낮은 피사계 심도의 동영상에서 포커스 된 비디오 객체의 자동 검출

(Automatic Extraction of Focused Video Object from Low Depth-of-Field Image Sequences)

박 정 우 † 김 창 익 ††
(Jungwoo Park) (Changick Kim)

요 약 영상을 낮은 피사계 심도로 찍는 카메라 기법은 전통적으로 널리 이용되는 영상 취득 기술이다. 이 방법을 사용하면 사진사가 사진이나 동영상을 찍을 때 영상의 관심 영역에만 포커스를 두어 선명하게 표현하고 나머지는 흐릿하게 함으로써 자신의 의도를 보는 이에게의 분명하게 전달 할 수 있다. 본 논문은 이러한 피사계 심도가 낮은 동영상 입력에 대하여 사용자의 도움 없이 포커스 된 비디오 객체를 추출하는 새로운 방법을 제안한다. 본 연구에서 제안하는 방법은 크게 두 모듈로 나뉜다. 첫 번째 모듈에서는 동영상의 첫 번째 프레임에 대해서 포커스 된 영역과 그렇지 않은 흐릿한 부분을 자동으로 구분하여 관심 물체만을 추출한다. 두 번째 모듈에서는 첫 번째 모듈에서 구한 관심 물체의 모델을 바탕으로 동영상 프레임에서의 관심 물체만을 실시간이나 실시간에 가깝게 추출한다. 본 논문에서 제안하는 방법은 가상현실(VR)이나 실감 방송, 비디오 인덱싱 시스템과 같은 여러 응용 분야에 효과적으로 적용될 수 있고, 이러한 유용성은 실험 결과를 통해 보였다.

키워드 : 관심 물체, 낮은 피사계 심도, 비디오 객체 분할, 실감 방송.

Abstract The paper proposes a novel unsupervised video object segmentation algorithm for image sequences with low depth-of-field (DOF), which is a popular photographic technique enabling to represent the intention of photographer by giving a clear focus only on an object-of-interest (OOI). The proposed algorithm largely consists of two modules. The first module automatically extracts OOIs from the first frame by separating sharply focused OOIs from other out-of-focused foreground or background objects. The second module tracks OOIs for the rest of the video sequence, aimed at running the system in real-time, or at least, semi-real-time. The experimental results indicate that the proposed algorithm provides an effective tool, which can be a basis of applications, such as video analysis for virtual reality, immersive video system, photo-realistic video scene generation and video indexing systems.

Key words : Object of interest, low depth of field, video object segmentation, immersive video

1. 서 론

컴퓨터 비전이나 영상 처리 분야에서 사용자의 도움 없이 자동으로 영상에서 주제가 되는 관심 영역을 추출하는 문제는 굉장히 어렵거나 까다로운 문제이다. 일반적으로 영상 분할(image segmentation)은 주어진 영상을 균일한 컬러나 비슷한 명암의 세기로 나누는 것을 의미한다. 이러한 배경과 물체를 구분하지 않는 전통적인 영상 분할 방법은 관심 물체만을 배경에서 분리하여 인식하거나 영상 합성을 하기에 부적합하다. 이를 위한 전처리 단계에 해당하는 영상 분할은 기본적으로 물리적인 특징을 이용하여 영상 내부를 균일한 영역으로 구분해 나누는 것뿐만 아니라, 의미상으로 동일한 관심영역과 그렇지 않은 배경영역을 분리하는 기능도 필요하다. 우리는 기존 연구[1]에서 그림 1과 같은 낮은 피사계 심도의 영상으로부터 포커스 된 관심 물체 영역만을 추출하는 방법을 제안하였으며 기존의 다른 방법에 비

· 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원 사업의 연구결과로 수행되었음(IITA-2006-C1090-0603-0017)
† 학생회원 : 한국정보통신대학교 공학부
jwpark@icu.ac.kr
†† 정 회 원 : 한국정보통신대학교 공학부 교수
ckim@icu.ac.kr
논문접수 : 2005년 12월 1일
심사완료 : 2006년 9월 4일

그림 1 낮은 피사계 심도의 영상

해 더욱 향상된 성능을 갖는 것을 확인하였다. 영상 픽셀의 밝기나 텍스처를 고려하는 기존의 영상 추출 방법과는 달리 낮은 피사계 심도 영상의 포커스 정보는 사용자의 도움 없이 자동으로 관심 영역(OOI: object-of-interest)을 추출하는데 중요한 역할을 한다. 본 논문은 기존의 낮은 피사계 심도를 가진 정지 영상에서 관심 영역을 추출하는 방법에 대한 연구[1]를 낮은 피사계 심도의 동영상으로 확장해서 포커스 된 비디오 객체를 효율적으로 추출하는 것을 목적으로 한다.

낮은 피사계 심도의 영상에서 관심 대상에만 포커스를 주어 사진을 찍는 방법은 영상에 인위적인 깊이감을 줌으로써 사람이 좀 더 사진을 잘 이해 할 수 있도록 도와주는데 널리 사용되는 기술이다. 이러한 낮은 피사계 심도의 영상에 존재하는 포커스 단서는 영상의 관심 물체만을 추출하기 위한 중요한 특징으로 사용할 수 있다. 피사계 심도가 낮은 영상은 선명하게 포커스 된 영역과 그렇지 않은 영역으로 나뉜다. 포커스가 맞지 않아 흐릿하게 보이는 영상의 배경이나 물체의 모델링을 위해 다음과 같은 2차원 가우시안 함수를 이용할 수 있다.

$$G_\sigma(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (1)$$

σ는 영상의 흐릿한 (defocusing or blurring) 정도를 조절하는 필터 매개 변수로 사용된다. 그래서 (x,y) 픽셀 위치에서 포커스 되지 않아 흐릿한 영상 $I_d(x,y)$는 식 (2)처럼 원래 포커스 된 영상 $I_f(x,y)$와 가우시안 함수 $G_\sigma(x,y)$를 선형 컨볼루션(linear convolution)을 한 형태로 모델링 할 수 있다.

$$I_d(x,y)=G_\sigma(x,y)*I_f(x,y) \quad (2)$$

포커스 되지 않은 흐릿한 영역은 식 (2)와 같이 영상에 저역 통과 필터(low pass filter)를 적용하여 높은 주파수를 가진 부분이 제거되는 과정으로 나타낼 수 있다. 그래서 피사계 심도가 낮은 영상에서 확실하게 포커스 된 부분은 그렇지 않은 영역보다 높은 주파수 성분을 갖는다는 가정은 주파수의 크기를 비교함으로써 포커스 된 선명한 영역과 그렇지 않은 영역을 구분하는 단서를 제공해준다.

