

동영상 내용 분석을 위한 관심 객체 추출

박소정[†], 김민환^{††}

요약

동영상에서의 관심 객체를 추출하는 것은 비디오 내용 분석과 비디오 검색 및 압축의 성능을 개선시키는데 큰 역할을 한다. 관심 객체는 단순히 사람 눈의 시선을 끄는 대상물이 아니라 내용전개의 중심이 되거나 제작자가 표현하려고 하는 핵심 객체를 의미한다. 이러한 관심 객체는 움직이는 객체뿐만 아니라 정지해 있는 객체도 될 수 있으나, 사람의 관심을 절차적으로 표현하는 것이 어렵기 때문에 관심 객체를 명확하게 정의하기가 곤란하다. 이에, 본 논문에서는 동영상 샷에서의 움직이는 객체의 위치, 크기, 움직임 패턴의 변화에 대한 조건을 정의하여 필터링에 의해 사람의 관심을 끄는 움직임 관심 객체를 추출하는 방법을 제시하고, 아울러 동영상 샷에서 정지되어 있는 객체에 대해서도 컬러/텍스처 특이성, 위치, 크기, 출현 빈도 등에 대한 조건을 정의하여 정지 관심 객체도 추출할 수 있는 방법을 제안한다. 제안한 방법을 50개의 동영상 샷에 대하여 실험한 결과, 사람이 선정한 움직임 및 정지 관심 객체를 84% 정도 추출할 수 있음을 확인할 수 있었다.

Segmentation of Objects of Interest for Video Content Analysis

Sojung Park[†], Minhwan Kim^{††}

ABSTRACT

Video objects of interest play an important role in representing the video content and are useful for improving the performance of video retrieval and compression. The objects of interest may be a main object in describing contents of a video shot or a core object that a video producer wants to represent in the video shot. We know that any object attracting one's eye much in the video shot may not be an object of interest and a non-moving object may be an object of interest as well as a moving one. However it is not easy to define an object of interest clearly, because procedural description of human interest is difficult. In this paper, a set of four filtering conditions for extracting moving objects of interest is suggested, which is defined by considering variation of location, size, and moving pattern of moving objects in a video shot. Non-moving objects of interest are also defined as another set of four extracting conditions that are related to saliency of color/texture, location, size, and occurrence frequency of static objects in a video shot. On a test with 50 video shots, the segmentation method based on the two sets of conditions could extract the moving and non-moving objects of interest chosen manually on accuracy of 84%.

Key words: Object of Interest(관심 객체), Visual Attention(시선), Video Content Analysis(동영상 내용 분석), Content Based Video Retrieval(내용기반 비디오 검색)

* 교신저자(Corresponding Author) : 김민환, 주소 : 부산
시 금정구 장전동 산 30(609-735), 전화 : 051)510-2423,
FAX : 051)517-2431, E-mail : mhkim@pnu.kr

접수일 : 2007년 1월 11일, 완료일 : 2007년 6월 21일

[†] 준희원, 부산대학교 컴퓨터공학과

(E-mail : sjpark@pnu.kr)

^{††} 종신회원, 부산대학교 컴퓨터공학과

* 이 논문은 부산대학교 자유과제 학술연구비(2년)에 의하여 연구되었음

1. 서 론

일반적으로 사람은 정지영상이나 동영상에서 배경보다는 어떤 관심 대상을 중심으로 내용을 이해하는 것으로 알려져 있다. 정지영상에서의 관심 대상은 촬영자가 표현하고자 하는 객체를 의미하며, 동영상에서의 관심 대상은 내용 전개에 있어서 중심이 되는 객체를 의미한다고 볼 수 있다. 사람이나 물체와 같이 특정 객체 형태로 나타나는 이러한 관심 대상을 의미적 객체(semantic object) 또는 관심 객체(object of interest)라고도 한다. 관심 객체는 기존의 영상 처리 및 인식 분야에서 사용해 오던 관심 영역(region of interest)이라는 용어와 구분할 필요가 있다. 관심 영역은 응용 목적에 적합한 영역을 의미하는 것으로서, 동영상에서의 워터마킹[1], 압축[2], 통신[3], 원격 화상회의[4] 등을 처리하기에 적합한 영역 등을 예로 들 수 있다. 한편, 자연 경관을 찍은 정지영상이나 어떤 주제의 내용을 전개하기 위한 도입부로서 전체 배경 장면을 원거리에서 촬영한 동영상 샷(shot)에서는 특별한 관심 객체가 없다고 볼 수 있다. 또한 동영상 샷에서 주인공 및 중심인물을 중심으로 한 내용 전개를 보조하기 위한 보조출연자 및 소품 등도 관심 객체라고 볼 수 없다.

관심 객체는 영상물에서 보다 의미 있는 정보 및 내용을 분석하는데 유용하게 활용할 수 있다. 예를 들어, 정지영상에 대한 내용기반 검색 분야에서 기준의 컬러 분포, 모양(shape), 텍스처 등과 같은 저수준의 특징 값만으로는 사용자의 검색 의도나 생각을 잘 표현할 수 없어 검색 성능을 향상시키는데 한계가 있음이 지적되어 왔는데, 소위 개념 차이(semantic gap)라고 불리는 이러한 한계를 극복하기 위한 방법으로서 관심 객체를 추출하여 검색에 활용할 수 있다 [5]. 동영상에서는 특정 대상물(어린이 통학버스, 공격용 헬리콥터, 유명 배우 등)이 등장하는 비디오를 검색하거나 또는 주인공 및 핵심 등장인물을 위주로 비디오 내용을 요약(summarization)하려는 경우에 관심 객체를 추출하여 유용하게 활용할 수 있다.

그러나 주어진 영상물에서 관심 객체를 자동으로 추출하는 것은 매우 어렵다. 그 원인은 관심 객체에 대한 명확한 정의의 어려움과 정확한 추출의 어려움으로부터 기인한다고 볼 수 있다. 비록 관심 객체를 영상물 제작자가 표현하려고 의도하거나 또는 내용

전개에 있어서 중심이 되는 객체로 정의할 수 있지만, 사전 정보가 없는 상황에서 주어진 영상물에서의 관심 객체를 추출하기 위해서는 관람자 입장에서 결정해야 한다. 즉, 관람자의 생각이나 의지, 사전 지식 및 경험, 전후 상황(context) 정보 등이 종합적으로 결부되어 주관적으로 결정된다고 볼 수 있다. 따라서 사람의 고수준 인지 과정 및 주관적 판단 과정을 소수의 특징(feature)을 이용하여 절차적으로 나타내어 결정해야 하므로 매우 어려운 일이다. 한편, 영상물에서의 관심 객체는 관념적인 대상(사랑, 정의, 죽음 등)이 아니라 특정 형태를 갖는 객체(사람, 자동차, 얼룩말 등)를 나타내는 것으로서, 영상물에서 여러 개의 특징 영역으로 분할(segmentation)되더라도 하나로 묶어 추출해야 하는 어려움이 있다. 이러한 정의 및 추출의 어려움으로 인해 사용자 또는 관람자가 수작업으로 분할된 영역을 선택하여 관심 객체를 지정[6-8]해 주거나 마우스 등과 같은 입력장치로 관심 객체의 개략적인 범위를 지정[9]해 주는 방식을 이용하기도 하였다.

정지영상에서 관심 객체를 정의하는 것은 동영상에 비해 상대적으로 수월하다고 볼 수 있다. 이것은 정지영상 촬영자가 표현하고자 하는 대상물을 영상 중앙 부분에 상대적으로 크게 나타내는 경향이 있으며, 또한 이 대상물은 주위 배경과 식별되는 컬러 및 텍스처를 갖는 경향도 있기 때문이다[5]. 정지영상에서의 관심 객체를 자동으로 결정하기 위해 이루어져 왔던 접근방법으로는 먼저 배경과 같이 명확한 비관심 영역을 제거하는 방식[10-12]이 시도되어 왔으며, 근래에는 인간의 시선(visual attention) 처리 방식을 모의(simulation)하여 구한 시선 집중도(degree of visual attention) 정보를 기반으로 관심 객체 영역을 추출하려는 방식[13-15]이 많이 시도되었다. 특히 Itti[13]는 인간시각시스템의 초기 처리 단계에서의 신경 구조 및 동작 특성을 감안하여 영상물에서의 컬러, 명암, 오리엔테이션 등의 대조(contrast)에 기반한 지역적 특이성(saliency map)을 구해 시선 정보를 양적으로 표현하고자 하였는데, 이 방법은 최근에도 정지영상[16]뿐만 아니라 동영상[17,18]에서 관심 객체 영역의 분석 및 추출에 많이 활용되고 있다. 한편, 관심 객체의 영역을 반복적인 에너지 최소화 방법을 적용하여 보다 정확하게 결정하기 위한 연구[19]도 이루어졌다.

