

# 대비 지도와 움직임 정보를 이용한 동영상으로부터 중요 객체 추출

## (Salient Object Extraction from Video Sequences using Contrast Map and Motion Information)

곽수영<sup>†</sup>      고병철<sup>\*\*</sup>      변혜란<sup>\*\*\*</sup>  
(Soo Yeong Kwak)    (Byoung Chul Ko)    (Hye Ran Byun)

**요약** 본 논문에서는 시공간 정보를 이용하여 동영상에서 움직이는 객체를 자동으로 추출하는 방법을 제안한다. 본 논문에서 제안하는 방법은 다른 영역과 구별되는 현저한 장소에 무의식적으로 집중되는 시각주의 특성을 컴퓨터 시스템에 도입한 대비 지도(contrast map)와 중요 특징점(salient point)을 적용한 것이 큰 특징이라고 할 수 있다. 대비 지도는 밝기(luminance), 색상(color) 그리고 방향성(direction) 3가지의 특징 정보 중 밝기와 방향성의 특징을 나타내는 밝기 지도(luminance map)와 방향성 지도(directional map)를 결합하여 대비 지도를 생성한다. 또한, 사람이 시각적으로 볼 때 의미 있다고 생각하는 중요 특징점을 웨이블릿 변환을 이용하여 찾아낸다. 이렇게 생성된 대비 지도와 중요 특징점을 이용하여 대략적인 집중윈도우(AW:Attention Window)의 위치와 크기를 결정한다. 다음으로, 동영상의 가장 큰 특징인 움직임 정보를 추정하여 집중윈도우를 객체에 가장 근사하게 축소시키고, 윤곽선 정보를 이용하여 객체를 추출한다. 윤곽선을 추출하기 위해 캐니에지(canny edge)를 사용하였으며, 배경의 윤곽선 제거를 위하여 윤곽선의 차이(DE:Difference of Edge)를 이용하여 가로 후보영역과 세로 후보영역을 추출한다. 추출된 2개의 후보영역을 AND연산과 모폴로지 연산을 이용하여 객체를 자동으로 추출하는 방법을 제안한다. 실험은 카메라가 고정된 상태에서 촬영한 동영상에 대해 이루어 졌으며, 객체와 배경이 효과적으로 분리되는 것을 확인하였다.

**키워드** : 대비 지도, 밝기 지도, 색상 지도, 방향성 지도, 중요 특징점, 집중윈도우, 윤곽선의 차이, 객체 추출

**Abstract** This paper proposes a moving object extraction method using the contrast map and salient points. In order to make the contrast map, we generate three-feature maps such as luminance map, color map and directional map and extract salient points from an image. By using these features, we can decide the Attention Window(AW) location easily. The purpose of the AW is to remove the useless regions in the image such as background as well as to reduce the amount of image processing. To create the exact location and flexible size of the AW, we use motion feature instead of pre-assumptions or heuristic parameters. After determining of the AW, we find the difference of edge to inner area from the AW. Then, we can extract horizontal candidate region and vertical candidate region. After finding both horizontal and vertical candidates, intersection regions through logical AND operation are further processed by morphological operations. The proposed algorithm has been applied to many video sequences which have static background like surveillance type of video sequences. The moving object was quite well segmented with accurate boundaries.

**Key words** : contrast map, luminance map, color map, directional map, salient points, Attention Window, difference of edge, object extraction

· 본 연구는 한국과학기술원 특정기초연구(R01-2005-000-10898-0)에서 지원받았습니다.

† 학생회원 : 연세대학교 컴퓨터과학과  
ksy2177@cs.yonsei.ac.kr

\*\* 정회원 : 계명대학교 정보통신학부 교수  
niceko@kmu.ac.kr

\*\*\* 종신회원 : 연세대학교 컴퓨터과학과 교수  
hrbyun@cs.yonsei.ac.kr

논문접수 : 2005년 2월 18일

심사완료 : 2005년 9월 22일

### 1. 서론

최근 컴퓨터 및 통신기술이 발달하면서 통신망 상에서 단순히 문자 정보만을 제공했던 것과는 달리 비디오, 오디오와 같은 멀티미디어 정보 제공에 대한 사용자의 욕구가 증가하고 있다. 많은 멀티미디어 데이터가 증가

함에 따라 보다 효율적으로 데이터를 관리하는 방법과 사용자가 원하는 데이터를 보다 쉽고 빠르게 사용할 수 있도록 멀티미디어 데이터를 효율적으로 검색할 수 있는 방법도 필요하게 되었다. 이에 국제 표준화 기구인 ISO(International Standards Organization)와 IEC(International Elector-technical Commission)는 멀티미디어 데이터에 대한 효율적인 검색을 제공하고자 국제 표준을 마련하였다. 이것이 바로 MPEG(Moving Picture Experts Group)이며, 그 중에서도 멀티미디어의 내용 기반 검색에 관한 표준을 마련하고자 구성된 것이 바로 MPEG-7(ISO/IEC JTC1/SC29/WG11)이다.

MPEG-7은 내용기반 멀티미디어 검색 문제에 관한 해결법을 제공하였다. 이러한 MPEG-7에는 두 가지 방향의 접근 및 해법이 제시되어 있다. 그 하나는 상위레벨 내용에 기반한 검색과 다른 하나는 하위 레벨 내용에 기반한 검색 방법이다. 하위레벨의 내용에는 키워드, 모양, 색, 크기, 위치, 방향성, 멜로디 등이 포함될 수 있고, 상위레벨의 내용에는 자동차, 집, 비행기와 같이 의미있는 정보들이 포함될 수 있다. 하위레벨의 내용에 의한 검색은 사용자 질의가 부자연스럽다는 단점이 있으나 상위레벨의 내용에 의한 검색은 좀 더 자연스러운 질의를 할 수 있다는 장점이 있다. 즉, 실제 사용자들은 색상, 질감, 모양 등과 같은 하위레벨의 특징보다는 자동차, 배, 사람, 독수리 등과 같은 상위레벨의 개념을 가지고 접근하기 때문에 사람과 검색 시스템 사이에는 검색결과에 대한 만족도의 차이가 발생 할 수밖에 없다. 이러한 문제점을 해결하기 위해 몇몇 시스템들은 동영상으로부터 의미있는 객체를 추출하려고 노력하고 있다. 하지만 복잡한 배경이 존재하는 자연동영상, 객체가 무수히 많이 존재하는 동영상, 특별히 눈에 띄는 객체가 존재하지 않는 영상 등과 같은 경우에는 의미 있는 객체를 사람의 수작업없이 자동으로 완벽하게 추출한다는 것은 지금까지의 기술을 비추어 보면 미흡한 점이 많이 존재한다. 지금까지의 몇몇 시스템들은 사람의 수작업을 거쳐서 객체를 추출한다거나, 객체가 중심에 존재한다는 가정을 두어 객체를 추출하는 시스템들이 많이 개발되었다[1,2]. 객체를 추출하는 시스템들은 멀티미디어 검색 이외에 MPEG-4, 워터마킹, 디지털 카메라 등과 같은 여러 가지 다른 응용분야에도 적용할 수 있다.

본 논문에서는 다양한 응용분야에 유용하게 적용될 수 있는 객체를 추출하는 새로운 방법을 제안한다. 그림 1은 제안하는 방법의 전체적인 흐름도이다. 먼저, 입력된 영상에서 밝기, 방향성 2가지의 특징 정보를 이용하여 생성된 각각의 특징 지도(feature map)를 결합하여 대비 지도를 생성한다. 사람이 시각적으로 볼 때 의미 있다고 생각하는 점들 즉, 중요 특징점을 웨이블릿 변환

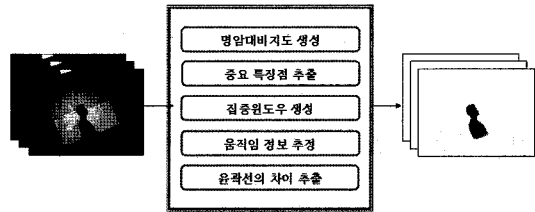


그림 1 제안하는 방법론의 흐름도

을 이용하여 찾아낸다. 이미 생성된 대비 지도와 중요 특징점을 결합하여 적절한 집중원도우의 위치와 크기를 결정한다. 또한, 동영상의 가장 큰 특징인 움직임 정보를 추정하여 집중원도우를 객체에 가장 근사하게 축소시킨다. 집중원도우 내부에 존재하는 영역에 대해서 윤곽선의 차이를 추출하여 가로 후보영역과 세로 후보영역을 찾은 후, AND 연산과 모폴로지 연산을 통하여 의미 있는 객체를 추출한다.

## 2. 관련 연구

비디오 시퀀스로부터 원하는 객체를 추출하기 위한 지금까지의 연구를 살펴보면 시공간 기반(spatial-temporal based) 객체 추출 방법, 동작(motion) 분할 객체 추출 방법, 모폴로지 기반(morphological based) 객체 추출 방법으로 크게 3가지로 분류할 수 있다.

시공간 기반의 객체 추출 방법은 공간영역의 정보와 시간 정보를 이용하여 동질영역(homogeneous region)을 추출하는 기술이다[3-5]. 이 방법은 시공간영역 정보를 이용하여 주변 영역과의 유사도를 측정하여 영역간의 관계를 나타내는 그래프를 생성한다. 이렇게 생성된 그래프는 영역을 반복적으로 결합할 때 사용된다. 이 방법은 크게 상향식 접근 방법, 하향식 접근 방법으로 나눌 수 있다. 상향식 접근 방법은 입력되는 영상의 여러 기본 특징들을 추출한 후, 이를 결합하고 전체 영상 중에서 각 영역의 중요도를 가리키는 특징을 추출하여 각 영상들 간의 유사도(similarity)를 측정하는 방법이다. 계산의 복잡도가 낮다는 장점이 있지만, 영상에서 여러 객체들 사이의 인위적인 계층관계를 생성하여야 한다는 문제점이 있다. 반면에, 하향식 접근 방법은 영역의 결합과 다른 영역과 구별되는 움직이는 객체 추출을 동시에 실시하는 방법이다.