일반적으로 사용자의 도움 없이 자동으로 일반 영상에서 의미 있는 관심 영역을 찾기는 매우 어렵다. 왜냐하면 영상에서 의미 있는 관심 영역의 판단은 사람마다 다른 매우 주관적인 요소이기 때문이다. 대개 사용자는 자신의 인지정보를 이용하여 일반 영상 내 의미 있는 관심 물체 영역을 직접 수동으로 정하는 방법으로 영상의 의미 있는 관심 영역을 추출하였다[3-5]. 그러나 낮은 피사계 심도의 카메라 기법을 적용하여 영상 데이타를 얻는다면, 사진사가 자신의 의도를 그림 1과 같이 포커스 정보를 이용하여 영상에 표현할 수 있다. 또한, 이러한 낮은 피사계 심도 영상의 포커스 정보는 사람이 2차원 영상에서 깊이감을 느낄 수도 있게 해줌으로써 영상을 더욱 이해하기 쉽게 해주는 특징을 갖는다[2]. 이와 같이 일반 영상이 아닌, 포커스 정보로 사진사의 의도가 반영된 낮은 피사계 심도 영상을 영상 분할에 사용한다면 사용자의 도움 없이 의미 있는 관심 영역을 정확하게 추출하는 일이 가능하다[1].

우리는 이러한 영상 분할 방법을 낮은 피사계 심도를 갖는 동영상 데이타로 확장하여 비디오 객체를 기반으로 하는 멀티미디어 응용 분야에 효과적으로 적용하고자 한다. 동영상은 정지 영상의 집합이므로, 낮은 피사계 심도 동영상의 매 프레임마다 기존의 낮은 피사계 심도를 가진 정지 영상에서 관심 영역을 추출하는 방법[1]을 적용하여 의미 있는 관심영역인 비디오 객체를 추출할 수 있다. 하지만 동영상 데이타가 가진 시간적, 공간적 중복성을 이용한다면 좀 더 효율적으로 비디오 객체를 배경에서 분리 할 수 있다[6-8]. 낮은 피사계 심도의 동영상에서 자동으로 의미 있는 포커스 된 물체를 추출하면 객체를 기반으로 하는 디지털 비디오 응용분야에 효율적으로 적용할 수 있다. 예를 들어, MPEG-4 부호화기는 객체의 정보를 이용한 동영상 압축 알고리즘을 사용하기 때문에 효율적인 객체 기반의 영상 압축을 위해서는 부호화 이전에 동영상 내에서 의미 있는 객체의 추출을 필요로 한다. MPEG-7의 경우에는 영상에 존재하는 물체에 대한 형태와 움직임 정보를 포함하는 기술자(descriptor)를 이용하여 정지 영상과 동영상에 관한 인덱싱과 검색을 하므로 각 프레임에 대한 의미 있는 객체 분할이 선행되어야 한다. 또한, 대용량 영상 데이타베이스에서 컨텍스트 기반의 영상 검출을 위한 영상 인덱싱, 전자 현미경을 사용한 분자나 세포 수준의 3차원 영상 복원 및 분석, 디지털 카메라를 위한 영상 개선(image enhancement), 깊이 추정을 하기 위한 거리 영상 분할(range segmentation)과 다양한 포커

스 정보를 가진 여러 장의 영상 합성[9]과 같은 다양한 분야에도 폭넓게 적용 할 수 있다.

이와 같이 기존에 취득된 낮은 피사계 심도를 가진 동영상으로부터 비디오 객체를 추출하는 작업 외에도, 추출된 비디오 객체를 컴퓨터 그래픽으로 생성된 다른 배경과 합성하는 응용이나 2차원 영상에서 3차원으로 재구성 가능한 실감 방송 등에의 응용 등을 염두에 두어 동영상 취득시에 낮은 피사계 심도 기법을 이용하여 촬영하는 것 또한 더욱 다양한 응용을 위해 가능하다. 정지 영상을 다루는 기존 연구[1]와는 달리 동영상에서 실시간으로 비디오 객체를 추출해야 하므로 고속의 효율적인 접근 방법이 필요하다. 이 방법은 다음과 같이 크게 두 부분으로 구성된다. 동영상에서 추출한 모델을 결정하기 위해, 첫 번째 프레임에 대해서 기존 연구[1]를 개선한 방법을 적용하여 포커스 된 관심 영역을 추출한다. 이렇게 추출한 관심 물체 모델을 바탕으로 동영상 데이타의 시간적, 공간적 특징을 이용하여 나머지 프레임들에 대해 포커스 된 관심 영역을 추적한다.

본 논문의 구성은 다음과 같다. 다음 장에서는 본 논문에서 제안하는 피사계 심도 낮은 동영상에서 포커스 된 관심영역을 추출하는 2가지 모듈로 구성된 방법들에 대해 자세히 알아보고, 3장에서는 제안한 방법 기법을 실제 낮은 피사계 심도 동영상에 적용한 결과를 보여주며 4장에서 결론을 맺는다.

2. 제안하는 알고리즘

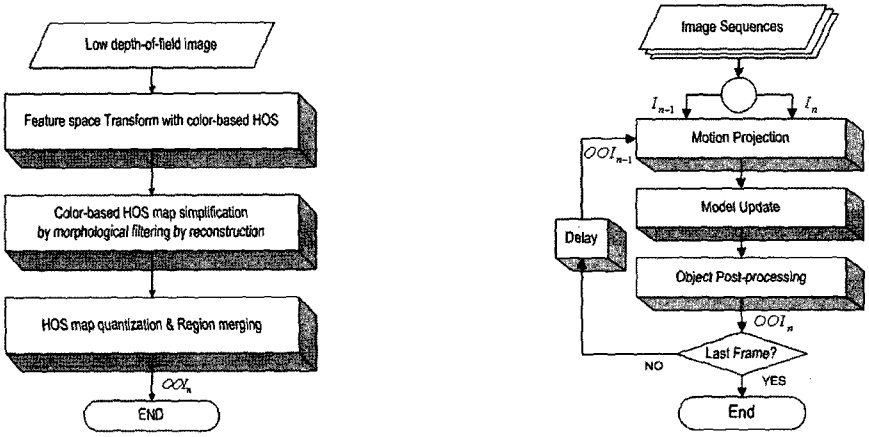
이번 절에서는 의미 있는 동영상 객체를 추출하기 위해서 사용자가 추출할 모델을 직접 설정하는 과정 없이 자동으로 분할하는 알고리즘에 대해 알아본다. 우선 첫 번째 프레임에서 주어진 동영상을 분할하기 위한 초기 모델을 설정하기 위해 그림 2(a)와 같이 포커스 된 하나 이상의 관심 영역을 추출한다. 이 방법을 통해 얻은 포커스 된 객체 모델을 기반으로 하여 그림 2(b)와 같이 연속된 다음 동영상 프레임에서 관심 영역만을 추적(OOI tracking)하여 추출할 수 있다. 즉, 현재 프레임에서 추출한 새로운 비디오 객체 모델은 프레임 간의 시간적인 중복성을 고려하여 다음 프레임의 관심 영역을 예측하는데 사용된다. 또한, 관심 영역의 움직임 예측을 통해 다음 프레임에서의 관심 영역 추출을 위한 후보 영역을 지정함으로써 불필요한 연산을 최소화 할 수 있다. 다음 장에서 모듈별로 각 단계에 해당하는 자세한 특징을 살펴보기로 한다.