반면에, 동영상에서 관심 객체를 정의하는 것은 정지영상에 비해 상대적으로 어렵다고 볼 수 있다. 다양한 촬영 및 편집 기법[20,21]에 의해 프레임 내에서 관심 객체의 움직임, 위치, 크기 등의 변화가 많기 때문이다. 기존의 보안감시, 이동물체 추적 및 인식, 비디오 이벤트(event) 분석, 비디오 장르(genre) 분류 등과 같은 응용 분야에서는 움직이는 대상을 영역을 주요 관심 대상 영역으로 정의하여 사용해 왔다. 일반적으로, 동영상에서의 움직임 객체는 사람의 시선을 많이 집중시키는 것으로서 실제적인 관심 객체가 될 가능성이 높다. 그러나 모든 움직임 객체를 관심 객체로 분류하기에는 무리가 있다. 예를 들어, 크기가 매우 작은 움직임 객체, 일시적으로 나타났다가 사라지는 객체, 프레임의 가장자리에서 주로 움직이는 객체, 프레임의 한쪽 가장자리에서 다른 쪽으로 지나가듯이 움직이는 객체 등은 관심 객체라고 볼 수 없다. 이에, 사용자 정의에 의해 동영상에서 보다 의미 있는 즉 관심 객체를 추출하여 사용하려는 시도가 이루어져 왔다[9,22]. Zhong[9]은 동영상에서 사용자에 의해 정의된 관심 객체를 정확하게 추출하면서 추적할 수 있는 방법을 제시하였다. Porikli[22]는 비디오 시퀀스에 해당하는 3차원 시공간(spatio-temporal space)에서 유사한 속성을 갖는 화소들을 묶어 기본 볼륨(volume)을 형성한 후, 이 기본 볼륨에 대한 계층적 그룹핑(hierarchical clustering) 과정에서 특정 관심 객체에 대한 사용자의 사전 지식을 적용하여 해당 관심 객체를 추출할 수 있다고 하였다. 한편, 동영상에서 어떤 정지되어 있는 객체가 비록 움직임 객체보다 시선은 덜 끌지만 관심 객체가 되는 경우가 있다. 예를 들어, 그림 1(a)의 자동차 전시회 동영상에서는 중앙의 큰 정지 객체인 자동차를 관심 객체라고 볼 수 있으며, 그림 1(b)의 어떤 야외 모임 동영상에서는 뒤에서 다가오고 있는 사람들뿐만 아니라 중앙에 서 있는 소녀도 관심 객체라고 볼 수 있다. Han[17]은 이러한 정지 객체가 의미 있는 객체가 될 수 있음을 주장하면서, 움직임 및 정지 관심 객체를 자동으로 추출할 수 있는 통합된 구조(unified framework)를 제시하였다. 여기에서는 정지 관심 객체를 추출하기 위하여 시선 집중도 정보[13]에다가 프레임 중앙에 보다 높은 가중치를 부여하기 위한 가우시안 분포 함수를 적용하는 방식을 사용하였다.

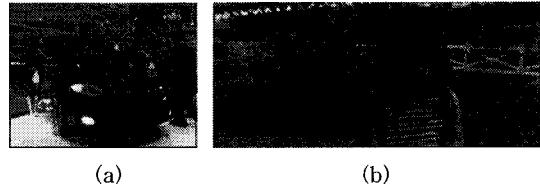


그림 1. 동영상에서의 정지 관심 객체 예 ((a) 자동차, (b) 소녀)

동영상에서 관심 객체를 자동으로 추출하기 위해선 프레임 시퀀스 전체에 대한 분석을 통해 대상을 이 연속으로 나타나는 프레임 비율과 더불어 대상을 의 크기 변화, 위치 변화, 움직임 패턴, 지역적 특이성 등의 정보를 종합하여 활용할 필요가 있다. 기존의 동영상 관련 연구에서도 움직임 객체의 추출 성능을 높이기 위해 이러한 정보들을 부분적으로 활용해 왔다. 예를 들어 Zhong[23]은 대상을 출현 프레임 비율이 낮거나 대상을 크기가 작은 것을 제거함으로써 잠시 지나치는 대상을이나 잡영 같은 것을 제거하고자 하였다. 한편, 기존 동영상 연구에서 관심 영역을 결정하기 위해 움직임 정보와 더불어 시선 정보를 많이 활용해 왔다[17,18]. 그러나 시선 정보는 사람의 의지와 상관없이 눈이 움직여 집중하게 되는 현상을 양적으로 표현한 것으로서, 사람의 생각이나 의지, 사전 지식 및 경험, 전후 상황(context) 정보 등과 깊게 연관된 사람의 관심을 표현하기에는 다소 미흡해 보인다. 또한 시선 정보는 한 프레임 내에서의 컬러, 명암, 오리엔테이션 등의 대조(contrast)에 기반한 지역적 특이성을 나타낸 것으로, 전체 프레임 시퀀스에 걸쳐 나타나는 대상을의 특성을 표현하는데도 미흡하다고 볼 수 있다.

본 논문에서는, 동영상 전체 프레임에서의 대상을의 크기 변화, 위치 변화, 움직임 패턴, 배경 대비 중앙의 특이성 등의 정보를 활용하여 움직임 및 정지 관심 객체를 결정하는 방법을 제시하고자 한다. 먼저 동영상에서의 움직임 객체가 관심의 대상이 되기 위해서 갖추어야 할 크기 변화, 위치 변화, 움직임 패턴 등에 대한 조건을 정의한다. 이 조건은 기존의 움직임 객체의 추출 성능을 높이기 위해 사용하던 조건과 개념적으로 일부 유사할 수도 있으나, 관심 객체를 추출한다는 목적에서 차이가 있으므로 구현 내용이 다르다. 구현 측면에서 보면, 보다 강화된 조건을 활용하여 관심 객체를 추출한다고 볼 수 있다. 한편,

동영상에서의 정지 관심 객체를 추출하기 위해서도 어떤 정지 영역이 만족해야 할 크기 변화, 위치 변화, 움직임 패턴 등에 대한 조건을 제시한다. 각 프레임에서 어떤 정지 영역을 추출하기 위해서는 시선 정보 대신에 배경 대비 중앙 부분의 컬러/텍스처 특이성 정보[5]를 사용한다.

2장에서 동영상에서의 움직임 및 정지 관심 객체를 추출하기에 적합한 조건에 대하여 정의하고, 3장에서는 이러한 정의에 따라 실제로 관심 객체를 추출하는 방법을 다룬다. 4장에서는 50개의 다양한 동영상 샷에 대한 실험결과를 보이고 문제점 및 개선사항을 논한다.

2. 동영상에서의 관심 객체 정의

동영상에서의 관심 객체는 단순히 시선을 끄는 움직임 객체라기보다는 내용 전개와 관련된 것으로서 영상물 제작자의 표현 의도와 밀접한 관계가 있는 것이라고 볼 수 있다. 한편 동영상을 관람하는 사람 입장에서 보면, 관심 객체 여부에 관한 판단은 사람의 고수준 인지과정을 통해 이루어지며, 또한 개인의 주관적인 판단에 의해 결과가 달라질 수 있다는 측면도 있어, 동영상에서의 관심 객체를 논리적으로나 수치적으로 명확히 정의하는 것은 매우 어렵다.