움직임 정보를 이용한 동작 분할 객체 추출 방법은 색깔, 밝기 등과 같은 특징정보로부터 획득한 움직임 파라메타에 의존하여 객체를 추출하는 방법이다. 움직임 파라메타에 의존한다는 것은 각 프레임마다 허프 변환(Hough transform)[6], 병합(merging)[7], 베이지안(Bayesian)[8], K-means[9]와 같은 기법을 이용하여 분

할된 영역들의 눈에 띄는 움직임 벡터를 사용하는 것이다. 대부분의 동영상의 경우 배경을 뒤로 하고, 전방객체가 움직이는 것을 볼 수 있다. 이 때 프레임 간 움직임 정보를 이용하여 전역적 움직임을 분리할 수 있는데, 이 부분을 제거하여 배경에서 전방객체를 추출한다. 이 방법에도 객체와 배경의 경계에서 서로 다른 움직임을 갖기 때문에 발생하는 틸(aperture)문제와 객체의 일시적인 가려짐 현상에 대한 문제점을 가지고 있다. 이러한 문제점을 해결하고자 Chang[10]과 Stiller[11]은 움직임 정보와 영역 분할을 결합하여 사용하는 방법을 제안하였으며, Tunce[12]는 색깔, 밝기, 움직임의 변화 정보를 사용하는 방법으로 위의 문제점을 해결하였다.

모폴로지 기반의 객체 추출 방법은 모폴로지 필터를 이용하거나[13,14], 또는 모폴로지와 인접한 화소간의 공간정보를 사용하는 워터셰드(watershed) 영역 분할을 사용하는 방법들[15,16]을 말한다. 이 방법은 여러 가지 모폴로지 연산을 사용하여 잡음과 작은 변화영역을 상쇄하고, 비교적 정확한 윤곽선을 살리는 단순화(simplification) 작업을 거친 다음, 단순화된 영상에서 같은 밝기값을 가지는 영역을 추출하는 마커 추출(marker extraction) 과정을 거친다. 마커를 제외한 영역은 불확정 영역으로 보고 추출된 마커 영역에서 지역경계 결정을 위해 워터셰드 알고리즘을 적용한다. 하지만 모폴로지 기반 객체 추출 방법은 계산의 복잡도가 높고 매우 집약된 계산과정의 수행을 필요로 하는 단점이 있다.

### 3. 제안하는 방법론

본 논문에서는 밝기와 방향성 두가지 특징 정보를 이용하여 생성된 대비 지도의 중요도(saliency)와 중요 특징점을 찾는 방법을 결합하여 중요객체가 있는 위치와 크기를 결정하는 방법을 제안한다. 또한, 동영상의 가장 큰 특징인 움직임 정보를 추정하고 윤곽선 정보를 이용하여 중요객체를 자동으로 추출하는 방법을 제안한다.

#### 3.1 대비 지도(contrast map)를 이용한 중요영역 추정

대비 지도는 Itti[17]가 제안한 방법으로 다른 영역과 구별되는 현저한 장소에 무의식적으로 시선이 집중되는 인간의 시각주의 특성을 컴퓨터 시스템 내에 도입한 것이라고 볼 수 있다. 대비 지도는 인간의 시각주의 기능에 영향을 미친다고 알려져 있는 몇 가지 특징들을 입력되는 영상의 모든 영역에서 추출하여 각각의 특징에 해당되는 특징 지도를 형성한다. 이렇게 형성된 특징 지도를 결합하여 전체 영상 중에서 각 영역의 중요도를 가리키는 특징을 추출하고, 이를 주의집중에 이용하는 방식이다. 따라서, 대비 지도의 중요도를 이용하면 중요객체가 영상의 어느 부분에 존재하는가를 판단할 수 있는 척도가 될 수 있다. 본 논문에서는 Itti[17]의 알고리

즘을 수정하여 밝기 정보에 대해 6개, 방향성 정보에 대해 18개의 특징 지도를 생성하고 각각의 특징에 대해서 밝기 지도, 방향성 지도를 완성한 뒤, 2가지 특징 지도를 같은 비율로 결합하여 하나의 대비 지도를 생성한다. 또한, 본 논문에서는 실험을 통하여 우리가 일반적으로 객체를 검출할 때 중요하다고 생각하는 색상 정보가 항상 좋은 특징이 될 수 있는 것은 아니라는 것을 밝히고, 색상 지도를 제외하고 밝기 지도와 방향성 지도만을 결합하여 하나의 대비 지도를 완성하여 사용한다. 이에 대한 실험결과는 본 논문 4.2절에 기술되어 있다.

##### 3.1.1 밝기 지도(luminance map)

밝기 지도는 밝기 정보를 이용하여 생성된 특징 지도로써, 이를 생성하기 위해서 명도영상을 사용하였다. 우선 영상의 잡음을 제거하기 위해서 가우시안 스무딩 필터링(gaussian smoothing filtering)을 적용하였고, 영상의 크기와 필터 크기를 달리하여 모두 6개의 특징 지도를 생성하여 결합시켜 준다. 영상의 크기와 필터의 크기를 달리해 주는 이유는 각각의 특징 지도는 이를 생성할 때 사용되는 필터의 크기에 매우 민감하기 때문이다. 객체의 크기가 필터의 크기보다 클 경우에는 중요도가 낮게 나올 수 있고, 배경이 객체보다 오히려 높은 중요도를 가질 수도 있다. 이러한 문제점을 보완하기 위해 본 논문에서는 영상의 크기를 축소하여 여러 가지 크기의 필터를 적용시켜 나온 결과들을 모두 결합하여 하나의 밝기 지도를 만들도록 하였다. 이러한 방법을 수식 (1)로 정의하였다.

$$\bar{L} = \frac{1}{6} \sum_{c=2s}^{c+4} \sum_{s=c+3}^{c+4} N(G(c,s)) \quad (1)$$

수식 (1)에서  $G(c,s)$ 는 가우시안 스무딩 필터링을 적용한 명도영상을  $1/c$ 배의 크기로 축소하고,  $s$ 크기의 필터를 적용한 영상을 나타낸다. 함수  $N(\cdot)$ 은 괄호안의 영상의 크기를 정규화한다는 것을 의미한다. 즉, 각 영상의 크기를  $1/2$ 배,  $1/3$ 배,  $1/4$ 배로 축소하고 각각의 축소된 영상마다 필터의 크기를 달리하여 6개의 특징 지도를 생성한다. 이 6개의 특징 지도를 모두 결합하여 하나의 밝기 지도를 생성하게 된다. 이 때 주의할 점은 각각의 영상은 다른 크기로 축소되었기 때문에 결합할 때 영상의 크기를 정규화하여 결합하여야 한다. 가장 작은 영상의 크기가  $1/4$ 배로 축소된 것이기 때문에 위에서 생성된 각각의 지도는 모두 원영상의  $1/4$ 배로 정규화하여 결합한다. 그림 2에서 보듯이, 축소된 영상마다 각각 2개의 영상이 생성되므로 모두 6개의 특징 지도가 생성된다. 결합할 때에는 평균 영상을 구하기 위해서 모두 같은 가중치를 주어 6으로 나누어준다.

##### 3.1.2 색상 지도(color map)

색상 지도를 생성하기 위해서 CIE(Commission Inter-

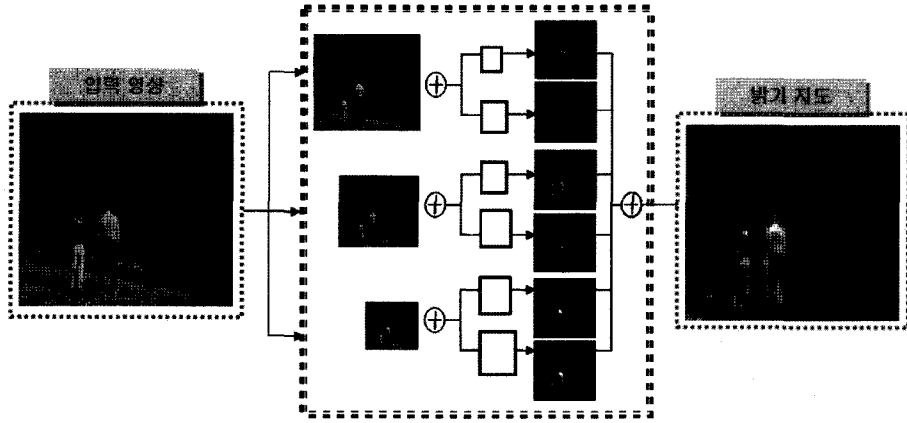


그림 2 밝기 지도 생성도

national d'Eclairage:국제조명위원회)의 L\*a\*b\* 칼라모델을 사용한다. L\*a\*b\* 칼라모델은 XYZ 칼라모델의 색과 색의 거리가 인간이 가지고 있는 인식과 대응하고 있지 않다는 단점을 보완하기 위해 개발된 칼라모델로서 인간의 색상지각을 가장 잘 나타내어 주는 색상모델이다.

