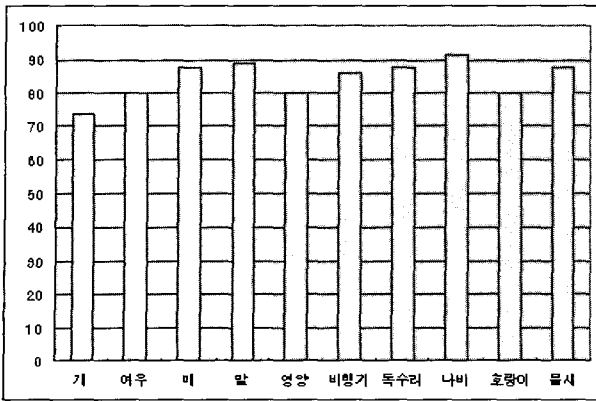


그림 17. 객체 추출에 대한 정확율  
Fig. 17. Precision of object extraction.

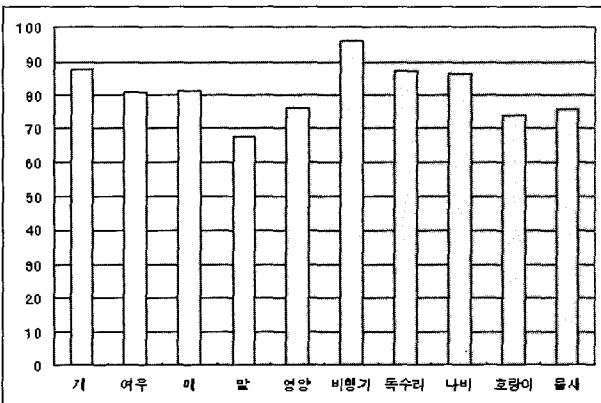


그림 18. 객체 추출에 대한 재현율  
Fig. 18. Recall of object extraction.

으로 이루어지기 때문에 어떻게 추출하느냐에 따라서 상당히 상반된 결과를 얻을 수 있다는 것이다. 즉, 기준이 되는 실제 객체 영역을 얼마나 정확하게 추출할 수 있느냐가 성능에 큰 영향을 준다는 것을 의미한다. 두 번째는 영상 분할 알고리즘의 성능에 따라 실험결과가 달라질 수 있다는 것이다.

#### IV. 결론 및 향후 과제

본 논문에서는 영상에 대한 사전 정보 없이 영상의 배경의 복잡도나 객체의 위치, 그리고 객체의 수에 관계없이 영상 내에 존재하는 중요 객체를 자동적으로 추출하는 방법을 제안하였다. 본 논문에서는 에지 지도와 두 종류의 색상 지도를 이용해 중요 객체를 추출하는 방법을 제안 하였으며, 배경의 영향을 줄이기 위해서 참조 지도와 결합 지도를 생성하는 방법을 제안하였다. 실험 결과 정확율이 84.3%와 재현율이 81.3%로서 영상으로부터 중요 객체를 효과적으로 추출할 수 있음을 확인할 수 있었다. 하지만 객체 후보 영역 추출 결과를 객

관적으로 평가할만한 척도가 아직까지 제시된 바가 없어 객관적인 평가를 하지 못하였다. 그러므로 향후 실험 결과에 대한 객관적인 평가 기준을 마련해야 할 필요가 있으며, 성능 개선을 위해 중요 객체를 추출하는 과정에서 적용 가능한 최적의 영상 분할 알고리즘 개발에 대한 연구가 진행되어야 할 것이다.