이에 본 논문에서는, 샷(shot) 단위의 동영상 시퀀스에 대하여 사람들이 공통적으로 선정하는 관심 객체에 대한 실험적 관측을 통해, 공학적인 측면에서 관심 객체를 효과적으로 추출하는데 유용한 조건들을 정의하고자 한다. 주로 전체 프레임에 걸쳐 나타나는 대상물의 크기, 위치, 움직임 변화와 배경과 대비한 컬러/텍스처 특이성 등에 관한 정보를 활용한다. 이때 동영상 샷은 인물의 허리 이상이 보이는 미디엄 샷, 인물의 발끝부터 머리끝까지 보이는 롱 샷 등을 대상으로 하며, 너무 많이 클로즈업 되었거나 너무 멀리서 촬영한 샷은 대상에서 배제한다. 수십 명의 인물이 보이는 군중 샷에서는 특정한 관심 객체를 결정하기 어려우므로 대상에서 배제한다. 따라서 축구나 농구에서와 같이 전체 선수들의 움직임이 나타나는 스포츠 동영상도 배제된다. 관심 객체는 움직임 관심 객체와 정지 관심 객체로 구분할 수 있는데 [17], 서로 다른 속성을 가지고 있으므로 각각 별도로 정의한다.

2.1 움직임 관심 객체 정의

동영상에서의 움직임 관심 객체는 다음과 같은 조건을 모두 만족시키는 대상물로 정의할 수 있다.

- (M1) 배경과 구분되는 움직임을 가져야 한다.
- (M2) 여러 프레임에 걸쳐 자주 나타나야 한다.
- (M3) 영역의 크기가 지속적으로 매우 작게 나타나지 않아야 한다.
- (M4) 프레임 가장자리에 지속적으로 너무 치우쳐 나타나지 않아야 한다.

동영상에서의 움직임 객체는 배경과 구별되는 움직임을 가질 때에 추출할 수 있다. 이 조건은 대부분의 움직임 객체 추출 방법에서도 기본적으로 사용하는 것이다.

어떤 움직임 객체가 사람의 관심을 끄는 관심 객체로 판정받기 위해서는 일정 비율 이상의 프레임에서 지속적으로 나타날 필요가 있다. 잠시 나타났다가 사라지는 움직임 객체는 시선은 끌지 모르나 관심의 대상이 되지는 못한다. Zhong[23]은 움직임 객체의 추출 성능을 높이기 위한 방안으로서 객체 영역 그룹핑의 부정확성을 보완하고 잠시 정지했다가 다시 움직이는 객체도 정확하게 추출하기 위해 이러한 비율을 사용하기도 하였다.

움직임 관심 객체는 여러 프레임에 걸쳐 커지거나 작아질 수 있으나, 크기가 작은 상태로 계속 나타나는 안 된다. 영상을 제작자는 어떤 장면에서의 주인공이나 중요 대상물이 프레임에서 큰 영역을 차지하도록 제작하는 것이 일반적이며, 또한 관람자 입장에서 볼 때도 지속적으로 작게 나타나는 물체는 비록 움직임이 있더라도 사람이 관심 대상으로 인지하지 못하기 때문이다. 기존의 움직임 객체 추출 방법[23]에서는 프레임 내에서의 잡영을 제거하기 위해 모폴로지 연산 등을 통해 크기 필터링을 하기도 하였다.

움직임 관심 객체는 프레임 가장자리에 너무 치우쳐서 계속적으로 나타나지 않아야 한다. 영상을 제작자는 주인공이나 중요 대상물을 영상의 중앙 근처에 위치하도록 제작하는 것이 일반적이다. 움직임 관심 객체가 프레임 가장자리로부터 나타나거나 프레임 가장자리로 사라질 경우에는 프레임 가장자리에 짧은 시간 동안 나타날 수도 있다.

샷 단위의 동영상 시퀀스에서 움직임 관심 객체를 추출하기 위해서는, 먼저 각 프레임에서의 움직임 영

역을 구한 후 전체 프레임에 걸쳐 클러스터링 한다. 이후에 각 클러스터에 대하여 위의 네 조건을 만족시키지 못하는 것을 제거하는 방식을 통해 움직임 관심 객체를 결정하게 된다. 이러한 필터링을 통한 제거 방식은 기존의 움직임 객체 추출 방법에서도 움직임 객체의 추출 성능을 높이기 위해 유사한 조건을 부분적으로 사용하였으나, 본 논문에서는 움직임 관심 객체를 추출하기 위한 목적에서 보다 강화된 조건을 활용한다고 볼 수 있다. 실제 구현과정에서도, 클러스터 형성이 용이한 Zhong[23] 방법에 의해 추출된 움직임 객체들에 대하여 위 네 조건의 동시에 만족여부를 조사하는 방식을 사용한다.

2.2 정지 관심 객체 정의

동영상에서의 정지 관심 객체는 한 프레임 내에서의 지역적 시선 집중도[13] 기반으로 정지된 대상물을 탐색하는 방식[17]으로 결정하기 보다는, 전체 프레임에 걸쳐 대상물의 크기, 위치, 움직임 변화와 배경과 대비한 컬러/텍스처 특이성 등에 관한 정보를 종합 분석하여 다음과 같은 조건을 모두 만족시키는 것으로 정의할 필요가 있다.

- (S1) 움직임 객체가 자주 등장하지 않는 부분 영역에 나타나야 한다.
- (S2) 주변과 식별되는 컬러 및 텍스처 특징을 가져야 한다.
- (S3) 전체 프레임에 걸쳐 지속적으로 나타나야 한다.
- (S4) 비교적 크고 고정된 크기를 가져야 한다.

정지 관심 객체는 일반적으로 각 프레임에서 움직임 객체가 자주 등장하지 않는 영역에 고정된 위치에서 나타난다. 영상을 제작자는 자신이 관심 있는 정지 객체가 움직이는 객체로 인해 오랫동안 가려지는 것을 원하지 않기 때문이다. 만일 어떤 정지 객체가 움직임 객체로 인해 자주 보이지 않는다면 그 객체는 배경의 일부분일 가능성이 크다. 그러므로 정지 관심 객체는 프레임에서 움직임이 상대적으로 적은 부분 영역에서 찾아 추출할 필요가 있다.

어떤 한 프레임 내에서 사전지식 없이 관심영역을 추출해 내는 것은 매우 어려운 과제이다. Han[17]은 컬러, 명암, 오리엔테이션 등의 대조(contrast)에 기반한 지역적 시선 집중도[13]를 이용하여 프레임

중앙 부분의 특이 영역을 추출하려 하였다. 그러나 지역적 시선 집중도는 사람의 시선을 표현하는데 유용하지만 사람의 관심을 표현하는 데는 다소 부적합함을 실험을 통해 확인할 수 있었다. 한편, Kim[5]은 정지영상 중앙에서의 주변과 식별되는 컬러 및 텍스처 분포 정보를 이용하여 관심 객체를 구분해내고자 하였는데, 프레임의 중앙 부분과 주변(배경) 부분 간의 전체적 분포 정보를 활용하므로 본 논문에서 추출하려는 관심영역을 결정하는데 보다 유용하였다. 이에, 조건 (S2)에서는 정지 관심 객체의 컬러 및 텍스처 분포가 주변 부분과 차별화 되도록 정의하였다.

정지 관심 객체는 전체 프레임에 걸쳐 지속적으로 나타나는 경향이 있다. 이것은 영상을 제작자가 어떤 정지된 객체를 해당 샷에서의 관심 객체로 표현하고자 할 때에 그 객체를 프레임 중앙 부분에 지속적으로 나타나도록 촬영하는 경향이 있기 때문이다.

한편, 정지 관심 객체는 일반적으로 움직임 관심 객체보다 크며 전체 프레임에 걸쳐 고정적인 크기를 가지는 경향이 있다. 이것 또한 영상을 제작자가 정지 관심 객체를 강조하기 위해 사용하는 방식이라고 볼 수 있다.