색상 지도에서는 L\*a\*b\* 칼라모델에서 밝기값을 나타내는 L\*성분만을 제외하고 녹색/빨강, 파랑/노랑을 나타내는 a\*와 b\* 칼라모델을 이용한다. 색상 지도를 생성하기 위해 밝기 지도를 생성하는 방법과 같이 영상의 잡음을 제거한 다음, 영상의 크기와 필터의 크기를 달리하여 a\* 색상체계에 대해 6가지의 특징 지도, b\* 색상체계에 대해 6개의 특징 지도를 생성한다. 그림 3에서 보듯이 12가지의 특징 지도가 생성된다. 12가지의 특징 지도도 영상의 크기가 각기 다르므로 가장 작은 영상크기인 원영상의 1/4배로 줄여서 모두 결합하여 하나의 색상 지도를 완성한다. 이 특징 지도를 생성할 때에도 전처리 단계로 영상처리기법 중 잡음을 제거하기 위해 가우시안 스무딩 필터링을 취해준다. 이러한 방법을 수식 (2)로 정의한다. 수식 (2)에서 A(c,s)와 B(c,s)는 L\*a\*b\* 칼라모델에서 a\*칼라모델과 b\*칼라모델을 1/c배의 크기로 축소하고, s크기의 필터는 적용한 영상을 나타낸다. 함수 M(·)은 괄호안의 영상을 정규화 한다는 것을 의미한다. 즉, 밝기 지도와 같은 방법으로 a\* 색상체계에 대해 6가지의 특징 지도, b\* 색상체계에 대해 6개의 특징 지도를 생성한 뒤, 이를 정규화하고 결합하여 색상 지도를 획득한다.

$$\bar{C} = \frac{1}{12} \sum_{c=2s}^{c+4} \sum_{s=c+3}^{c+4} [M(A(c,s)) + M(B(c,s))] \quad (2)$$

일반적으로는, 색상 정보는 객체와 배경을 구분하는 중요한 특징으로 알려져 있지만 항상 좋은 특징이 될 수 있는 것은 아니다. 색상 정보는 객체와 배경의 색에

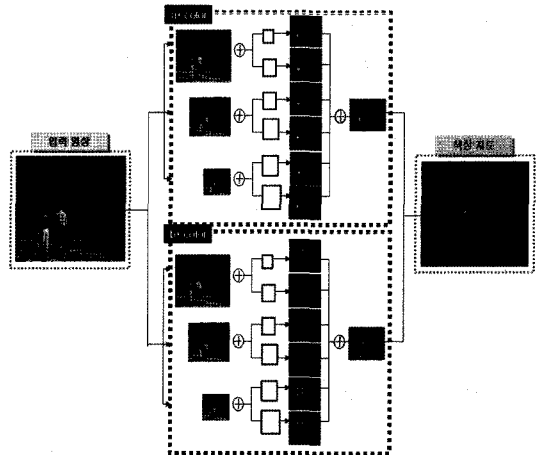


그림 3 색상 지도 생성도

매우 의존적이기 때문이다. 그림 4의 (a)는 검은 바탕에 빨강색의 객체를 나타내고 (b)는 빨강 바탕에 분홍색의 객체를 나타내는 그림이다. (a)는 다른 영역과 구별되는 현저한 장소에 무의식적으로 집중되는 시각적 주의 특성에 의해 빨강색 객체에 초점이 맞춰지는 반면 (b)는 배경의 대부분이 빨강색으로 채워져 있기 때문에 중앙의 분홍색 객체에 초점이 가지 않고 빨강색 배경에 초점이 맞춰진다는 것을 알 수 있다. 이러한 현상은 색상 정보는 주위 환경에 따라 달라지므로 인간의 시각 주의 특성을 결정짓는 결정적인 요인은 될 수 없다는 것을 보여준다. 이러한 특성을 본 논문 4.2절에서 정지영상에서 색상 정보를 포함한 3가지 특징정보를 사용하여 객체를 추출한 방법과 색상 정보를 제외한 2가지 특징정보를 사용하여 객체를 추출한 방법에 대해 실험을 통하여 색상 정보가 항상 중요한 특징이 될 수 없음을 증명하였으며 그림 5에서는 대비 지도를 생성할 때 나쁜 영

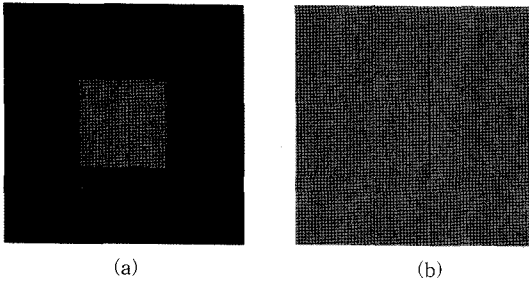


그림 4 색채 대비 (a) 검은 바탕의 빨강 객체 (b)빨강 바탕의 분홍 객체

향을 미칠 수 있는 색상 지도의 예를 보여주고 있다. 또한, 자연영상에서 동물은 보호색으로 인해 주변 배경과 유사한 색상을 가짐으로 오히려 대비 생성에 좋지 못한 영향을 미칠 수 있다.

3.1.3 방향성 지도(directional map)

중요한 영역은 주변 지역에 비해 눈에 띄는 질감정보를 가지고 있다. 본 논문에서는 이러한 질감정보를 이용하여 방향성 지도를 생성하기 위해 빠른 수행 능력과 여러 가지 질감을 쉽게 얻을 수 있는 웨이블릿 변환을 이용하였다. 1단계 웨이블릿 변환을 하면 저주파 부밴드, 수평방향(LH), 수직방향(HL), 대각선방향(HH)의 4개의 부밴드가 생성된다. 이렇게 생성된 부밴드 중 방향성 성분을 나타내는 수평, 수직, 대각선 부밴드를 특징 정보로 이용하여 방향성 지도를 생성한다. 밝기 지도와 색상 지도를 생성하는 방법과 같이 방향성분을 나타내고 있는 영상에 대해서도 잡음을 제거하기 위해 가우시안 스무딩 필터링을 취한 뒤, 각각 다른 크기로 축소하고 필터의 크기를 달리하여 특징 지도를 생성한다. 결합하기 전 영상을 정규화 할 때 주의할 점은 웨이블릿 변환을 통하여 생성된 영상은 이미 1/2배의 크기로 축소된 영상이므로 원영상의 1/3배, 1/4배로 크기를 축소하여 결합하여야 한다. 이렇게 획득한 특징 지도를 선형

결합하여 하나의 방향성 지도를 완성한다. 위에서 설명한 방법을 수식 (3)으로 정의하였다.

$$\bar{O} = \frac{1}{18} \sum_{HH, HL, LH} \sum_{c=2s}^{c+4} \sum_{s=c+3} N(O(c, s)) \quad (3)$$

방향성 지도는 3개의 방향에 대해 각각 6개의 특징 지도가 생성되므로 그림 6에서 보듯이 모두 18개의 특징 지도가 완성된다. 이들의 평균 영상을 획득하기 위해 18로 나누어 준다.

수식 (1)(2)(3)에 의하여 밝기 지도, 색상 지도 그리고 방향성 지도가 완성된다. 이 3가지 특징 지도 중 본 논문에서는 색상 지도를 제외한 2가지 특징지도에 대해서