2.1 첫 번째 프레임에서 포커스 된 의미 있는 관심 모델 추출

기존의 연구[10]에서는 동영상에서 비디오 객체를 자동으로 추출하기 전에 사용자가 영상에서 분할을 하기 위한 관심 물체를 선택해서 수동으로 그 해당 영역을 제한하는 방법을 사용한다. 이렇게 사용자가 의미 있는 관심 영역을 설정함으로써 동영상 데이타에서 효율적으로 비디오 객체를 추출한다. 그러나 이와 같은 사용자의 개입은 비디오 객체를 완전 자동으로 추출하는 응용 시스템을 개발하는데 장애 요소로 작용한다. 다음 절에서는 기존에 제안했던 영상 분할 방법[1]을 개선하여, 동영상의 첫 번째 프레임에서 의미 있는 관심 영역을 추출하는 방법에 대해서 알아본다.

2.1.1 컬러 기반의 HOS (higher order statistics) 지도를 사용한 낮은 피사계 심도 영상의 분할

기존에 제안했던 영상 분할 알고리즘[1]은 포커스 정보를 활용하여 낮은 피사계 심도 영상을 포커스 된 관심 영역과 그렇지 않은 배경이나 물체가 포함된 영역으



(a) 낮은 피사계 심도 영상의 초기 프레임에 대한 영상 분할 (n=0) (b) 관심 영역 추적을 이용한 영상 분할 방법 (n>0)

그림 2 블록 다이어그램

로 분리 가능하다. 본 논문에서는 낮은 피사계 심도의 동영상의 첫 번째 프레임에 대해 [1]에서 제안된 방법을 더욱 향상시킨 알고리즘을 적용하여 포커스 된 의미 있는 관심 영역 모델을 자동으로 추출한다. 사용자의 도움을 받지 않고 의미 있는 비디오 객체 모델을 추출하기 위해서는 그림 2(a)와 같이 3단계로 구성된 영상 추출 방법을 사용한다. 첫 번째 단계에서는 낮은 피사계 심도의 영상을 포커스가 존재하는 영역과 포커스가 맞지 않아 흐릿한 배경이나 물체 영역으로 구분할 수 있는 특징 공간으로 변환한다. 영상의 포커스 정보를 구분할 수 있는 이러한 특징 공간은 낮은 피사계 심도 영상의 각 픽셀에 대해 고차 통계(HOS: higher-order statistics)를 계산을 하여 얻을 수 있다. 이렇게 하여 생성된 HOS 지도는 영상에 존재하는 흐릿한 배경이나 물체와 같은 가우시안 노이즈로 구성된 모델과 비가우시안 정보를 가진 영상을 대표하는 포커스 된 물체를 구분하기 위한 좋은 지표가 된다[16].

흑백 영상의 밝기만을 고려한 기존 연구[1]의 HOS 지도는 많은 경우에 있어서 영상을 분할하기 좋은 입력 지표를 제공해 주지만, 그림 3(a)의 컬러 입력 영상에서와 같이 서로 뚜렷이 구분되는 다른 색상임에도 밝기 값이 비슷하여 흑백 영상 내에서 경계 부분이 불명확한 경우(그림 3(b)의 원 부분)에는 그림 3(c)의 경우와 같이 제대로 된 HOS 지도를 제공하지 못하는 한계가 있고, 이러한 HOS 지도는 그림 3(d)와 같이 관심 영역을 제대로 검출 할 수 없는 최악의 결과를 가져올 수도 있다. 그래서 본 논문에서는 개선된 지표를 얻기 위해 입력영상의 RGB 세 채널을 모두 고려한 컬러기반의 HOS 지도를 작성하는 방법을 제안한다. 개선된 HOS 지도를 얻기 위해서는 RGB 입력 영상의 각 채널에 대해서 픽셀당 4차 모멘트를 각기 계산한다. 이러한 고차 모멘트의 이용은 가우시안 노이즈에 대한 뛰어난 압축효과와 비가우시안 정보에 대한 보존 능력이 우수하므로 검출 및 분류의 해법으로 많이 사용된다[1].

R 채널의 한 픽셀 (x,y)에 대한 4차 모멘트는 다음과 같이 정의할 수 있다.

$$\hat{m}_R^{(4)}(x,y)=\frac{1}{N_\eta}\sum_{(s,t)\in\eta(x,y)}(I_{red}(s,t)-\hat{m}_R(x,y))^4 \quad (3)$$

여기서, (x,y)는 픽셀 (x,y)를 기준으로 한 이웃 픽셀들이고, $\hat{m}_R(x,y)$은 입력 영상 $I(x,y)$ 의 red 채널 $I_{red}(x,y)$의 샘플 평균이고 (즉, $\hat{m}_R(x,y)=\frac{1}{N_\eta}\sum_{(s,t)\in\eta(x,y)} I_{red}(s,t)$), η의 크기는 N_η이다. 세 채널 4차 모멘트 $\hat{m}_R^{(4)}(x,y)$, $\hat{m}_G^{(4)}(x,y)$, $\hat{m}_B^{(4)}(x,y)$의 값들 중 다음과 같이 픽셀당 최대값을 구한다.