3. 동영상에서의 관심 객체 추출 방법

동영상에서 관심 객체를 추출하기 위해서는 전체 프레임에서의 객체의 크기 변화, 위치 변화, 움직임 패턴 등의 정보를 종합하여 활용해야 하므로, 그림 2(a)에서와 같이 어떤 한 동영상 샷을 3차원 시공간 V 로 표현할 필요가 있다. 이러한 시공간에서 각 프레임에서의 움직임 객체 영역을 묶어 그림 2(b)에서와 같이 클러스터(Cluster) C_m 으로 표현 할 수 있다. 또한 2.1절의 동영상 관심 객체에 대한 조건 (M4)에서의 프레임 가장자리를 부분을 그림 2(b)와 같이 가장자리 시공간 V_b 로 표현할 수 있는데, 본 논문에서는 한 프레임의 폭 및 높이의 각각 0.9배에 중앙 부분을 제외한 나머지 부분에 대응하는 시공간으로 정의하였다. 한편, 정지 관심 객체를 결정하기 위해서는 2.2절의 조건 (S2)에서 프레임의 중앙 부분과 주변 부분에 대한 정의가 필요한 데, 개념적으로 프레임의 중앙 부분을 시공간 V 에서 표현하면 그림 2(c)에서와 같이 중앙 시공간 V_s 로 나타낼 수 있다. 정지영상에

서 중심 객체를 찾기 위한 Kim[5]의 연구에서는 프레임 중앙 부분을 프레임 너비와 높이의 각각 0.5배에 해당하는 중앙의 사각형 부분으로 정의하였는데, 본 연구에서는 V_s 의 절단면에 해당하는 프레임 중앙 부분을 프레임 너비의 0.6배와 프레임 높이의 0.5배로 정의하였다. 이것은 동영상에서의 정지 관심 객체가 보다 옆으로 넓은 범위에 걸쳐 나타날 수 있음을 감안한 것이다. 한편, 각 프레임에서 조건 (S2)를 만족시키는 정지 관심영역을 묶어 그림 2(c)에서와 같이 클러스터 C_s 로 나타낼 수 있다.

3.1 움직임 관심 객체의 추출

움직임 관심 객체를 추출하기 위해서는 먼저 조건 (M1)을 만족시키는 움직임 객체를 결정하여야 한다. 본 논문에서는 움직임 객체 추출 방법을 별도로 개발하는 대신에 Zhong[23]의 방법을 활용한다. 이 방법에서는 먼저 각 프레임에서 분할된 영역 단위로 움직임 정보를 구하여 유사한 움직임 영역들을 병합한 후 전경(foreground) 및 배경 레이어(layer)들로 분류한다. 다음에는, 전경 레이어를 대상으로 프레임 간의 움직임 기반의 유사 레이어를 병합하여 시간적 (temporal) 레이어를 형성한다. 마지막으로, 시간적 레이어에 대한 상대적 출현 길이(duration), 프레임 간의 중심 및 크기 변화, 프레임 내에서의 잡영 제거 등을 통해 유용한 움직임 객체에 대한 시간적 레이어를 최종 결정한다. 이러한 최종 시간적 레이어는 본 논문에서의 움직임 객체 클러스터(C_m)에 해당하는 것으로서, 2.1절에서 제시한 조건 (M2)~(M4)에 대한 만족 여부를 조사하는데 매우 유용하게 활용할 수 있다.

3.1.1 움직임 객체 클러스터의 길이 필터링

움직임 객체 클러스터 C_m 의 길이를 $Length(C_m)$, 주어진 동영상 샷의 프레임 수를 N 이라 하면, 클러스터 길이의 비율 $Length_Ratio_Moving(C_m)$ 은 식 (1)과 같이 정의된다. 만약 이 비율이 실험적으로 구한 임계값 T_{LM} (=0.5)보다 작다면, 조건 (M2)를 만족시키지 못하는 것으로 결정하였다. 즉, 움직임 관심 객체가 되기 위해서는 동영상 샷의 절반 이상에 걸쳐 나타나야 한다는 것을 의미한다.

$$Length_Ratio_Moving(C_m) = \frac{Length(C_m)}{N} \quad (1)$$

3.1.2 움직임 객체 클러스터의 크기 필터링

조건 (M3)은 움직임 객체 클러스터의 체적 $Volume(C_m)$ 을 이용하여 판단 가능하다. 그림 2(a)의 3차원 시공간 V 는 주어진 동영상 샷에 따라 변하므로 $Volume(C_m)$ 은 정규화 하여 사용할 필요가 있다. 정규화된 체적 $Size_Ratio_Moving(C_m)$ 은 식 (2)와 같이 정의되며, 이 정규화된 체적이 실험적으로 결정한 임계값 T_{SM} (= 0.01)보다 작으면 조건 (M3)을 만족시키지 못하는 것으로 결정하였다. 이 임계값은 움직임 객체 클러스터의 길이 비율과 프레임 간 움직임 객체 영역의 크기 변화 등을 감안하여 매우 낮게 설정할 필요가 있었다.

$$Size_Ratio_Moving(C_m) = \frac{Volume(C_m)}{Volume(V)} \quad (2)$$

3.1.3 움직임 객체 클러스터의 위치 필터링

조건 (M4)는 움직임 객체 클러스터의 체적 C_m 과

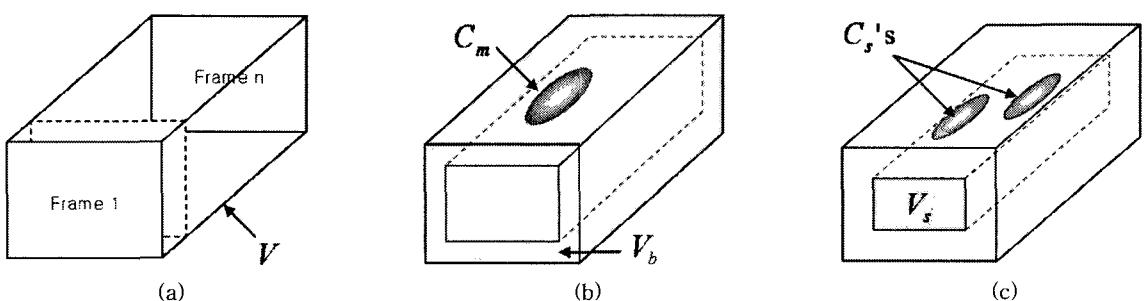


그림 2. (a) 동영상 샷의 3차원 시공간(V) 표현, (b) 움직임 객체 클러스터(C_m) 및 프레임 가장자리 시공간(V_b), (c) 정지 관심영역의 탐색을 위한 중앙 시공간(V_s) 및 정지 객체 클러스터(C_s)

그림 2(b)의 가장자리 시공간 V_b 가 교차하는 체적이 매우 작아야 함을 의미한다. 클러스터 체적(*Volume* (C_m))에 대한 교차된 체적의 비율 *Border_Ratio_Moving*(C_m)은 식 (3)과 같으며, 실험적으로 구한 임계값 T_{BM} (= 0.5)보다 크면 C_m 에 대응되는 움직임 객체는 걸려지도록 하였다. 이 임계값은 개략적으로 어떤 객체의 움직임이 50% 이상 프레임 가장자리에서 일어나는 경우를 의미한다고 볼 수 있다.

$$\text{Border_Ratio_Moving}(C_m) = \frac{\text{Volume}(C_m \cap V_b)}{\text{Volume}(C_m)} \quad (3)$$

3.2 정지 관심 객체의 추출

정지 관심 객체는 조건 (S1)에 의해 움직임 객체가 자주 등장하지 않는 부분 영역에서 찾아서 추출되어야 한다. 먼저 이러한 부분 영역을 그림 2(c)의 중앙 시공간 V_s 에서 움직임 객체의 출연 빈도를 고려하여 결정한다. 다음에는 이 부분 영역을 기반으로 중심 객체(central object) 추출 방법[5]을 활용하여 조건 (S2)를 만족시키는 정지 객체를 추출한 후, 정지 객체 클러스터 C_s 를 구성한다. 만약 어떤 정지 객체 클러스터가 조건 (S3)과 (S4)를 모두 만족시킨다면, 그에 대응하는 정지 객체는 정지 관심 객체로 선택된다.

정지영상에서 중심 객체를 추출하는 Kim[5]의 방법에서는 먼저 컬러 분포의 유사도를 측정하는 한 방법인 코렐로그램(correlogram)[24] 방법을 활용하여 정지영상의 중앙 1/4 영역과 나머지 주변 영역 간의 컬러/텍스처 특이 화소를 결정한 후, JSEG[25] 방법으로 분할된 영역 중에서 특이 화소를 최대로 많이 포함하는 핵심 영역(core region)을 결정하였다. 다음에는 핵심 영역의 이웃한 영역 중에서 핵심 영역과 유사한 컬러 분포를 갖는 것과 배경 영역과 판이하게 다른 컬러 분포를 갖는 것을 합병함으로써 중심 객체를 결정하였다. 본 논문에서는 동영상의 특성상 정지 객체가 가로로 보다 넓은 범위에 걸쳐 나타날 수 있다는 것을 감안하여 프레임의 중앙 부분으로서 프레임의 가로 0.6배, 세로 0.5배에 해당하는 영역을 사용하였으며, 핵심 영역만을 정지 객체로 결정하였다. 이에 따라 정지 객체의 경계가 Kim[5]방법에 비해 상대적으로 부정확할 수 있다.