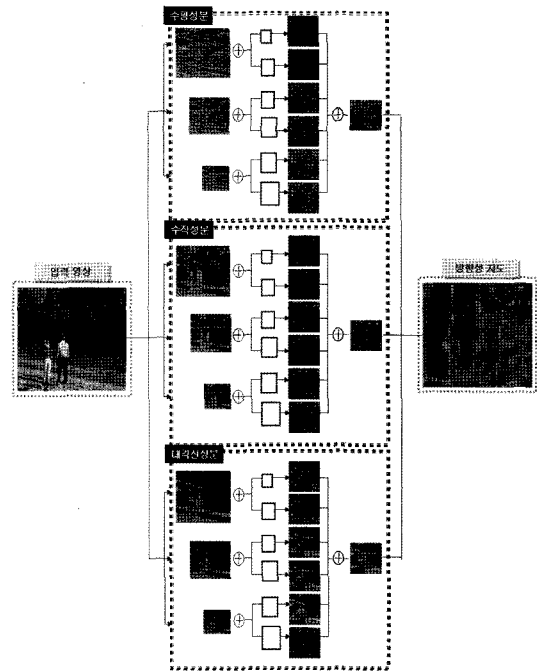


그림 6 방향성 지도 생성도

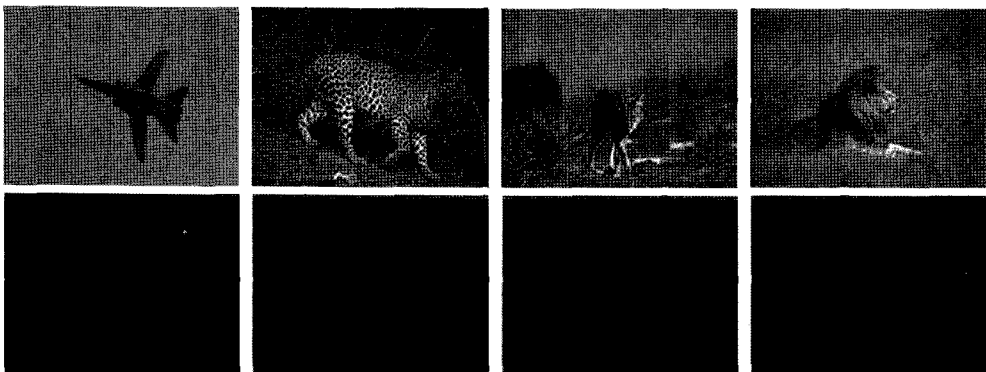


그림 5 대비 지도를 생성할 때 나쁜 영향을 미칠 수 있는 색상 지도

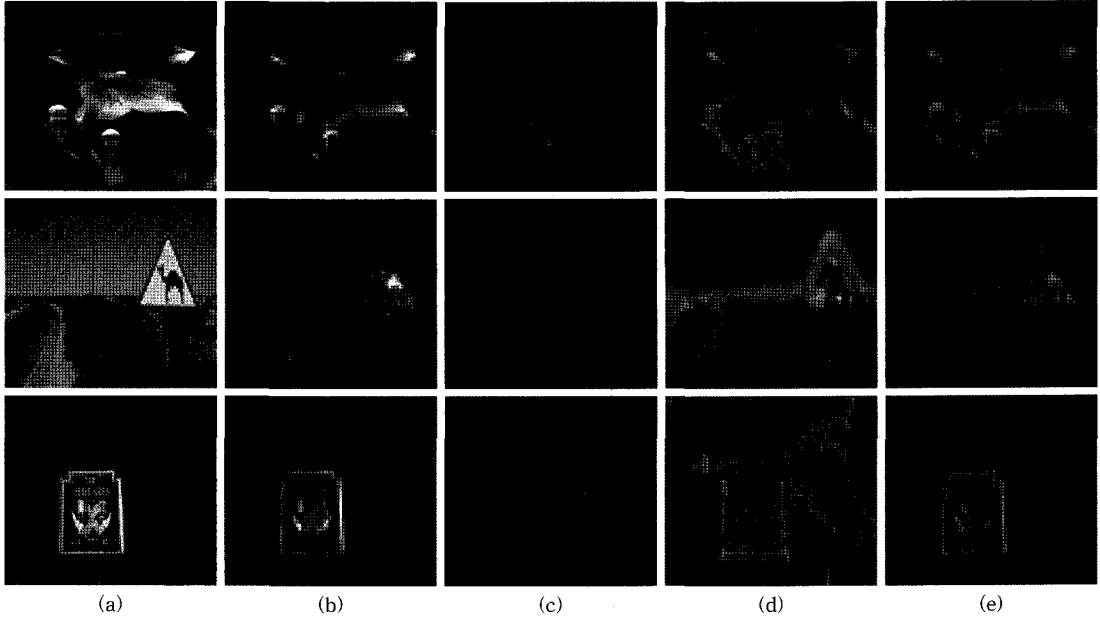


그림 7 대비 지도 생성 과정 (a) 원 영상 (b) 밝기 지도 (c) 색상 지도 (d) 방향성 지도 (e) 최종 대비 지도

만 수식 (4)와 같이 모두 같은 비율로 선형 결합하여 하나의 대비 지도를 완성한다. 선형 결합된 영상은 원영상의 1/4배로 축소된 영상이므로 한 화소 값을 4개의 화소에 채워주어 원영상의 크기와 같게 하여 준다. 그렇기 때문에 그림 7처럼 블록현상이 나타남을 알 수 있다. 그림 7는 원 영상을 수식 (1),(2),(3)을 이용하여 생성된 3가지의 특징 지도, 즉 밝기 지도, 색상 지도 그리고 방향성 지도의 결과와 수식 (4)에 의해 2가지의 영상을 선형 결합하여 생성된 대비 지도를 보여 주고 있다.

$$S = \frac{1}{2}(\bar{L} + \bar{O}) \tag{4}$$

**3.2 중요 특징점을 이용한 집중원도우 설정**

대비 지도를 생성 후 중요객체가 존재하는 영역을 찾기 위해 집중원도우의 위치를 찾는다. 본 논문에서는 집중원도우가 고정적으로 중앙에 위치하는 Kim[2]의 방법과는 달리, 집중원도우가 영상의 특성에 따라 가변적으로 변화도록 설계하였다. 본 논문에서는 적절한 집중원도우의 위치 및 크기를 추정하기 위해 영상으로부터 중요 특징점을 추출하는 방법을 사용하였다.

**3.2.1 중요 특징점**

중요 특징점은 영상에서 눈에 띄는 점들의 집합이라고 할 수 있다. 이 점들은 주파수의 변화가 심한 영역에서 많이 발생함으로 객체와 같이 배경에서 두드러지게 분리되는 경계부분에 많은 중요 특징점을 가지게 된다. 따라서 중요한 객체는 항상 다수의 중요 특징점을 가지게 되고 이러한 사실을 이용하여 중요 특징점 분포가

가장 많은 영역을 찾아 집중원도우의 위치를 결정한다. 본 논문에서는 Loupias[18]가 제안한 웨이블릿 변환을 이용하여 중요 특징점을 찾는 방법을 사용하였다. 입력된 영상은 웨이블릿 변환을 하게 되면 저대역 부밴드와 고대역 부밴드로 나뉘게 된다. 고대역 부밴드는 수평, 수직, 대각선 성분이 주로 남게 되는데 이들의 상관계수(coefficient)를 계산하여 상관계수가 높은 부분만을 추출하여 중요 특징점이라 한다. 웨이블릿 변환은 3단계까지 실시하였으며 3단계에서 추출된 중요 특징점을 위의 단계로 차례로 매핑하여 원 영상에 중요 특징점이 나타나게 한다(그림 8). 윗 단계로 매핑 할 때마다 영상의 크기가 다르다는 것을 고려하여야 한다. 저대역 부밴드와 고대역 부밴드의 계수를 어떻게 설정하느냐에 따라 Haar나 Daubechies 웨이블릿으로 나누어지는데 본 논문에서 웨이블릿은 수행속도가 빠른 Haar 웨이블릿을 사용하였다. 그림 9의 노란색 점들은 Haar 웨이블릿을 사용하여 중요 특징점을 추출한 결과이며, 객체와 배경의 경계부분과 객체의 두드러지는 점에서 많이 나타남을 알 수 있다.

**3.2.2 집중원도우 생성**

집중원도우는 영상에서 눈에 띄는 부분이나 중요객체가 존재하는 영역의 위치를 알려주는 틀이다. 집중원도우는 중요하지 않은 객체 또는 배경을 제거하는 역할을 하기 때문에 중요객체를 추출하는 시간을 줄여줄 뿐만 아니라 보다 정확한 위치를 파악하게 해 주는 역할을 한다. 이러한 이유로 집중원도우는 객체의 크기에 가장

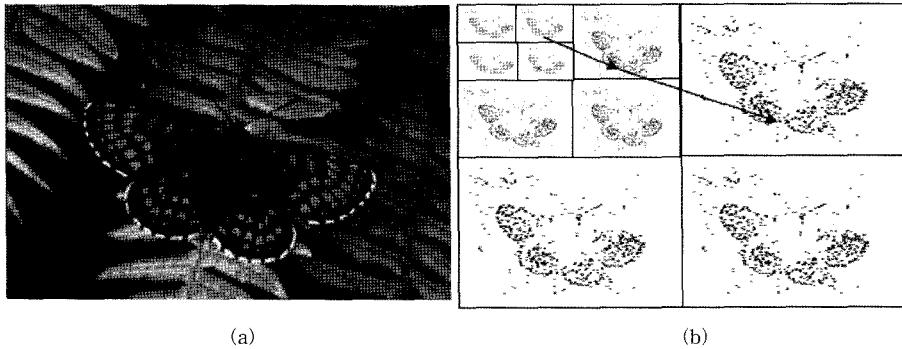


그림 8 중요 특징점 추출 방법 (a) 원영상 (b) 중요 특징점 매핑 방법

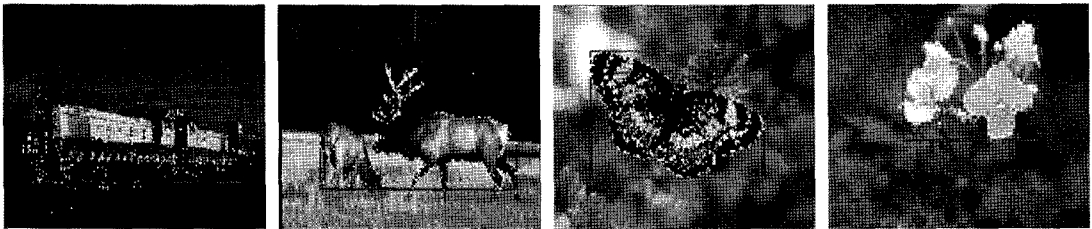


그림 9 대비 지도와 중요 특징점을 이용한 집중윈도우 생성 및 축소 결과

근사하게 결정될수록 객체를 추출하는데 효과적이라고 할 수 있다. 정확한 집중윈도우의 위치를 찾기 위해 본 논문에서는 중요 특징점과 대비 지도를 결합한 방법을 제안한다.

객체에 가장 근사하는 집중윈도우를 찾기 위해 집중윈도우의 초기 크기는 입력영상의 3/4으로 초기화 하였으며, 대비 지도의 중요도 값이 특정 임계값 이상이고 해당 중요도 값이 나타나는 부분에 중요 특징점이 동시에 존재할 경우 해당 영역으로 집중윈도우의 크기와 위치를 결정시켜 준다. 이러한 방법으로 생성된 집중윈도우의 결과를 그림 9에 나타내었다. 이렇게 결정된 집중윈도우는 이후 움직임 검출을 통해 객체의 크기와 가장 유사하게 다시 축소된다.

**3.3 움직임 정보 추정 및 객체 추출**

**3.3.1 움직임 정보 추정**

동영상의 객체들은 각각 다른 방향과 속도로 움직이기 때문에 움직임 정보는 동영상을 분석하기 위한 중요한 특징이다. 이와 관련하여 광류흐름(Optical flow)측정법, 블록 매칭(Block matching) 등 다양한 접근방식으로 움직임 추정에 관한 연구가 진행되었다. 전통적으로는 지역적인 움직임 정보를 표현하기 위해 광류를 많이 사용하였으나 광류 방법은 잡음에 민감할 뿐만 아니라 같은 질감이나 색상이 존재하는 영역에 대해서는 움직임이 존재함에도 불구하고 움직임 정보가 표현되지 않는 단점이 있다. 그림 10에서 빨강색 나타낸 점들이 광류에 의한 움직임 정보를 나타내는 점들이고, 그림에서 보는 바와 같이 사람의 옷 부분은 움직임이 존재하였음에도 불구하고 움직임 정보가 정확히 표현되지 않



그림 10 광류에 의한 움직임 추출

은 것을 확인 할 수 있다. 또한, 계산량이 상당히 많은 단점도 있다.