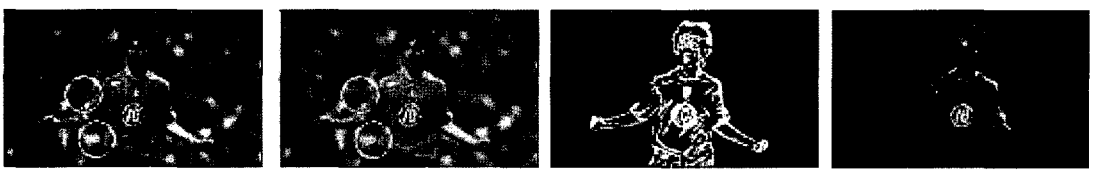
$$tempHOS(x,y)=\max(\hat{m}_R^{(4)}(x,y),\ \hat{m}_G^{(4)}(x,y),\ \hat{m}_B^{(4)}(x,y)) \quad (4)$$

이렇게 구한 HOS가 갖는 값의 범위는 영상에 따라 결과 값이 굉장히 다양하므로 각 픽셀에 대한 값은 0에서 255값을 갖는 범위를 갖도록 정규화 한다. 이렇게 정규화 된 컬러 기반의 HOS 지도의 각 픽셀 값은 다음과 같이 정의 한다.

$$colorHOS(x,y)=\min(255, tempHOS(x,y)/DSF) \quad (5)$$

DSF는 다운 스케일 인자(down scaling factor)로써 결과 값의 범위를 [0, 255]로 정규화 하는데 사용한다. 다양한 영상을 테스트한 결과, DSF 값은 300이 적절함을 알았고, 본 논문에서는 모든 영상에 대하여 이와 동일한 값을 적용하여 실험하였다. 흑백 기반의 HOS 지도(그림 3(c))와 컬러 기반의 HOS 지도(그림 4(a))를 비교해 보았을 때, 후자의 경우 포커스된 영역의 경계면이 더욱 뚜렷하게 나타남을 알 수 있다.

다음으로 모폴로지 필터(morphological filter *by reconstruction*)를 사용하여 앞에서 구한 컬러 기반의 HOS 지도의 내부에 포함된 홀과 지도 외부의 노이즈를 제거하는 HOS 지도의 단순화 과정을 수행한다[1]. 본 연구에서 사용한 모폴로지 필터는 닫힘 필터(morphological closing *by reconstruction*)를 먼저 HOS 지도에 적용한 후 열림 필터(morphological opening *by reconstruction*)를 적용하는 모폴로지 닫힘-열림 필터(morphological closing-opening *by reconstruction*)을 사용하였다[17]. 이 필터의 특징은 필터가 적용되는 영



(a) (b) (c) (d)

그림 3 영상의 밝기에 기반한 HOS 지도: (a) 낮은 피사계 심도의 컬러 입력 영상, (b) 흑백 채널로 변환한 영상, (c) 흑백 영상에 기반한 HOS 지도, (d) (c)의 결과를 적용하여 영상 분할한 결과

역의 주변의 외관(boundary)을 변형시키지 않고 보존하면서 내부의 홀과 영역 밖의 노이즈를 제거할 수 있다는 점이다. 또한, 다른 모폴로지 필터와 동일하게 홀을 채우는 범위와 노이즈를 제거하는 범위는 구조자(structuring element)에 따라 달라진다. 그림 4(b)에서 볼 수 있듯이, 모폴로지 필터를 적용하면 포커스 된 관심 영역의 내부는 잘 채워지는 반면 관심 영역 밖의 노이즈 성분은 제거되는 것을 알 수 있다. 마지막으로 포커스 된 관심영역과 배경영역을 분리하기 위해 단순화된 HOS 지도를 바탕으로 영역 병합(region merging)을 수행한다. 기존의 방법[1]보다 좀 더 빠르고 효율적인 영역 병합을 하기 위해서 단순화된 HOS 지도는 그림 4(c)와 같이 세 가지 단계로 구성된, 즉, 흰색 픽셀을 갖는 관심 영역과 회색 영역의 관심 영역 후보, 검은색 영역의 배경 영역으로 양자화 된다.

(a) (b) (c)

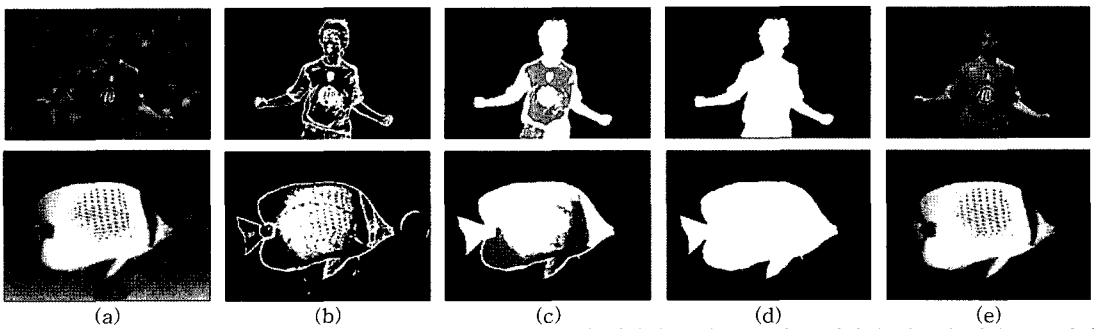
그림 4 컬러 기반 HOS 지도의 단순화와 양자화: (a) 컬러 기반 HOS 지도, (b) 모폴로지 연산을 이용하여 (a)를 단순화한 지도, (c) (b)의 양자화 결과

포커스 된 관심 영역의 최종 결정은 기존 방법[1]의 영역 병합 방법과 같이 회색 영역을 둘러싸고 있는 흰색의 절대 포커스 영역이 절반 이상 존재할 때, 이 회색 관심 영역 후보지를 흰색의 포커스 된 관심 물체영역으로 통합함으로써 이루어진다. 그림 5는 주어진 동영상의 첫 번째 프레임에 대해 개선한 알고리즘의 도식화를 보여준다.

동영상은 일련의 연속된 정지 영상으로 구성된다. 낮은 피사계 심도의 비디오 데이타에서 포커스 된 관심 영역만 추출하기 위해서는 모든 프레임에 이 절에서 제안한 방법을 적용하는 것이 가능하다. 그러나 연속된 프레임 간의 시간적인 중복성을 고려하지 않고 동영상에서 관심 영역을 추출 하는 것은 매우 비효율적인 접근 방법이다. 그래서 초기 프레임에 대한 영상 분할이 완료된 후, 이어지는 프레임에 대해서는 계산 복잡도가 높은 정지 영상 분할 알고리즘을 각 동영상 프레임에 적용하는 대신에 이전 프레임에서 추출한 관심 물체에 대한 모델을 기반으로 하여 현재 프레임의 새로운 관심 물체를 고속으로 찾아내는 물체 추적 방법을 이용하였다. 다음 절에서는 이렇게 프레임 간의 시간적인 중복성을 고려하여 이전 프레임에서 추출한 관심 모델을 기반으로 하여 현재 프레임 내의 관심 물체의 위치를 추정하는 물체 추적 방법을 소개한다.