3.2.1 중심 움직임 계산

먼저 MO_f 를 프레임 f 에서 움직임 객체의 위치를 나타내는 움직임 발생 배열(Motion occurrence array)이라고 하자. 만약 프레임 f 의 픽셀 (i,j) 에 움직임 객체가 존재한다면 $MO_f(i,j)$ 를 1로 정하며 그렇지 않으면 0으로 정한다. 그러면 식 (4)에서 보듯이, 주어진 동영상 샷의 프레임 수 N 에 대해 모든 움직임 발생 배열 $MO_f(i,j)$ 를 누적함으로써 각 픽셀의 위치 (i,j) 에서 움직임 빈도 $MF(i,j)$ 를 계산할 수 있다.

$$MF(i,j) = \sum_{f=1}^N MO_f(i,j) \quad (4)$$

$$MO_f(i,j) = \begin{cases} 1, & (i,j) \in \text{moving object} \\ 0, & \text{otherwise} \end{cases}$$

움직임 빈도 $MF(i,j)$ 는 식 (5)에서와 같이 임계값을 통해 $SMF(i,j)$ 로 간략화 할 수 있다. 임계값 T_{SMA} 는 실험을 통해 구해진 값으로 (0.5 x N)을 사용한다. 다음에는 식 (5)에서 보듯이, 정지 객체를 찾는 기본 검색 영역(default searching area, DSA)에서의 움직임 정도 비율로서 중심 움직임 정도(central motion activity) CMA 를 정의할 수 있다. 기본 검색 영역 DSA는 프레임 너비의 0.6배, 높이의 0.5배로 정의되는 프레임 중앙 부분을 의미하며, 그림 2(c)에서 중앙 시공간 V_s 의 절단면에 해당한다. 식 (5)에서의 S_{DSA} 는 DSA의 크기 즉 면적을 의미하는 것으로서, CMA 를 정규화 하는데 사용하였다. DSA에 적은 움직임 객체가 존재한다면, CMA 는 낮아지며 정지 관심 객체가 발생할 가능성은 높아진다고 볼 수 있다.

$$SMF(i,j) = \begin{cases} 1, & MF(i,j) > T_{SMA} \\ 0, & \text{otherwise} \end{cases} \quad (5)$$

$$CMA = \frac{1}{S_{DSA}} \sum_{(i,j) \in DSA} SMF(i,j)$$

두 개의 동영상 샷에 대하여 그림 3(b)에 움직임 빈도 $MF(i,j)$ 와 그림 3(c)에 간략화 된 움직임 빈도 $SMF(i,j)$ 의 예를 보이고 있다. 그림 3(d)에서 프레임 중앙의 큰 사각형은 기본 검색 영역(DSA)을 나타낸다. 권투 동영상 샷의 CMA 는 매우 높지만 자동차 전시회 동영상 샷의 CMA 는 매우 낮음을 알 수 있다. 따라서 자동차 전시회 동영상 샷에 정지 관심 객체가 존재할 가능성이 높음을 알 수 있다.

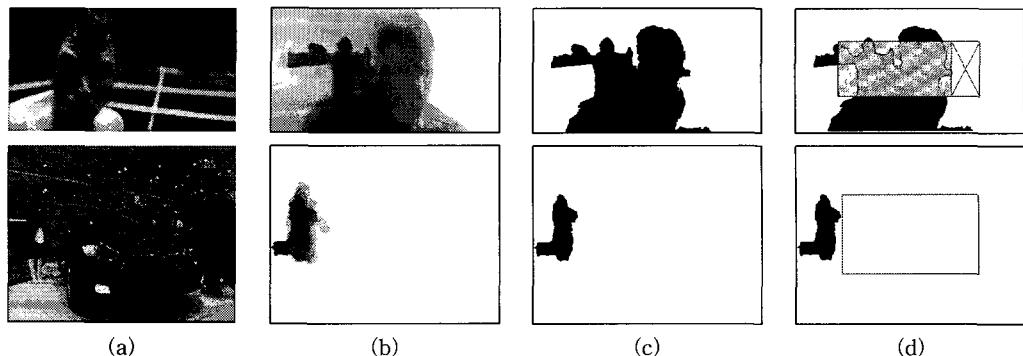


그림 3. 움직임 분석을 위한 두 개의 동영상 샷 예제. (a) 동영상 샷, (b) 움직임 빈도, (c) 간략화 된 움직임 빈도, (d) 기본 검색 영역 내에서의 정지 객체 검색을 위한 부분 영역 선택 예 (권투 동영상: 높은 움직임 정도 및 크기가 작은 영역의 제거에 의해 선택된 부분 영역이 없음. 자동차 전시회 동영상: 기본 검색 영역 전체가 부분 영역으로 선택됨)

3.2.2 정지 객체 검색을 위한 부분 영역 선택

조건 (S1)을 만족시키기 위해서는 정지 객체 검색을 위한 부분 영역이 결정되어야 한다. 먼저 기본 검색 영역 DSA를 열 단위로 차단(mask)한다. 즉, 높은 움직임 정도가 나타나는 열은 전체를 차단한다. 그림 3(d) 권투 동영상의 DSA에서 빛금 친 부분 영역은 움직임 정도가 높아 열 단위로 차단된 것을 나타낸 것이다. 만약 비교적 큰 부분 영역이 차단되지 않았다면, 이 영역을 정지 객체 검색을 위한 부분 영역으로 선택한다. 열 단위로 차단을 할 경우에, DSA의 중앙을 제외한 양 쪽에서 부분 영역이 선택될 수도 있다. 그림 3(d) 권투 동영상에서 'X' 표기된 부분 영역은 비록 차단되지는 않았지만 상대적으로 너무 작아서 정지 객체 검색을 위한 부분 영역으로 부적합하므로 제외되었음을 나타낸 것이다. 반면에, 그림 3(d) 자동차 전시회 동영상에서는 DSA의 모든 영역이 정지 객체 검색을 위한 부분 영역으로 선택되었음을 알 수 있다.

3.2.3 주요 정지 객체 추출 및 정지 객체 클러스터 생성

각각의 프레임에서 조건 (S2)를 만족시키는 정지 객체들을 추출하여야 한다. 본 논문에서는 위에서 구한 정지 객체 검색을 위한 부분 영역에 대하여 중심 객체 추출 방법[5]의 핵심 영역 추출 기법을 사용하여 주요 정지 객체들을 결정한다. 핵심 영역 추출 기법은 코렐로그램[24]을 이용하여 주위 배경 부분과 칼라 및 텍스처가 식별되는 영역을 결정하는 것으로서, 그림 4의 세 번째 행에 보인 바와 같이 주요 정지

객체를 잘 추출할 수 있다. 반면에, 시선과 관련한 지역적 특이성[13]은 컬러, 명암, 오리엔테이션 등의 지역적 대조에 기반하여 구하는 것으로서, 그림 4의 두 번째 행에 보인 바와 같이 한 프레임 내에서 상대적으로 크게 나타나는 주요 정지 객체를 결정하는데 적합하지 못함을 알 수 있다. 정지 객체 클러스터 C_s 는 각 프레임에서 추출한 주요 정지 객체들을 묶어서 시공간 V 내에서 생성한 것이다.