블록 매칭 알고리즘은 현재 프레임의  $n \times m$  크기를 갖는 여러 개의 기준 블록(reference block)으로 분할하여 각각의 기준 블록을 이전 프레임의 탐색 영역(search window) 내에서 가장 닮은 블록(정합블록)을 찾아내어, 기준 블록에 대한 정합 블록의 상대 위치를 움직임 벡터로 정한 후 두 블록간의 화소 값 차이와 움직임 벡터만을 전송하는 기법이다. 기준 블록과 가장 닮은 즉 정합 블록을 찾아내는 판단기준 즉, 유사도를 측정하는 기준은 수식 (5)의 SSE(sum of square error)를 사용한다. SSE가 최소인 블록이 정합 블록이 되는 것이다. (수식 (5)에서  $y_i$ 는 관찰된 프레임을 나타내고  $y'_i$ 는  $y_i$ 의 다음 프레임을 나타낸다.) 하지만 블록 매칭 알고리즘은 조명의 변화 또는 카메라의 움직임이 조금이라도 발생하는 경우 배경과 객체의 움직임의 크기 정보를 비교하기 어렵다는 단점이 있다.

$$SSE = \sum_{i=1}^n \sqrt{(y_i - y'_i)^2} \quad (5)$$

본 논문에서는 Shi[5]가 제안한 움직임 정보 추출방법에서 착안하여 각 화소의 움직임 크기 정보를 영상에 나타내어 주변 화소와의 상관도(correlation)를 측정하고 움직임 정보의 유사도를 계산하는 방법을 제안한다. 제안된 움직임 정보 추출방법은 그림 11에 나타내었다. 즉, 기준 블록의 평균값과 탐색영역 내부에 존재하는 모든 비교 블록의 평균값의 유사도를 측정하여 가장 유사한 비교 블록을 찾는다. 이때 유사도를 측정하기 위해 수식 (6)을 정의하여 사용하였으며, 유사도가 높은 블록은  $S_{(x,y)}$ 의 값이 최소가 되는 블록을 선택하였다. 가장

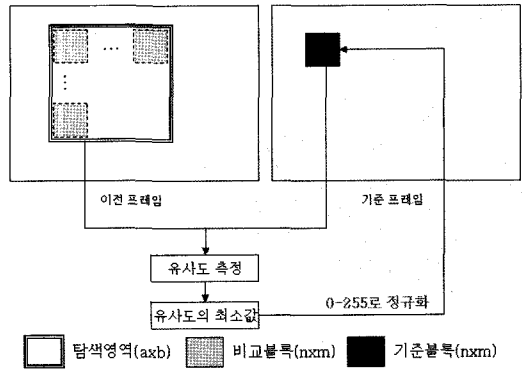


그림 11 제안하는 움직임 정보 추출 방법

유사하다고 판단되는 비교 블록의 유사도의 값을 0-255로 정규화 하여 기준 블록에 입력하여 주면 그 값이 움직임 정보의 크기값으로 결정되는 것이다.

$$S_{(x,y)} = \frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n \left( \sqrt{(\mathcal{C}(i,j) - P_{(x,y)}(i,j))^2} \right) \quad (6)$$

where  $i \leq x \leq i+a$   
 $j \leq y \leq j+b$

기준 블록의 크기가  $n \times m$ 이고, 탐색영역 크기가  $a \times b$ 일 때, 수식 (6)에서  $\mathcal{C}(i,j)$ 는 기준이 되는 현재 프레임의 기준 블록을 나타내고,  $P_{(x,y)}(i,j)$ 는  $\mathcal{C}(i,j)$ 의 바로 이전 프레임의 비교 블록을 나타낸다.  $(i,j)$ 는 기준 블록과 비교 블록의 범위이며  $(x,y)$ 는 탐색영역의 범위이다. 수식 (6)에 정의한  $S_{(x,y)}$ 는 기준 블록과 비교 블록 사이의 유사도를 측정하는 수식으로, 기준 블록의 움직임 크기를 나타내기 위해서는 탐색영역 범위 안에서  $S_{(x,y)}$  값이 최소가 되는 값을 0-255로 정규화 하여 나타

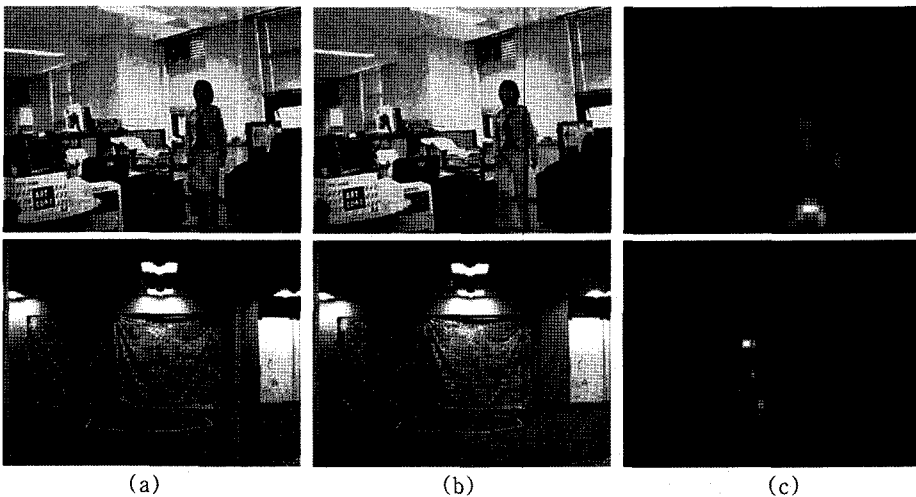


그림 12 움직임 추정 결과 (a) 이전 프레임 (b) 현재 프레임 (c) 움직임 크기정보를 나타낸 영상



낸다. 이 값이 클수록 움직임의 변화가 크게 일어났다는 것을 알 수 있으며 영상에서는 정규화 하여 표현하였기 때문에 흰색(255, 255, 255)에 가깝게 즉, 밝은 값으로 표현된다.

본 논문에서 제안한 움직임 추정 방법을 집중윈도우 내부영역에 대해서만 추출하고 움직임의 크기 정보에 따라 추출하고자 하는 객체에 가장 근사하게 집중윈도우의 크기를 축소하여 가장 적합한 크기와 위치를 결정해 준다. 정지영상에서는 움직임 정보를 얻을 수 없기 때문에 본 논문 3.1절과 3.2절에서 설명한 대비 지도와 중요 특징점을 이용하여 영상에서 공간적인 정보만으로 결정된 집중윈도우의 정보로써 객체를 추출하지만 연속적인 영상을 가지고 있는 동영상에서는 움직임 정보를 추가하여 정확한 집중윈도우를 결정한다. 3.3.2절에서 설명할 객체 추출 과정은 3.2절의 결과로 나온 집중윈도우의 내부영역에 대해서만 고려하기 때문에 대부분의 배경을 제거해 줌으로써 객체 추출을 효율적으로 할 수 있게 도와준다.

3.3.2 객체 추출

움직이는 객체를 추출하기 위하여 3.3.1절에서 서술한 움직임 정보와 객체의 윤곽선(contour)를 이용한다. 윤곽선은 에지(edge)를 추출한 후 이들을 연결한 선을 나타낸다. 윤곽선 정보는 객체를 배경으로부터 구분해 내는 중요한 정보일 수 있으나, 추출하고자 하는 객체뿐만 아니라 배경의 윤곽선 정보도 함께 나타날 수 있다는 문제점이 있다. 이러한 문제를 해결하기 위해 비디오 시퀀스에서 이전 프레임과 현재 프레임과의 윤곽선의 차이를 이용한다. 현재 프레임과 이전 프레임과의 윤곽선 차이를 구하면 그림 13의 (c)에서 보는 바와 같이 배경의 윤곽선을 대부분 제거할 수 있다. 본 논문에서는 앞에서 설명한 윤곽선 차이를 DE(Difference of Edge)라고 칭하고 수식 (7)로 정의하겠다. 즉, 이전 프레임과 현재 프레임에 가우시안 스무딩 필터링을 취하고 캐니 에지를 추출한 다음, 이전 프레임과 현재 프레임의 캐니 에

지의 차를 구하여 윤곽선 차이를 구할 수 있다. 수식 (7)에서  $f_n$ 은 현재 프레임을 나타내고  $f_{n-1}$ 은 이전 프레임을 나타낸다.  $\nabla G \times f$ 는 가우시안 스무딩 필터링을 취한 프레임이고,  $\Phi(f)$ 는 f 프레임에 대한 캐니 에지 추출을 나타낸다.

$$DE = |\Phi(\nabla G \times f_{n-1}) - \Phi(\nabla G \times f_n)| \quad (7)$$

윤곽선 정보를 이용하여 객체를 추출하기 위해서는 끊어진 에지를 이어주고 불필요한 배경의 에지를 제거하는 윤곽선 보정작업이 필요하다. 이러한 보정작업은 모폴로지 연산을 이용하였다. 모폴로지 연산 중 팽창연산과 침식연산 순으로 조합하여 수행하는 닫힘 연산을 사용하였다. 닫힘 연산은 팽창연산으로 객체를 확장하기 때문에 객체 내 빈 공간이 배운 후, 침식연산으로 원 영상의 크기로 돌아가게 한다. 즉, 팽창연산의 효과로 인해 배경보다 객체를 더 강조하게 된다. 이러한 방법으로 윤곽선을 보정한 다음, 움직임 정보와 추출된 객체의 윤곽선을 함께 이용하여 가로 후보영역과 세로 후보영역을 추출한다. 가로 후보영역은 왼쪽에서 오른쪽으로 이동하면서 움직임이 크게 존재하고 윤곽선이 나타나는 부분의 위치를 저장해 두고, 오른쪽에서 왼쪽으로 이동하면서 움직임이 있고 윤곽선이 나타난 부분의 위치를 찾는다. 이렇게 찾아낸 두 위치를 이어주어 가로 후보영역을 추출하게 된다. 세로 후보영역 추출도 가로 후보영역 추출과 방법은 같으나 가로로 이동하지 않고 위에서 아래로 이동하면서 움직임과 윤곽선이 같이 나타나는 초기 위치를 찾고, 아래에서 위로 이동하면서 위치를 찾아 두 위치의 사이를 이어준다. 이렇게 찾아낸 가로 후보영역과 세로 후보영역은 AND 연산을 실시하여 중요 객체를 추출한다. 이렇게 추출한 객체 또한 잡음이 존재함으로 모폴로지 연산 중 객체를 강조해 주는 팽창연산을 수행하여 최종적으로 움직이는 객체를 추출한다. 그림 14는 3.3절에서 설명한 방법을 차례로 나타낸 결과이다.