2.2 비디오 객체 추적 모듈

기존의 연구[1]에서 포커스 된 관심 영역을 구하기 위한 모폴로지 필터의 반복된 사용은 낮은 피사계 심도 영상의 분할에 있어 매우 중요한 역할을 함에도 불구하고 높은 연산량 때문에 동영상을 다루기에는 적합하지 않다. 고속의 비디오 관련 응용 시스템을 개발하기 위해서는 이와 같은 모폴로지 필터를 사용하는 과정은 더욱 효율적인 객체 분할 알고리즘으로 대체되어야 한다. 일단 첫 번째 프레임에서 포커스 된 관심 물체를 추출하고 나면, 동영상의 두 번째 프레임부터 이전 프레임에서 추출된 모델을 바탕으로 포커스 된 영역을 추적(object tracking)한다. 효율적으로 물체를 추적하기 위해 연속된 두 프레임 간의 비디오 객체의 움직임을 추정하고 위치를 추적하는 방법을 설계하였다. 그래서 비디오 객체를 추적하는 모듈은 이전 프레임의 관심 물체 모델을 가지고 현재 프레임의 대략적인 위치를 추정하는 모션



(a) (b) (c) (d) (e)

그림 5 컬러 기반 HOS 지도를 바탕으로한 낮은 피사계 심도의 영상의 분할: (a) 낮은 피사계 심도의 영상 (두 번째 줄의 "Bream" 테스트 영상의 첫번째 프레임은 실험을 위해 배경에 가우시안 블러링 연산 적용); (b) 컬러 기반 HOS 지도 (c) HOS 지도의 양자화 결과 (d) 최종 관심영역의 분할을 위한 영역 병합 결과 (e) 포커스 된 관심 영역의 추출

예측 단계와, 예측된 모션과 기존의 모델을 이용하여 새로운 현재 관심 영역 모델로 갱신하는 단계를 거친다. 마지막으로, 수정된 모델을 이용하여 현재 프레임에서 포커스 된 물체를 찾아내고 찾아낸 관심 영역 주변의 노이즈를 제거하는 단계를 거친다. 그림 2(b)의 블록 다이어그램과 같이, 이전 비디오 프레임에서 추출한 관심 물체 모델은 현재 프레임에서 포커스 된 관심 비디오 객체(video object)를 추정하는데 사용한다. 본 논문에서는 기존의 연구[10]에서 비디오 객체 추적에 사용한 방법을 개선하여 적용하였다.

2.2.1 모션 예측

비디오 객체를 추적하려면 프레임 간의 대응되는 관심 영역의 상관 관계를 규명하는 것이 그 바탕이 된다. 먼저 물체 추적을 하기 위하여 이전 프레임의 주어진 모델을 가지고 현재 프레임에서 그 물체가 위치할 대략적인 영역을 추정하는 모션 예측 과정을 필요로 한다. 연속된 두 장의 프레임 간의 관심 물체에 대한 모션 필드(motion field)는 다음과 같은 변환 함수에 의해서 가능하다.

$$I_n(x,y) = I_{n-1}(f(x,y), g(x,y)) \tag{6}$$

여기서 $I_n(x,y)$ 와 $I_{n-1}(x',y')$ 사이의 기하 관계는 $x' = f(x,y)$, $y' = g(x,y)$ 와 같은 변환 함수에 의해서 정의할 수 있다. 조밀한 모션 추정을 위해 사용하는 계층적인 블록 매칭 방법[14]은 높은 정확도를 가지고 고속으로 다음 물체가 위치할 영역을 예측하는데 사용한다. 그래서 영상의 각 픽셀(x,y)에 해당하는 변환 함수는 다음과 같이 나타낼 수 있다.

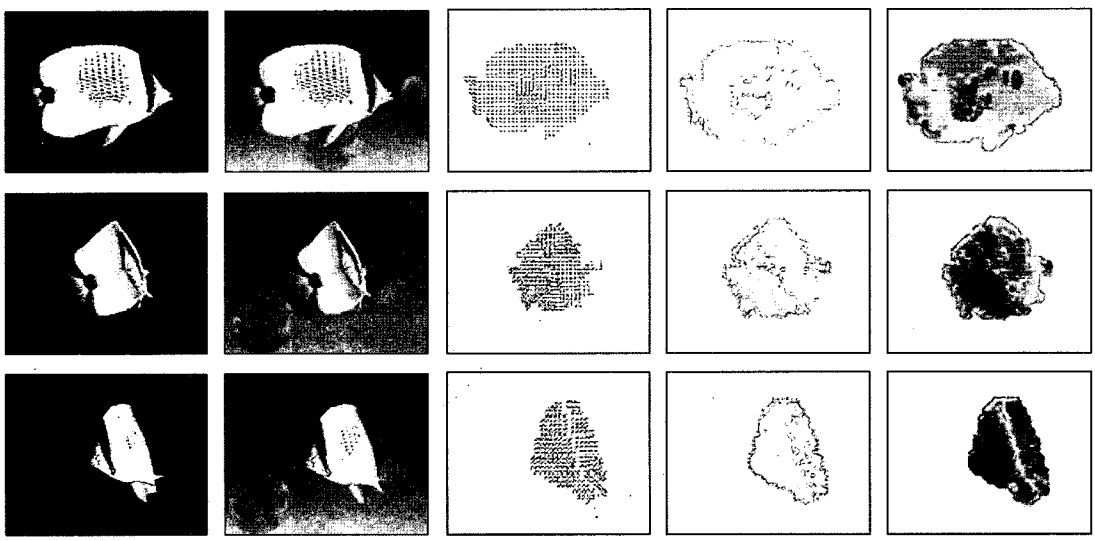
$$f(x,y) = x - u(x,y),\ g(x,y) = y - v(x,y) \tag{7}$$

여기서 $(u(x,y), v(x,y))$는 픽셀 (x,y)에서 예측한 모션 벡터이다.