3.2.4 정지 객체 클러스터의 길이 필터링

먼저 동영상 시퀀스를 나타내는 3차원 시공간 V

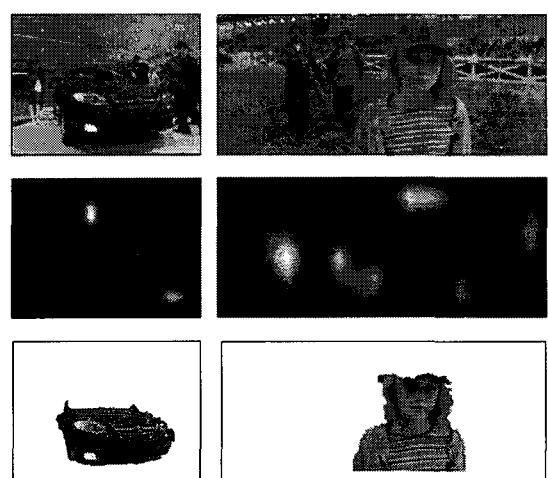


그림 4. 두 동영상 샷에 대한 정지 객체의 추출 예. 두 번째 행은 주어진 동영상에 대한 지역적 특이성(saliency map)[13]을 나타내며, 세 번째 행은 중심 객체 추출 방법[5]을 활용해 추출한 주요 정지 객체를 나타냄.

의 길이 N 과 정지 객체 클러스터 길이 $Length(C_s)$ 간의 비율 $Length_Ratio_Static(C_s)$ 을 식 (6)과 같이 계산한다. 만약 이 비율이 실험적으로 구한 임계값 T_{LS} (= 0.6)보다 작으면, 정지 객체는 조건 (S3)을 만족시키지 못하는 것으로 결정한다.

$$Length_Ratio_Static(C_s) = \frac{Length(C_s)}{N} \quad (6)$$

3.2.5 정지 객체 클러스터의 크기 및 균일성 필터링

정지 객체 클러스터 C_s 의 정규화된 체적 $Size_Ratio_Static(C_s)$ 은 식 (2)에서와 유사하게 정의한다. 만약 식 (7)에서와 같이 정의된 정규화된 체적이 실험적으로 구한 임계값 T_{SS} (= 0.05)보다 작으면, 대응하는 정지 객체는 조건 (S4)를 만족시키지 못하는 것으로 결정한다.

$$Size_Ratio_Static(C_s) = \frac{Volume(C_s)}{Volume(V)} \quad (7)$$

또한, 만약 정지 객체 클러스터 내 정지 객체 크기의 프레임 간 차이의 절대값 평균이 실험적 임계값 T_{VS} (= 0.05)보다 크면, 고정된 크기를 유지하지 못하는 것으로 해석하여 조건 (S4)를 위배하는 것으로 결정한다.

4. 실험결과 및 토의

본 논문에서 제안한 관심 객체 추출 방법을 10편의 영화 및 드라마 동영상 중에서 그림 5에서와 같은 50개 샷(shot)에 대하여 실험하였다. 이때, 동영상 샷은 인물의 허리 이상이 보이는 미디엄 샷, 인물의 발 끝부터 머리끝까지 보이는 롱 샷 등을 대상으로 하였으며, 대상물이 클로즈업 되어 프레임에서 너무 크게 나타나거나 너무 멀리서 촬영함에 따라 대상물이 매우 작게 나타나는 샷은 실험 대상에서 배제하였다. 한편, 스포츠 동영상에서와 같이 많은 대상물(인물)이 나타나는 군중 샷도 특정한 관심 객체를 결정하기 어려우므로 대상에서 배제하였다. 동영상 샷은 해상도가 (320×136) , (320×178) , (320×240) 인 것을 각각 18, 22, 10개 사용하였으며, 동영상 샷의 프레임 수는 약 20 - 200개로서 샷의 정의 속성상 많은 차이가 있었다.

실험결과의 평가를 위해 5명의 일반 실험자에게 의뢰하여 수작업으로 주어진 50개 동영상 샷으로부터 관심 객체를 선정하도록 하였는데, 선정된 것에 대하여 다수결 원칙을 적용한 결과 총 57개의 관심 객체가 결정되었다. 이 중에서 움직임 관심 객체는 그림 5에서 사각형으로 표기한 바와 같이 35개이었으며, 동그라미로 표기된 나머지 22개는 정지 관심 객체이었다.

한편, 제안한 방법으로 움직임 관심 객체를 결정하기 위해서는 먼저 동영상 샷으로부터 움직임 객체를 추출한 후에 2.1절의 (M2)-(M4) 조건에 맞는 가여부를 조사해야 한다. 본 실험에서는 별도로 움직임 객체 추출 알고리즘을 개발하는 대신에 Zhong[23]의 방법을 사용하였다. 그 결과 수작업에 의해 결정된 것 중에서 31개의 움직임 관심 객체를 포함하여 총 44개의 움직임 객체를 추출할 수 있었다. 그러나 수작업으로 결정된 것 중에서 4개의 움직임 관심 객체에 해당하는 것은 추출하지 못하였다.

그림 6에 여섯 개의 동영상 예제 샷에 대한 실험결과를 보이고 있다. 그림 6(b)는 Zhong[23] 방법에 의해 추출된 움직임 객체들을 나타내며, 그림 6의 (c)와 (d)는 제안한 방법에 의해 추출된 움직임 관심 객체와 정지 관심 객체를 각각 나타낸 것이다. 대부분의 모든 관심 객체가 추출되었으며, 무의미한 움직임 객체들도 잘 걸러졌음을 알 수 있다. 그림 6(b)에서 동그라미로 표시된 움직임 객체들은 클러스터 크기 필터링으로 인해 제거된 것이며, 사각형으로 표시된 것들은 클러스터 위치 필터링을 통해 제거된 것을 나타낸다. 삼각형으로 표시된 움직임 객체는 클러스터 길이 필터링에 의해 제거된 것을 의미한다.

그림 7에 잘못 추출된 관심 객체의 대표적인 예를 나타내었다. 그림 7(b)는 Zhong[23] 방법에 의해 추출된 움직임 객체를 나타낸 것이며, 그림 7(c)는 제안 방법에 의해 추출된 정지 관심 객체를 나타낸 것이다. 위쪽의 드라마 동영상 샷에서 중앙의 드라마 주인공은 주어진 50개의 동영상 샷에 대하여 관심 객체를 선정하는데 참여했던 실험자들은 정지 관심 객체로 판정하였다. 그러나 Zhong[23] 방법에서는 계속내리고 있는 비(rain)와 카메라 패닝(panning)으로 인해 움직임이 있는 것으로 판단하여 주인공의 일부분을 움직임 객체로 추출하였다. 이에 따라, 제안한 방법에서는 정지 객체 검색을 위한 부분 영역이 설정



그림 5. 실험에 사용된 50개의 동영상 샷. 사각형과 동그라미로 표기된 객체는 일반 실험자들에 의해 수작업으로 선정된 움직임 관심 객체와 정지 관심 객체를 각각 나타냄