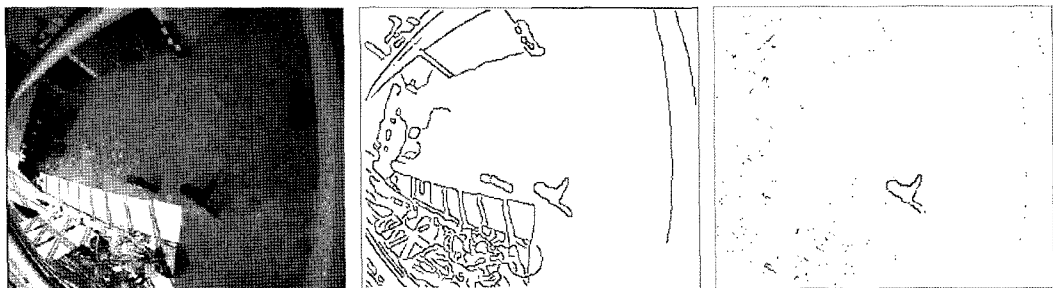


그림 13 윤곽선의 차이를 추출한 결과 (a) 원 영상 (b) 캐니에지 (c) 윤곽선의 차이 영상

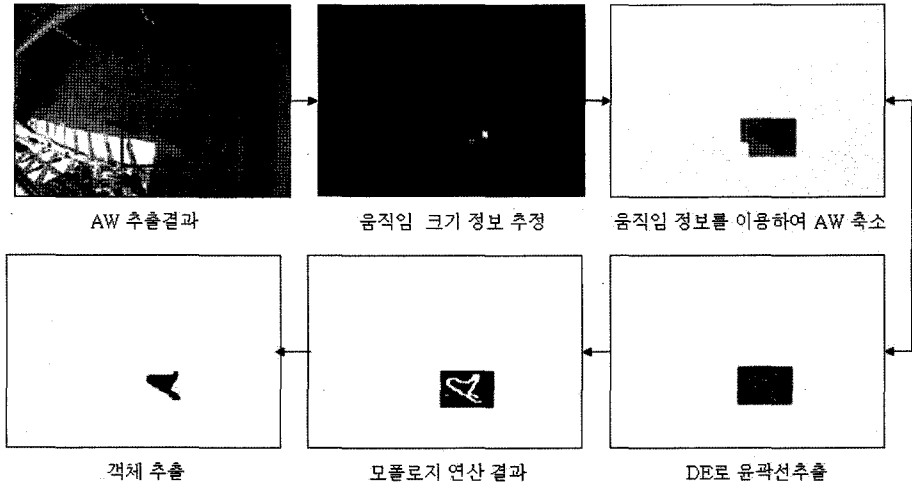


그림 14 움직임 정보와 윤곽선의 차이를 이용한 객체 추출 흐름도

4. 실험결과 및 평가

4.1 실험 환경

본 논문의 실험은 아래와 같은 실험환경에서 시도되었다.

표 1 실험 환경

CPU	Pentium4 2.40GHz
메모리	512MB
운영체제	Microsoft Window XP Professional version 2002
실험도구	Microsoft Visual C++ 6.0

4.2 대비 지도의 구성요소 중 색상 지도에 대한 실험

본 세부 실험은 Itti[17]가 제안한 대비 지도를 구성하는 밝기, 색상, 방향성 3가지 특징 지도 중 색상 지도가 중요한 특징정보가 될 수 있는 가를 평가하는 실험이다. 이 실험을 하기 위해 본 논문 3장에서 제안한 알고리즘에 영상을 분할하는 Ko[19]의 영역분할 방법을 추가하여 정지영상에서 객체를 추출하였다. 실험데이터는 Corel Photo CD에서 중요한 객체가 존재하는 영상을 임의로 300개를 선정하였다.

선정된 영상은 그래픽 툴(Adobe Photoshop 6.0)을 이용하여 그림 16의 (b)와 같이 추출하고자 하는 객체를 정확하게 잘라 내고 배경부분을 흰색(255,255,255)으로 채워주어 MEI(Manually Extracted Image)를 생성하고, 본 시스템을 통해 나온 결과물은 그림 16의 (c)와 같이 배경을 검정색(0, 0, 0)으로 채워 AEI (Automatically Extracted Image)를 생성한다. 생성된 MEI영상과 AEI영상을 비교하여 성능을 평가하게 된다. 성능평가 시스템은 정확히 추출된 결과(MEI) 보다 더 많은

영역이 추출된 경우와 추출되어야 하는 영역이 추출되지 않은 경우에 대한 에러율(error rate)을 계산하여 정확도를 평가하는 방법이다. 이 방법론을 수식 (8), (9)로 정의하여 사용하였다. 수식 (8)은 앞에서 설명한 2가지 경우에 대한 에러율이고, 수식 (9)는 정확도를 나타내는 백분율 값이다.

$$S_U = \frac{M - (M \cap E)}{S_M}, S_O = \frac{E - (M \cap E)}{S_E} \quad (8)$$

$$A = \{1 - (S_U + S_O)\} \times 100 \quad (9)$$

수식 (8)에서  $M$ 은 그래픽 툴로 중요한 객체만 잘라낸 영상(MEI)의 화소 개수를 나타내고,  $E$ 는 시스템을 통해 추출된 객체(AEI)의 화소 개수를 나타낸다.  $S_M$ 과  $S_E$ 는  $M$ 과  $E$ 의 개수(cardinality)를 의미한다.  $S_U$ 는 그림 15의 (a)의 빗금 친 부분을 의미하는 MEI영역에서 MEI와 AEI의 공통되는 화소들을 제거하여 추출되지 않은 영역 즉, 하위 추출영역(under-extraction)에 대한 에러율을 나타내고,  $S_O$ 는 그림 15의 (b)의 빗금 친 부분을 의미하는 AEI영역에서 MEI와 AEI의 공통되는 화소들을 제거하여 추출되지 않아도 되는 배경이 추출된 영역 즉, 상위추출영역(over-extraction)에 대한 에러율

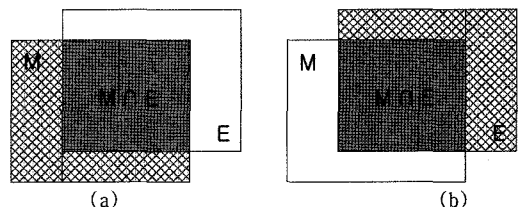


그림 15 평가 시스템 (a) 하위 추출영역, (b) 상위 추출 영역

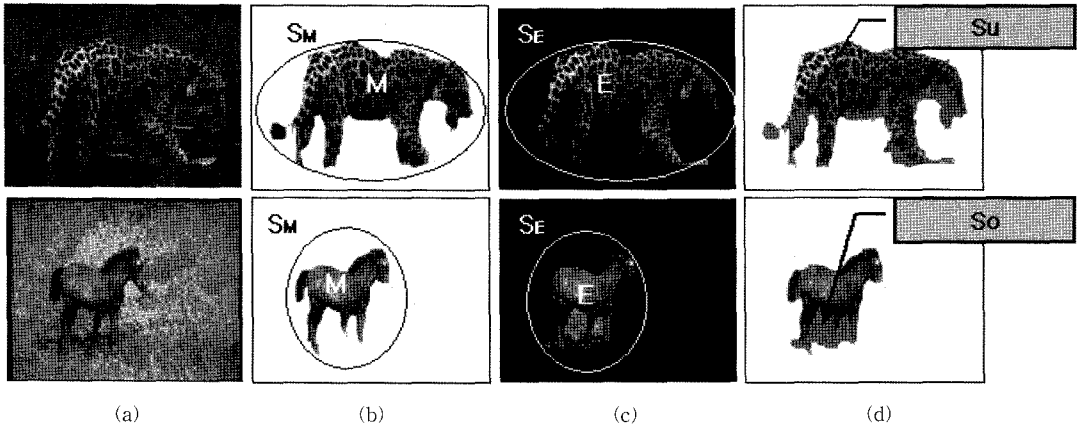


그림 16 평가시스템의 예 (a) 원영상 (b) MEI (c) AEI (d) 하위추출영역과 상위추출영역의 예

을 수식으로 나타낸 것이다. 수식 (9)는 수식 (8)의 에러율의 여집합으로 그래픽 툴로 중요 영역을 추출한 영상과 시스템을 통해 추출된 중요한 영역사이의 정확성을 백분율로 나타낸 값이다. 객체가 정확하게 추출되었다면 에러율을 나타내는  $S_0$ 와  $S_o$ 는 0에 가까운 값을 나타내고 정확성을 나타내는  $A$ 는 100에 가까운 값을 나타낸다.