동영상의 이전 프레임에서 추출한 포커스 된 물체 모델을 이용해서 현재 프레임의 모션 예측을 한다. $I_{n-1}(x,y)$ 와 $I_n(x,y)$ 는 각각 $n-1$ 과 n 번째 동영상 프레임이고 OOI_n 는 현재 프레임의 관심 물체(OOI: object-of-interest)를 의미한다. 이전 $n-1$번째 프레임에서 추출한 포커스 된 관심 영역인 OOI_{n-1} 과 현재 n 번째 영상의 상관관계는 모션 정보에 기반하여 OOI_{n-1}을 현재 프레임에서 예측함으로써 추정이 가능하다. 이전 프레임에서 추출한 관심 영역 OOI_{n-1} 을 현재 프레임 $I_n(x,y)$에 추정한 영역 P는 다음과 같이 나타낼 수 있다.

$$P = \left\{ (x+dx, y+dy) \,\middle|\, (x,y) \in OOI_{n-1} \right\} \tag{8}$$

여기서 (dx,dy)는 포커스 된 물체 OOI_{n-1} 의 모션 변위를 의미한다. 그림 6은 다음과 같이 계층적인 블록 매칭 과정을 보여준다. 그림 6(a)는 이전 프레임에서 추출한 관심 물체 OOI_{n-1} 모델이고 (b)는 현재 프레임을 나타낸다. 그림 6(c), (d), (e)는 이전 프레임의 관심 물체를 바탕으로 현재 프레임의 관심 영역을 추적하기 위하여 계층적인 블록 매칭을 사용하여 단계적으로 모션 예



(a) (b) (c) (d) (e)

그림 6 계층적인 블록 매칭 알고리즘을 적용한 과정: (a) 관심 영역에 해당하는 모델 (프레임 1, 113, 119), (b) 현재 영상 (프레임 2, 114, 120), (c), (d), (e): 블록 매칭한 결과 (대략적으로 추정한 모션 벡터에서부터 상세하게 예측한 모션 벡터를 차례로 보여준다.)

측을 한다. 그림 6은 대략적으로 모션 예측을 하는 단계에서부터 상세하게 구하는 단계까지 보여준다. 1초에 30 프레임으로 구성된 비디오 영상에서 관심 영역을 실시간에 가깝게 고속으로 정확하게 추출하기 위해서는 계산 복잡도는 낮으나 영상 분할 성능은 좋은 알고리즘을 사용해야 한다. 그래서 본 논문에서는 모션 추정을 이용하여 대략적인 물체의 위치를 추정하기 위해 그림 6(c)와 같이 낮은 단계의 계층적인 블록 매칭 알고리즘을 사용하였다. 이러한 블록 매칭 방법의 사용은 낮은 연산량으로 다음 프레임에서의 관심 영역 위치를 대략적으로 예측할 수 있다.

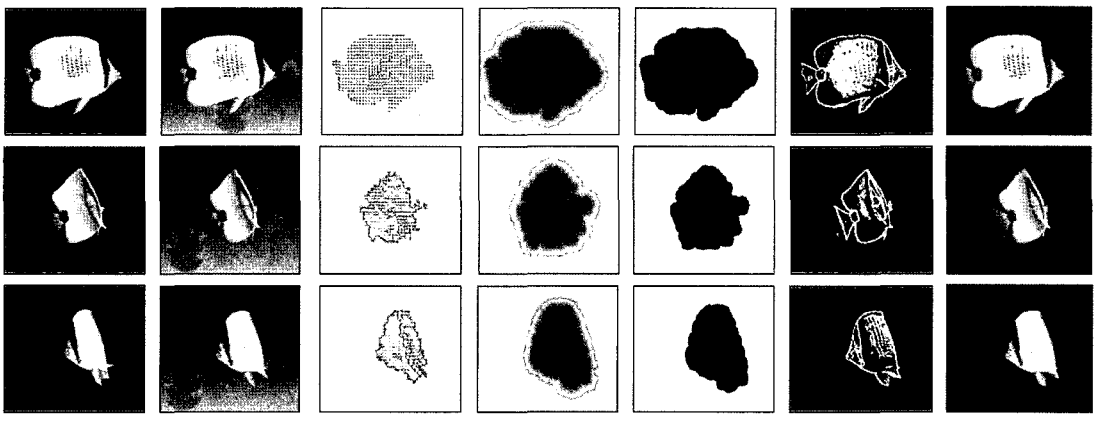
그러나 모션 정보만을 이용하여 현재 비디오 프레임의 관심 영역의 위치를 정확히 추정하는 것은 한계가 있기 때문에 이렇게 예측한 관심 영역을 영역 제한 알고리즘을 사용하여 수정할 수 있다. 일반적으로 컬러 영상의 정보나 흑백 영상의 세기를 바탕으로 하는 영상 분할의 경우에는 영상을 분할하여 예측한 영역과 분할된 영역을 비교하여 매치한다. 그러나 대부분의 경우 영상의 의미 없는 과도한 분할로 인해 이것을 처리하기 위한 높은 계산량을 필요로 한다. 그래서 [11,12]와 같은 기존의 연구에서는 계산량을 줄이기 위해 유사한 컬러 영역을 기반으로 영역 병합(region merging)을 수행한다. 그러나 이러한 영상의 컬러 정보나 세기 정보에 기반한 영역 병합은 우리의 관심 영역인 포커스 된 영역과 그렇지 않은 흐릿한 배경 영역을 하나의 영역으로 만들 가능성도 있다. 다음 절에서는 단순하면서도 효율적으로 모션 예측에 의해서 추정한 관심 영역의 크기를 제한하는 방법을 소개한다.

2.2.2 모델 갱신

이전 프레임의 관심 영역 모델을 바탕으로 계층적인 모션 예측 방법을 이용해서 현재 프레임에 존재하는 대략적인 관심 영역을 추정할 수 있다. 이렇게 모션 예측을 통해 얻은 영역을 거리 변환(distance transform)을 적용하여 현재 관심 영역을 수정한 확장 모델을 얻을 수 있다. 특징을 포함한 픽셀과 그렇지 않은 픽셀로 구성된 이진 영상을 대상으로 거리 변환을 수행하면, 각 픽셀의 값은 가장 가까운 거리에 있는 특징 픽셀까지의 거리로 정의된 그림 7(d)와 같은 거리 변환 영상(*DT*: distance transformed image)을 얻을 수 있다. 본 논문에서는 다음과 같이 예측한 영역 모델 P (그림 7(c))는 특징 픽셀로, 그렇지 않은 영역은 특징을 가지지 않은 픽셀로 정의하였고, 다음과 같은 식을 이용하여 추정한 영역 모델에 대해 수정한 결과를 얻을 수 있다.

$$\begin{cases} D(x,y)=1 & \text{if } DT(x,y) \le T_{\text{chamfer-3-4}} \\ D(x,y)=0 & \text{otherwise} \end{cases} \quad (9)$$