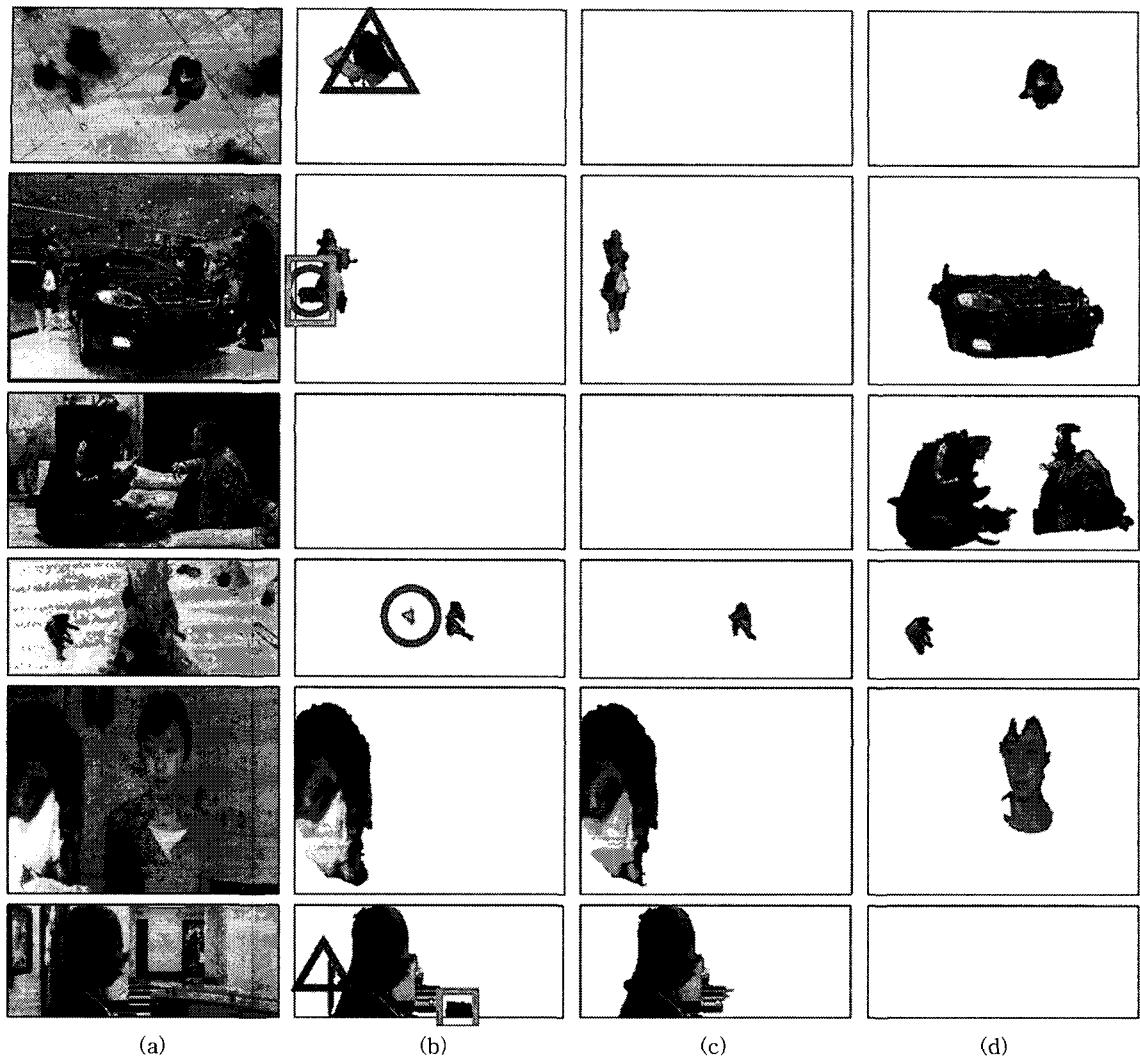


그림 6. 제안한 방법을 통한 관심 객체 추출 예. (a) 여섯 개의 동영상 샷 예, (b) Zhong(23) 방법에 의해 추출된 움직임 객체, (c) 제안 방법에 의해 추출된 움직임 관심 객체, (d) 제안 방법에 의해 추출된 정지 관심 객체

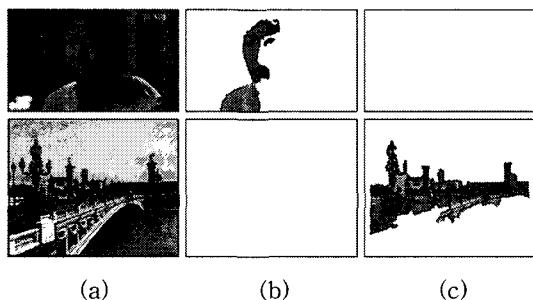


그림 7. 실험에서 잘못 추출된 관심 객체의 예. (a) 두 개의 동영상 샷 예, (b) Zhong(23) 방법에 의해 추출된 움직임 객체, (c) 제안 방법에 의해 추출된 정지 관심 객체

되지 않아 주인공을 정지 관심 객체로 추출하지 못하였다. 한편, 아래쪽의 자연전경 동영상 샷에서는 실험자들이 다리 부분을 관심 객체로 선정하지 않았음에도 불구하고, 제안한 방법에서는 정지 관심 객체로 추출하였다.

표 1은 50개의 동영상 샷에서 수작업으로 추출한 57개의 관심 객체에 대하여 실험결과를 혼동 행렬(Confusion Matrix)[26] 형태로 정리한 것이다. 7개의 잘못된 움직임 관심 객체(False Positive)가 추출되었고, 2개의 잘못된 정지 관심 객체가 추출되었다. 또한 4개의 움직임 관심 객체와 5개의 정지 관심 객체

체를 추출하지 못하였다. 한편, 2.1절에서 제시한 조건 (M2)-(M4)을 이용하여 움직임 관심 객체를 추출하기 위한 필터링 방법의 효용성을 알아보기 위해, Zhong[23] 방법으로 추출한 움직임 객체와 비교하여 표 2에 나타내었다. 제시한 필터링 방법은 움직임 관심 객체보다는 주로 움직임 비관심 객체를 제거하는데 유용함을 알 수 있다. 그러나 움직임 비관심 객체 중에서 46% 정도만을 제거하는 수준이므로, 향후에 보다 지능적인 필터링 조건이 추가로 개발될 필요가 있음을 알 수 있다. 한편, 표 1에서 추출하지 못한 움직임 관심 객체(False Negative) 4개는 Zhong[23] 방법으로 해당하는 움직임 객체가 추출되지 않아 기인되었음도 표 2에서 알 수 있다.

표 3은 실험결과에 대하여 재현율(Recall = TP/(TP+FN)), 정확률(Precision = TP/(TP+FP)), F-측정값(F-measure = $2 \times \text{Recall} \times \text{Precision} / (\text{Recall} + \text{Precision})$) 값을 구한 것이다. 움직임 관심 객체 추출이 정지 관심 객체 추출에 비해 낮은 정확률을 보이는데, 이것은 움직임 객체에 대한 필터링 조건을 보다 지능적인 것으로 개선할 필요가 있음을 의미한다. 반면에, 정지 관심 객체의 재현율은 움직임 관심 객체보다 낮다는 것을 알 수 있는데, 이것은 정지 객

표 1. 동영상 장면 50개에서 수작업으로 선정한 57 (35, 22) 개의 관심 객체(움직임, 정지)에 대하여 본 논문에서 제안한 방법으로 관심 객체를 추출한 결과

관심 객체 (움직임, 정지)		Predicted Class	
		Yes	No
Actual Class	Yes	TP = 48 (31, 17)	FN = 9 (4, 5)
	No	FP = 9 (7, 2)	TN

* TP(true positive), FN(false negative), FP(false positive), TN(true negative)

표 2. 움직임 관심 객체 추출을 위한 필터링 방법의 효용성 분석

	수작업 선정	움직임 객체 by Zhong[23] (A)	필터링 결과(B)	필터링 비율 ((A-B)/A)
관심 객체	35	31	31	0.00
비관심 객체	-	13	7	0.46
소계	35	44	38	0.14

표 3. 본 논문에서 제안한 방법의 성능 분석

	재현율 (Recall)	정확률 (Precision)	F-측정값 (F-measure)
전체 관심 객체	0.84	0.84	0.84
움직임 관심 객체	0.88	0.79	0.83
정지 관심 객체	0.77	0.89	0.83

체를 사전지식이 없이 추출하는 것이 어렵다는 것을 반영한 결과라고 판단한다. 한편, 전체적인 추출 성능을 나타내는 척도 중의 하나인 F-측정값은 움직임 및 정지 관심 객체가 모두 0.83이며, 전체 관심 객체에 대한 F-측정값은 0.84로 나타났다. 따라서 제안한 방법의 전체적인 성능은 대략 84%의 성공률을 보임을 알 수 있다.

본 논문에서는 여러 가지 필터링 조건의 만족 여부를 조사하기 위해 3절에서 T_{LM} , T_{SM} , T_{BM} , T_{SMA} , T_{LS} , T_{SS} , T_{Vs} 등의 실험적 임계값을 활용하였으며, 또한 가장자리 시공간(V_b)과 중앙 시공간(V_s)도 실험적으로 정의하였다. 이러한 임계값 및 정의값은 서로 연관되어 있어 각각에 대한 효용성 분석이 매우 어려웠다. 그러나 비록 필터링 조건에 대한 만족 여부를 조사하기 위해 임계값을 사용하는 것이 필수적이기는 하나, 향후에 실험적으로 설정한 각 값에 대하여 타당성 분석을 할 수 있는 방안을 강구할 필요가 있다고 판단한다.

한편, 추출된 정지 관심 객체의 경계 부분이 정확하지 못한 측면이 있는데, 이것은 중심 객체 추출 방법[5]에서의 핵심 영역만을 추출하여 주요 정지 객체로 결정하였기 때문이다. 따라서 방법[5]에서 배경 정보를 활용하고 있는 나머지 단계까지 적용하면 보다 적확한 객체 경계를 구할 수 있을 것으로 판단한다. 또한 그림 7의 자연환경 동영상 샷에서 잘못 추출된 정지 관심 객체(다리)와 같은 것은 사전에 관심 객체의 포함여부를 판단할 수 있는 영상 분류 방법[27]을 활용함으로써 불필요하게 추출되는 것을 방지할 수 있을 것으로 예상한다.