그림 17은 색상 정보가 중요한 특징정보가 될 수 있는 가를 Corel Photo CD의 데이터베이스의 영상으로 실험한 결과이다. 대비지도를 생성할 때 밝기, 색상, 방향성 3가지 특징정보를 결합(L+C-O)하여 객체를 추출한 결과와 색상 정보를 제외한 밝기, 방향성 특징만을 결합(L+O)하여 객체를 추출한 결과를 비교한 것이다. 실험결과는 수식 (8), (9)에서 정의한 성능평가 수식을 이용하였다. 그림 17의 결과 값에서 보듯이 3가지 특징정보를 결합한 대비 지도(L+C+O)가 상위추출영역( $S_o$ )에 대한 에러가 더 높다는 것을 알 수 있으며, 색상 정

보를 제외한 밝기, 방향성 특징만을 결합한 대비 지도(L+O)가 정확성을 나타내는  $A$ 값이 더 높음을 알 수 있다. 즉, 색상정보는 중요 객체를 구분하는 중요한 성분일수도 있지만 응용 분야에 따라서는 오히려 객체 구분을 어렵게 만들기 때문에 본 논문에서는 이러한 색상 정보를 제외하여 중요 객체 추출의 성능을 개선하였고 이를 실험을 통해 증명해 보였다. 따라서 본 시스템에서 밝기 지도와 방향성 지도만을 결합하여 대비 지도를 생성하여 실험하였다.

### 4.3 대비 지도와 중요 특징점의 결합성에 대한 실험

본 세부실험은 논문에서 제안한 대비 지도가 영상에서 객체를 추출할 때 도움이 되는가를 알아보는 실험이다. 본 실험을 한 이유는 집중윈도우 내부영역에 대해서만 움직이는 객체를 추출하므로 집중윈도우가 얼마나 정확한 크기와 위치를 가지는가는 아주 중요한 문제이다. 따라서 집중윈도우 생성의 정확성을 따져보기 위하여 본 실험을 하였다.

본 실험을 위해 4.2절의 실험과 동일한 데이터베이스를 사용하였고 중요 특징점 만으로 집중윈도우를 생성하여 객체를 추출한 방법, 명암대비 지도만으로 집중윈도우를 생성하여 객체를 추출한 방법, 그리고 명암대비 지도와 중요 특징점을 결합하여 집중윈도우를 생성하고 객체를 추출한 방법으로 나누어 실험을 하였다. 실험의 성능평가를 위해 4.2절에서 제안한 수식 (8), (9)를 이용하여 그림 18에 실험결과를 나타내었다. 그림 18의 3가지 방법론에 관한 성능 평가를 보면 하위추출영역( $S_u$ )의 에러는 비슷하지만 상위추출영역( $S_o$ )에 대한 에러가 더 높다는 것을 알 수 있으며, 명암대비 지도와 중요 특징점을 결합하여 중요객체를 추출한 방법이 다른 2가지 방법에 비해 정확성을 나타내는  $A$ 값이 더 높음을 알 수 있다. 즉, 사람이 중요하다고 생각하는 객체를 추출하기

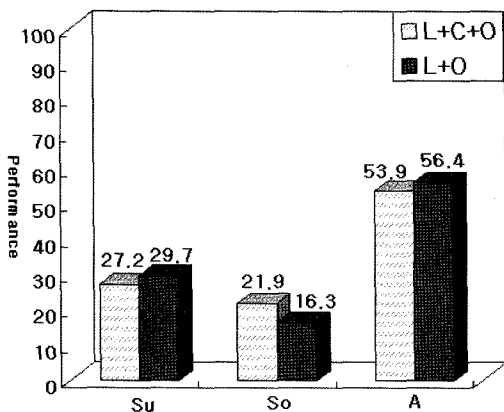


그림 17 대비 지도의 성능평가 비교

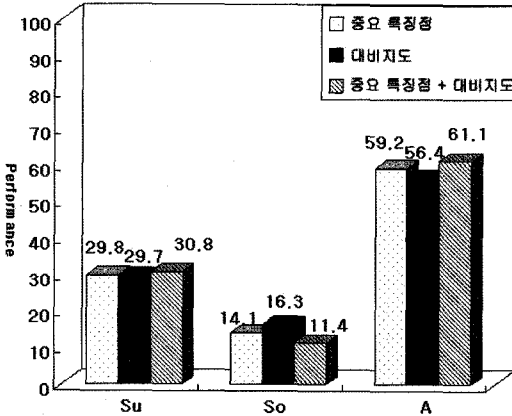


그림 18 3가지 방법론에 관한 성능 평가 비교

위해 현저한 장소에 무의식적으로 시선이 집중되는 시각적 주의 특성을 명암대비 지도와 중요 특징점을 결합하여 적용함으로써 객체의 대략적인 위치를 판단함에 있어 좋은 특징이 될 수 있다고 말할 수 있다.

4.4 동영상에서의 객체 추출

본 논문에서 제안한 알고리즘을 동영상에 적용하여 객체를 추출하였다. 실험 영상은 University of Edinburgh에서 공개한 CAVIAR데이터[12], CIPR(Center for Image Processing Research)[20]에서 공개한 데이터와 웹 캠(web cam)으로 직접 촬영한 320×240크기의 동영상을 사용하였다. 또한 능동(Active) 카메라를 이용

하여 아무런 제약 없이 획득한 동영상에서 객체를 찾아 인식하고 추적하기 위해서는 상당한 어려움이 따르기 때문에 실험할 동영상에는 몇 가지 제약을 두었다. 감시 시스템처럼 데이터를 획득하기 위해 고정(Static) 카메라를 사용하였고, 획득한 영상에는 사람이 주시할 만한 객체가 하나이상 존재하며, 배경과 물체가 비교적 잘 구분되는 영상을 주로 사용하였다. 그림 19의 데이터는 본 연구에서 직접 웹 캠으로 입력받은 동영상에서 객체 추출 결과를 보여주고 있다. 이 동영상은 카메라의 움직임은 없지만, 객체의 움직임이 크다는 특징을 갖고 있다. 웹 캠을 통하여 생성된 영상이기 때문에 잡음을 많이 포함하고 있다는 단점이 있음에도 불구하고 정확하게 움직이는 객체를 분할하는 것을 확인하였다. 그림 20은 CAVIAR데이터의 첫 번째 셋의 walking 데이터를 이용한 것이다. 이 동영상은 감시 카메라가 높게 고정되어 있으며 사람이 움직이는 영상이다. 카메라가 객체와 멀리 떨어져 있고 왼쪽에는 조명의 변화가 있음에도 불구하고 작은 객체를 잘 찾아내는 것을 볼 수 있다. 그리고 그림 21은 객체 추출에서 많이 사용하는 CIPR의 claire 동영상이다. 이 데이터는 배경이 단순지만 객체의 주된 움직임이 얼굴을 중심으로 나타나기 때문에 얼굴만 추출될 수 있다. 하지만 본 논문에서 제안한 방법은 객체 전체를 잘 찾아내는 것을 볼 수 있다.