$T_{chamfer-3-4}$ 라는 임계값은 수정한 모델의 크기를 조절하는 요소로 사용한다. 본 연구에서는 유클리디안 거리 측정방법을 모사한 *Chamfer-3-4* 방법[14]을 사용하였다. 여기서 *Chamfer-3-4* 근사 방법에 기인한 거리는 대응하는 유클리디안 거리보다 3배 정도 더 큰 값을 가진다. 예를 들어, 식 (9)에서 임계값 $T_{chamfer-3-4}$을 30으로 정한다면 두 지점 간의 거리는 대략 10 픽셀 정도임



(a) (b) (c) (d) (e) (f) (g)

그림 7 제안하는 관심 영역을 추적하는 알고리즘: (a) 이전 프레임에서 추출한 포커스 된 관심 오브젝트 모델 (1, 113, 119 프레임); (b) 현재 피사계 심도가 낮은 영상 (프레임 2, 114, 120); (c) 이전 모델(a)을 바탕으로 현재 영상에 대략적인 모션 예측, P; (d) 거리 변환된 (distance transformed) 영상, DT, 영상의 세기가 밝을수록 모델에서부터 거리가 멀다 (e) DT(x,y)로부터 구해진 1차 예측된 관심 영역, D; (f) D의 영역에 대하여 컬러 기반의 HOS 지도 작성 (g) 영역 채우기 기법과 후처리 과정을 수행하여 얻은 최종 OOI.

을 뜻한다. 본 연구에서 CIF 포멧을 갖는 비디오 영상에 대해 75를 갖는 임계값을 적용하였다.

이렇게 거리 변환에 의해서 현재 프레임의 $D(x,y)$=1인 영역에 안에서 대해서 컬러 기반의 HOS 지도를 구한다. 이렇게 다음 프레임의 포커스 된 관심 영역의 위치를 대략적으로 제한해줌으로써 불필요한 연산을 대폭 줄이는 것이 가능하다. 또한, 고속으로 물체를 추출하기 위해서 기존 연구[1]에서 사용했던 모폴로지 필터의 반복적인 사용을 피하고 [10]에서 제안한 '영역 채우기 기법'(filling-in technique)을 사용하였다.

2.2.3 후처리 과정

본 논문에서 제안한 방법을 사용하여 추출한 비디오 객체는 움직이는 물체에 관한 전체적인 형태를 제공해 주지만 배경과 포커스 된 물체 사이의 경계 영역을 울퉁불퉁하게 추출되거나 관심 영역에서 분리된 작은 영역인 노이즈 영역도 함께 추출될 수도 있다. 이렇게 잘못 추출된 영역은 모션 추정 시 잘못된 예측 결과를 주어 전체 시스템의 성능을 떨어뜨리기도 한다. 그래서 이번 단계에서는 에러 출력들을 제거하거나 최소화하고 추출한 관심 영역의 경계 영역을 부드럽게 만드는 연산을 한다. 이를 위해 두 종류의 열림, 닫힘 연산을 하는 모폴로지 필터를 사용한다. 이렇게 필터를 적용하면 관심 물체 분할을 위한 최종 이진 마스크를 얻을 수 있고, 이 마스크를 바탕으로 포커스 된 관심 영역에 대한 최종 비디오 객체 분할을 할 수 있다.

3. 실험 결과

사용자의 도움 없이 본 논문에서 제안한 자동으로 비디오 객체를 추출하는 알고리즘을 피사계 심도가 낮은 테스트 동영상에 적용하였다. 동영상 데이타의 첫 번째 프레임에서 포커스 된 관심 영역을 추출하는 영상 분할 모듈은 사용자의 도움 없이 나머지 동영상 프레임에서 추출할 물체의 모델을 결정한다. 좀 더 정확하게 의미 있는 관심 영역의 분할을 위해서 입력 영상의 RGB 각 채널을 다 고려하여 컬러 기반의 HOS 지도를 사용한다. 식 (3)에서 컬러 기반 HOS지도를 구하기 위해 n는 기존 연구[1]와 동일한 3×3 크기의 주변 픽셀을 고려하였다. 그리고 모폴로지 필터를 사용하기 위해 31×31 크기의 구조자(structuring element)를 사용하였다. 단순화 된 HOS 지도를 양자화 하고 영역 병합을 수행함으로써 의미 있는 포커스 된 관심 물체의 모델을 자동으로 추출해 낼 수 있었다.

계층적인 모션 추정 방법을 적용하여, 이전 프레임에서 추출한 관심 영역 모델을 바탕으로 현재 비디오 프레임의 관심 물체의 대략적인 위치를 예측한다. 이전 프레임에서 추출한 관심 객체 모델과 현재 프레임의 예측한 관심 영역 사이의 변위 벡터를 추정하기 위해서 64×64 크기의 측정 윈도(measurement window)를 사용하였고, $T_{chamfer-3-4}$ 값을 75로 설정하여 예측한 새로운 관심 영역의 크기를 제한하였다. 이렇게 찾은 후보 관심 영역(candidate of OOI)에 대하여 컬러에 기반한 HOS 지도를 작성한 후, 계산량이 낮은 '영역 채우기 기법'을 사용하여 복잡한 배경에서 포커스 된 비디오 오브젝트만을 추출해 낼 수 있었다. 그림 7은 본 논문에서 제안한 알고리즘을 바탕으로 실험한 결과를 보여준다. 여러 장의 테스트 영상에 대한 각 단계의 결과를 볼 수 있다. 그림 8과 9는 일련의 비디오 프레임에서 포커스 된 오브젝트만을 추출한 결과를 나타낸다. 그림 8에서 사용된 동영상은, MPEG4의 실험 동영상인 CIF 포맷의 'Bream' 비디오에서 미리 분리되어 제공되고 있는 배경 객체만을 가우시안 필터링하여 실험에 이용하였으며, 이 실험 동영상은 포커스 된 관심 물체 외에 여러 개의 각기 다른 방향으로 움직이는 배경 객체를 포함한다.

실험은 인텔 펜티엄-IV 3.4GHz PC에서 수행되었고, CIF 포맷의 영상에 대해 제안한 알고리즘의 평균 수행시간은 프레임 한 장당 0.77 초가 소요되는 것을 알 수 있었다. 제안한 알고리즘을 구성하는 두 종류의 영상 분할 모듈과 오브젝트 추적 모듈의 각 단계에 해당하는 평균 수행시간을 표 1에 정리하였다. 첫 번째 영상 분할 모듈을 사용하여 관심 영역을 추출하는 경우에 있어서, 높은 정확도를 가지고 관심 물체를 추출 할 수 있지만 포커스 된 객체를 찾기 위한 모폴로지 필터(morphological filters *by reconstruction*)의 반복된 사용으로 동영상에 적용하여 비디오 객체를 추출하는 방법으로는 적절하지 않다.