5. 결 론

본 논문에서는 동영상에서 움직이는 관심 객체와 정지된 관심 객체가 만족시켜야 할 조건을 정의하고,

이 조건에 따라 움직임 및 정지 관심 객체를 추출하는 방법을 제안하였다. 10편의 영화 및 드라마 동영상 중 50개 샷(shot)에 대하여 실험한 결과, 제안한 방법에 의해 일반 실험자가 수작업으로 선정한 관심 객체를 약 84% 성공률로 추출할 수 있었다. 정지 관심 객체의 추출 재현율(recall)은 움직임 관심 객체에 비해 상대적으로 낮았는데, 이것은 사전 정보 없이 어떤 프레임에서 정지 객체를 정의하고 추출하는 것이 어렵다는 것은 의미한다고 볼 수 있다. 반면에, 움직임 관심 객체의 추출 정확률(precision)은 79% 정도로 낮았는데, 이것은 본 논문에서 제시한 무의미한 비관심 움직임 객체의 필터링 조건에 지능적인 조건을 추가하여 보다 개선할 여지가 있음을 보여준 것이다. 한편, 본 논문에서는 관심 객체의 정의를 위한 조건들의 만족 여부를 조사하기 위해 여러 가지 임계값을 실험적으로 구하여 사용하였으나, 향후에는 각 임계값 설정을 분석에 의해 보다 타당하게 설정할 수 있는 방법을 강구할 필요가 있다. 제안한 관심 객체 추출 방법은 비디오 압축 및 내용기반 비디오 검색 분야에서 유용하게 활용될 수 있을 것으로 기대한다.

참 고 문 헌

- [1] R. Wang, Q. Cheng, and T. Huang, "Identify Regions of Interest(ROI) for Video Watermark Embedment with Principle Component Analysis," *ACM Multimedia 2000*, pp. 459-461, 2000.
- [2] A. Jerbi, J. Wang, and S. Shirani, "Error-Resilient Region-of-Interest Video Coding," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 15, No. 9, pp. 1175-1181, 2005.
- [3] M.J. Chen, M.C. Chi, C.T. Hsu, and J.W. Chen, "ROI Video Coding Based on H.263+ with Robust Skin-Color Detection Technique," *IEEE Trans. on Consumer Electronics*, Vol. 49, pp. 724-730, 2003.
- [4] X. Sun, J. Foote, D. Kimber, and B.S. Manjunath, "Region of Interest Extraction and Virtual Camera Control based on Panoramic Video Capturing," *IEEE Trans. on Multimedia*, Vol. 7, No. 5, pp. 981-990, 2005.
- [5] S. Kim, S. Park, and M. Kim, "Central Object Extraction for Object-Based Image Retrieval," *Lecture Notes in Computer Science*, Vol. 2728, pp. 39-49, 2003.
- [6] C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik, "Blobworld: A System for Region-Based Image Indexing and Retrieval," *VISUAL'99*, Amsterdam, Netherlands, pp. 509-516, 1999.
- [7] A.H. Kam, T.T. Ng, N.G. Kingsbury, and W.J. Fitzgerald, "Content Based Image Retrieval through Object Extraction and Querying," *IEEE Workshop on Content-based Access of Image and Video Libraries*, pp. 91-95, 2000.
- [8] W. Wang, Y. Song, and A. Zhang, "Semantics Retrieval by Region Saliency," *Int'l Conf. on Image and Video Retrieval*, pp. 29-37, 2002.
- [9] D. Zhong and S.F. Chang, "An Integrated Approach for Content-Based Video Object Segmentation and Retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 9, No. 8, pp. 1259-1268, 1999.
- [10] Q. Huang, B. Dom, D. Steels, J. Ashely, and W. Niblack, "Foreground / Background Segmentation of Color Images by Integration of Multiple Cues," *Int'l Conf. on Image Processing*, pp. 246-249, 1995.
- [11] Y. Lu and H. Guo, "Background Removal in Image Indexing and Retrieval," *Int'l Conf. on Image Analysis and Processing*, pp. 933-938, 1999.
- [12] S.B. Park, J.W. Lee, and S.K. Kim, "Content-Based Image Classification Using a Neural Network," *Pattern Recognition Letter*, Vol. 25, pp. 287-300, 2004.
- [13] L. Itti, C. Koch, and E. Niebur, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, pp. 1254-1259, 1998.

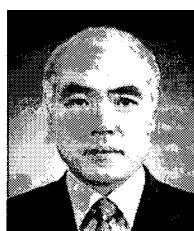
- [14] J.R. Serra and J.B. Subirana, "Texture Frame Curves and Regions of Attention Using Adaptive Non-cartesian Networks," *Pattern Recognition*, Vol. 32, pp. 503-515, 1999.
- [15] W. Osberger and A.J. Maeder, "Automatic Identification of Perceptually Important Regions in an Image," *IEEE Int'l Conf. on Pattern Recognition*, Vol. 1, pp. 701-704, 1998.
- [16] B.C. Ko and J.Y. Nam, "Object-of-Interest Image Segmentation Based on Human Attention and Semantic Region Clustering," *Journal of Optical Society of America*, Vol. 23, No. 10, pp. 2462-2470, 2006.
- [17] J. Han and K.N. Ngan, "Automatic Segmentation of Objects of Interest in Video: a Unified Framework," *Proc. Int'l Symposium on Intelligent Signal Processing and Communication Systems*, pp. 375-378, 2004.
- [18] W.H. Cheng, W.T. Chu, J.H. Kuo, and J.L. Wu, "Automatic Video Region-of-Interest Determination Based on User Attention Model," *IEEE Int'l Symposium on Circuits and Systems*, Vol. 4, pp. 3219-3222, 2005.
- [19] G. Hua, Z. Liu, Z. Zhang, and Y. Wu, "Iterative Local-Global Energy Minimization for Automatic Extraction of Objects of Interest," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 10, pp. 1701-1706, 2006.
- [20] 롤런드 루이스 (한인규 옮김), *비디오 촬영 테크닉*, 미진사, 1993.
- [21] 스티븐 디 캐츠 (김학순, 최병근 옮김), *영화연출론*(shot by shot), 시공사, 1997.
- [22] F. Porikli and Y. Wang, "Automatic Video Object Segmentation Using Volume Growing and Hierarchical Clustering," *EURASIP Journal on Applied Signal Processing*, Vol. 6, pp. 814-832, 2004.
- [23] D. Zhong and S.F. Chang, "Long-Term Moving Object Segmentation and Tracking Using Spatio-Temporal Consistency," *Int'l Conf. on Image Processing*, Vol. 2, pp. 57-60, 2001.
- [24] J. Huang, S.R. Kumar, M. Mitra, W.J. Zhu, and R. Zabih, "Image Indexing Using Color Correlogram," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 762-768, 1997.
- [25] Y. Deng, B.S. Manjunath, and H. Shin, "Color Image Segmentation," *IEEE Conf. on Computer Vision and Pattern Recognition*, Vol. 2, pp. 446-451, 1999.
- [26] I.H. Witten and E. Frank, *Data Mining*, Morgan Kaufmann Publishers, San Francisco, CA. USA, 2000.
- [27] S. Kim, S. Park, and M. Kim, "Image Classification into Object / Non-object Classes," *Lecture Notes in Computer Science*, Vol. 3115, pp. 393-400, 2004.



박 소 정

2003년 2월 부산대학교 정보컴퓨터공학부 학사
2005년 8월 부산대학교 컴퓨터공학과 석사
2005년 9월~현재 부산대학교 컴퓨터공학과 박사과정

관심분야 : 영상 처리 및 이해



김 민 환

1980년 2월 서울대학교 전기공학과 학사
1983년 2월 서울대학교 컴퓨터공학과 석사
1988년 8월 서울대학교 컴퓨터공학과 박사
1986년 3월~현재 부산대학교 컴퓨터공학과 교수

관심분야 : 영상분석 및 컴퓨터시각