본 논문의 방법론을 많은 데이터로 실험해 본 결과 집중윈도우의 크기가 유동적으로 변하기 때문에 객체크

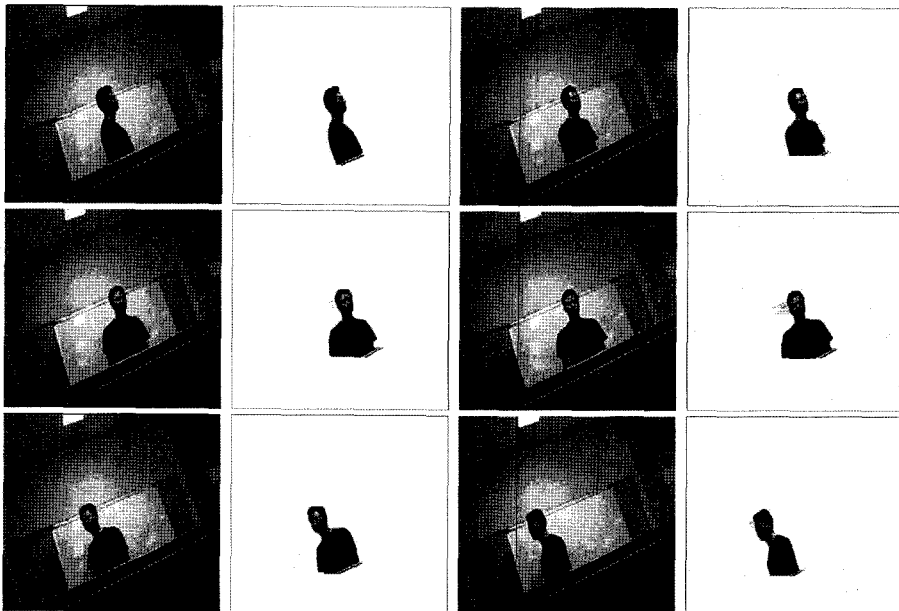


그림 19 객체 추출 결과(1)-직접 촬영한 데이터

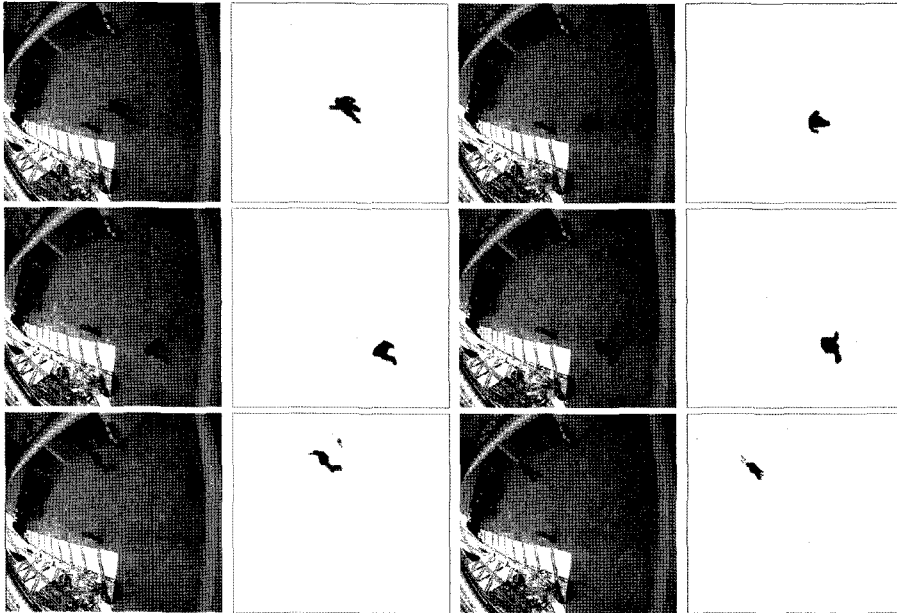


그림 20 객체 추출 결과(2)-CAVIAR데이터



그림 21 객체 추출 결과(3)-claire동영상

기에 무관하게 움직이는 객체가 잘 추출됨을 알 수 있다. 반면에 집중원도우가 생성될 때 잘못된 정보로 인하여 객체에 걸려지게 생성되거나 또는 다른 객체 위로 생성 될 경우 엉뚱한 영역의 움직임 정보와 윤곽선을 추출하기 때문에 잘못된 객체 추출 결과를 낼 수 있다. 또한, 윤곽선 정보를 사용하는 방법이기 때문에 객체 가까이 배경이 복잡하여 배경의 윤곽선 정보가 포함될 경우 문제가 발생하기도하는 문제점이 있다.

### 5. 결론 및 향후 연구 과제

본 논문에서는 시공간 정보와 움직임 정보를 결합하여 동영상에서 중요객체를 자동으로 추출하는 방법을 제안하였다. 본 시스템은 객체가 영상의 중심에 존재한다는 가정을 이용하여 실험한 Kim[2]등의 방법과 달리 아무런 사전 가정 없이 자동으로 전체영상에서 중요객체를 추출한다. 또한 Wang[21]의 방법과 달리 집중원도우의 크기가 고정되어 있지 않고 객체의 크기에 따라

유동적으로 변하기 때문에 좀 더 정확한 객체의 위치를 추정 할 수 있다는 장점이 있다. 그리고 다른 영역과 구별되는 현저한 장소에 무의식적으로 시선이 집중 되는 시각주의 특성을 적용했다는 것이 여러 객체 추출 방법론과 구별되는 본 시스템만의 특징이라고 볼 수 있다. 하지만 객관적으로 평가할 만한 방법론이 아직까지 제시된 바가 없어 객체 추출의 결과를 다른 방법론과 비교하여 평가하기 어렵다는 문제점이 있다. 또한, 복잡한 배경일 경우 배경의 윤곽선과 움직이는 객체의 윤곽선이 분리되지 않아 원하는 객체추출 결과를 얻지 못하는 단점이 있다.

본 알고리즘은 워터마킹, 영상검색, MPEG-4 등에 적용할 수 있으며 또한, 감시시스템(surveillance system)이나 제스처 인식(gesture recognition)에도 적용해 볼 수 있을 것이다. 하지만 본 논문에서 제안한 알고리즘을 실생활에 적용하기 위해서는 몇 가지 문제점이 존재한다. 일상적인 생활에서 촬영하는 대부분의 동영상은

카메라가 고정되어 있기 보다는 카메라가 움직임을 가지는 경우가 많다. 따라서 카메라가 이동할 때 중요객체를 추출하는 방법론 개발 또한 향후 연구방향으로 제시된다. 또한, 객체가 다른 객체에 의해 일시적으로 가려질 경우, 급격한 조명변화가 일어날 경우에 대해서도 고려해 보아야 할 문제이다. 이러한 문제점을 해결할 수 있는 방법들도 향후 흥미로운 연구 주제가 될 것으로 기대된다.

### 참 고 문 헌

- [1] Q. Tian, Y. Wu, and T.S. Huang, "Combine user defined region-of-interest and spatial layout for image retrieval," *Proceedings of International Conference on Image Processing*, Vol.3. pp.746-749, 2000.
- [2] S. Kim, S. Park, and M. Kim, "Central Object Extraction for Object-based Image Retrieval," *Proceedings of International Conference on Image and Video Retrieval*, pp. 39-49, 2003.
- [3] T. Meier and K. N. Ngan, "Video Segmentation for Content-based Coding," *IEEE Transaction on Circuits and Systems Video Technology*, Vol. 9, No. 8, pp. 1190-1203, 1999.
- [4] Y. Tsaig and A. Averbuch, "Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach," *IEEE Transaction on Circuits and Systems Video Technology*, Vol. 12, No. 7, pp. 597-612, 2002.
- [5] J. Shi and J. Malik, "Motion Segmentation and Tracking Using Normalized Cuts," *Proceedings of International Conference on Computer Vision*, pp. 1154-1160, 1998.
- [6] G. Adiv, "Inherent Ambiguities in Recovering 3-D Motion and Structure from a Noisy Flow Field," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 477-489, 1989.
- [7] G. D. Borshukov, G. Bozdagi, Y. Altunbasak, and A. M. Tekalp, "Motion Segmentation by Multistage Affine Classification," *IEEE Transaction on Image Processing*, Vol. 6, No. 11, pp. 1591-1594, 1997.
- [8] P. Bouthemy and E. Francois, "Motion Segmentation and Qualitative Dynamic Scene Analysis from an Image Sequence," *International Journal of Computer Vision*, Vol. 10, No. 2, pp. 157-182, 1993.
- [9] J. Y. A. Wang and E. H. Adelson, "Representing Moving Images with Layers," *IEEE Transaction on Image Processing*, Vol. 3, No. 5, pp. 625-638, 1994.
- [10] M. M. Chang, A. M. Tekalp, and M. I. Sezan, "Simultaneous Motion Estimation and Segmentation," *IEEE Transaction on Image Processing*, Vol. 6, No. 9, pp. 1326-1333, 1997.
- [11] C. Stiller, "Object-based Estimation of Dense Motion Fields," *IEEE Transaction on Image Processing*, Vol. 6, No. 2, pp. 234-250, 1997.
- [12] E. Tuncel and L. Onural, "Utilization of the Recursive Shortest Spanning Tree Algorithm for Video Object Segmentation by 2-D Affine Motion Modeling," *IEEE Transaction on Circuits and Systems Video Technology*, Vol. 10, No. 5, pp. 776-781, 2000.
- [13] S. R. Sternberg, "Grayscale Morphology," *Computer Vision, Graphics, and Image Processing*, Vol. 35, No. 3, pp. 333-355, 1996.
- [14] D. Wang, "Unsupervised Video Segmentation based on Watersheds and Temporal Tracking," *IEEE Transaction on Circuits and Systems Video Technology*, Vol. 8, No. 5, pp. 539-546, 1998.
- [15] J. G. Choi, S.W. Lee, and S. D. Kim, "Spatio-temporal Video Segmentation Using a Joint Similarity Measure," *IEEE Transaction on Circuits and Systems Video Technology*, Vol. 7, No 2, pp. 279-286, 1997.
- [16] P. Salembier, P. Brigger, J. R. Casas, and M. Pardas, "Morphological Operators for Image and Video Compression," *IEEE Transaction on Image Processing*, Vol. 5, No. 6, pp. 881-898, 1996.
- [17] 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.
- [18] E. Loupias, and N. Sebe, "Wavelet-based Salient Points for Image Retrieval," *Research Report RR 99.11, RFV-INSA Lyon*, 1999.
- [19] B. C. Ko, and H. Byun, "Region-based Image Retrieval: A New Method for Extraction of Salient Regions and Learning of Importance Scores," *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 17, No. 8, pp. 1349-1367, 2003.
- [20] <http://www.cipr.rpi.edu/resource/sequences/index.html>
- [21] J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLicity: Semantics-Sensitive Integrated Matching for Picture Libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 9, pp. 947-963, 2001.



곽수영

2003년 계명대학교 컴퓨터공학과 학사 졸업. 2005년 연세대학교 컴퓨터과학과 석사 졸업. 2005년~현재 연세대 컴퓨터과학과 박사과정. 관심분야는 영상처리, 컴퓨터 비전 및 패턴인식



**고 병 철**

1998년 경기대학교 전자계산학과 졸업 (이학사). 1998년~2000년, 2000년~2004년 연세대학교 대학원 컴퓨터과학과 졸업(공학석사, 공학박사). 2004년 3월~2005년 8월 삼성전자 통신연구소 책임연구원. 2005년 9월~현재 계명대학교 정보통신학부 전임강사. 관심분야는 내용기반 영상검색, 컴퓨터 비전 및 패턴인식



**변 혜 란**

1980년 연세대학교 수학과 학사 졸업  
 1987년 Univ. of Illinois at Chicago, Computer Science M.S. 1993년 Perdue University, Computer Science, Ph.D.  
 1994년 3월~1995년 2월 한림대학교 조교수. 1995년 3월~현재 연세대학교 컴퓨터과학과 교수. 관심분야는 멀티미디어, 컴퓨터비전, 영상처리, 인공지능, 패턴인식