#### 참고 문헌

- [1] Y. Rui and T.S.Huang, "Image Retrieval : Current Techniques, Promising Directions, and Open Issues," *Jouranal of Visual Communication and Image Representation*, vol. 10, pp.39-62, 1999
- [2] 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 Trans. Pattern Analysis and Machine Intelligence*, vol.22, pp.1349-1380, Dec. 2000.
- [3] S. Michael, "Next Generation Web Searches for Visual Content," *IEEE Computer*, pp.46-52, Nov. 2000.
- [4] M. Flicker, H. Sawhneyy, W. Niblack, J. Ashley and P. Yanker "Query by Image and Video Content: The QBIC System," *IEEE Computer Special Issue on Content Based Picture Retrieval System*, Vol. 28, pp.23-32, 1995.
- [5] W.Y. Ma and B.S. Manjunath "Netra: A Tool-box for Navigating Large Image Database," *IEEE Conference on Image Processing*, Vol. 1, pp.568-571, 1997.
- [6] J. Smith and S. Chang "VisualSEEK: A Fully Automated Content-Based Image Query System," *ACM Multimedia*, pp.87-98, 1996.
- [7] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein and J. Malik, "Blobworld: A System for Region-based Image Indexing and Retrieval," *International Conference of Visual Information System*, Vol. 3, pp.509-516, 1999.
- [8] W. Osberger and A.J. Maeder, "Automatic Identification of Perceptually Important Regions in An Image," *IEEE International Conference on Pattern Recognition*, pp.701-704, 1998.
- [9] J. Luo and A. Singhal, "On Measuring Low Level Saliency in Photographic Images," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp.84-89, 2000.
- [10] C. M. Privitera and L.W Stark, "Algorithms for Defining Visual Regions of Interest: Comparison with Eye Fixations," *IEEE Transaction on*

*Pattern Analysis and Machine Intelligence*, vol. 22, pp.970-982, Sep. 2000.

[11] J. Senders, "Distribution of Attention in Static and Dynamic Scenes," *Proceedings of SPIE Human Vision and Electronic Imaging II*, vol. 3016, pp.186-194, Feb. 1997.

[12] Y. Lu and H. Guo, "Background Removal in Image Indexing and Retrieval," *Proceedings of 10th International Conference on Image Analysis and Processing*, pp.933-938, Sep. 1999.

[13] Q. Huang, B. Dom, D. Steels, J. Ashely, and W. Niblack, "Foreground/ background segmentation of color images by integration of multiple cues," *International Conference on Image Processing*, vol. 1, pp.246-249, 1995.

[14] T. Tamaki, T. Yamamura, and N. Ohnishi, "Image Segmentation and Object Extraction Based on Geometric Features of Regions," *SPIE Conference on Visual Communications and Image Processing*, vol. 3653, pp.937-945, Jan. 1999.

[15] J. R. Serra and J. B. Subirana, "Texture Frame Curves and Regions of Attention Using Adaptive Non-cartesian Networks," *Pattern Recognition*, vol. 32, pp.503-515, Mar. 1999.

[16] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-based Visual Attention for Rapid Scene Analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.20, No. 11, pp.1254- 1259, 1998.

[17] A. M. Andrew, "Another Efficient Algorithm for Convex Hulls in Two Dimensions," *Information Processing Letters*, pp. 216-219, 1979.

[18] D. Wang, "Unsupervised Video Segmentation Based on Watersheds and Temporal Tracking," *IEEE Transaction on Circuits and System for Video Technology*, Vol. 2, pp.539-546, 1998.

[19] Y. Hu, X. Xie, W. Y Ma, L. T. Chia, and D. Rajan, "Salient Region Detection Using Weighted Feature Maps Based on The Human Visual Attention Model," *Proceedings of the Fifth IEEE Pacific-Rim Conference on Multimedia*, Vol. 2, pp.993-1000, 2004.

저 자 소 개



박 기 태(학생회원)  
 2000년 한양대학교 전자계산학과  
 학사 졸업.  
 2002년 한양대학교 컴퓨터공학과  
 석사 졸업.  
 2002년~현재 한양대학교 컴퓨터  
 공학과 박사 졸업.

<주관심분야 : 내용기반 영상검색, 객체 검출, 비  
 디오 색인 및 검색, 영상처리>



김 종 혁(정회원)  
 2004년 한양대학교 전자컴퓨터  
 공학부 학사 졸업.  
 2006년 한양대학교 컴퓨터공학과  
 석사 졸업.  
 2006년~현재 (주) 코디콤 연구원

<주관심분야 : 객체 검출, 내용기반 영상 검색>



문 영 식(평생회원)- 교신저자  
 1980년 서울대학교 전자공학과  
 학사 졸업.  
 1982년 한국과학기술원 전기 및  
 전자공학과 석사 졸업.  
 1990년 University of California  
 at Irvine / Dept. of  
 Electrical and Computer  
 Engineering 박사 졸업.

1982년~1985년 한국통신연구소 연구원  
 1990년~1992년 생상기술연구원 선임 연구원  
 1992년~현재 한양대학교 컴퓨터공학과 교수  
 <주관심분야 : 내용기반 영상검색, 비디오 색인  
 및 검색, 영상처리, 패턴인식>