본 실험에서는 비디오 객체 추출 알고리즘의 성능 평가를 위해 사용한 기존 연구[16]에서사용한 픽셀 기반의 측정 방법을 사용하였다. 관심 영역에 대한 기준 모델을

표 1 "Bream" 동영상에 대한 실험에서 각 모듈에 대한 단계별 평균 연산 시간

Segmentation module	Processing time (msec)	**Object tracking module**	Processing time (msec)
Color-based HOS map	143	Motion projection	532
Map simplification	3,321	Model update	142
Region merging	2	Post processing	98
Total	3,466	**Total**	772

바탕으로 추출한 포커스 된 관심 영역 사이의 왜곡 계산은 다음과 같다.

$$d(O^{est}, O^{ref}) = \frac{\sum_{(x,y)} O^{est}(x,y) \otimes O^{ref}(x,y)}{\sum_{(x,y)} O^{ref}(x,y)} \quad (10)$$

여기서 O^{est}와 O^{ref}는 관심 영역에 해당하는 이진 마스크와 MPEG 테스트 동영상에서 제공한 이진 마스크이고, ⊗연산은 이진 "XOR" 연산을 나타낸다. 표 2에서는 "Bream" 동영상에 본 논문에서 제안한 알고리즘을 적용하여 포커스 된 관심 영역을 찾는 영상 분할의 정확도와 에러율을 보여준다. 그림 8의 첫 번째 줄과 같이 프레임 간의 관심 모델의 모습이 거의 변화가 없고 비슷한 속도를 가진 변화(steady change)와 세 번째와 다섯 번째 줄과 같이 프레임 간에 관심 모델의 변화가 크고 속도도 일정하지 않은 변화(rapid change)를 구분하여 영상 분할 결과를 측정하였다. 이 결과에 따르면 제안한 알고리즘이 관심 모델의 변화에 관계없이 강인한 영상 분할 결과를 출력할 수 있다는 것을 알 수 있다.

또한 실제 자연에서 촬영한 영상 시퀀스를 이용한 실험 결과를 나타낸 그림 9에서는 카메라가 관심 객체인 새만을 추적하거나 개화되고 있는 꽃봉오리에 초점이 맞추어진 상황을 나타내고 있으며, 대부분의 프레임에서 포커스 된 객체의 추출이 원활히 이루어짐을 알 수 있었다.

4. 결 론

본 논문은 낮은 피사계 심도의 동영상 데이타를 이용하여 사용자의 도움 없이 비디오 객체만을 고속의 효율적인 방법으로 추출해 내는 알고리즘을 제안하였다. 정확한 영상 분할을 위해 기존에 제안했던 방법[1]을 개선하여 동영상 데이타의 첫 번째 프레임에 적용해 비디오 객체 분할을 위한 모델을 설정하였다. 그리고 고속의 동영상 객체 분할을 위해서 연속된 두 장의 프레임 사이의 시간의 중복성을 고려하여 효율적으로 관심 물체만을 추적하는 방법을 적용하였다. 포커스 된 비디오 객체는 현재 모델을 바탕으로 예측한 관심 영역만을 제한하여 계산함으로써 고속으로 추출 가능하다. 비디오 객체를 분할하는 속도는 영상의 포커스 된 관심 영역의 크기에 영향을 받는다.

향후 개선사항으로는 본 논문에서 제안한 동영상 객체 분할 시스템을 가상 현실(VR)이나 실감 방송과 같은

표 2 "Bream" 동영상에서 객체 추출의 정확도 분석

Bream Frames	Accuracy rate	Error rate (spatial distortion measures)
Steady change	94.3%	5.7%
Rapid change	92.2%	7.8%

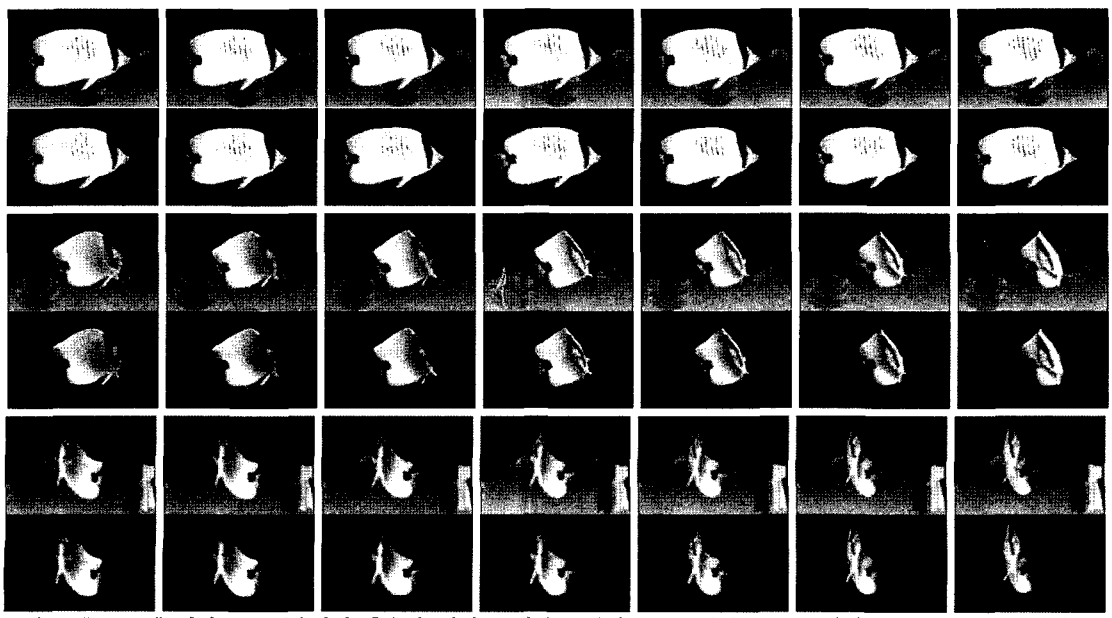


그림 8 "Bream" 시퀀스를 이용하여 추출한 비디오 객체 (1번째 줄: 프레임 1~7, 3번째 줄: 110~116, 5번째 줄: 216~222)

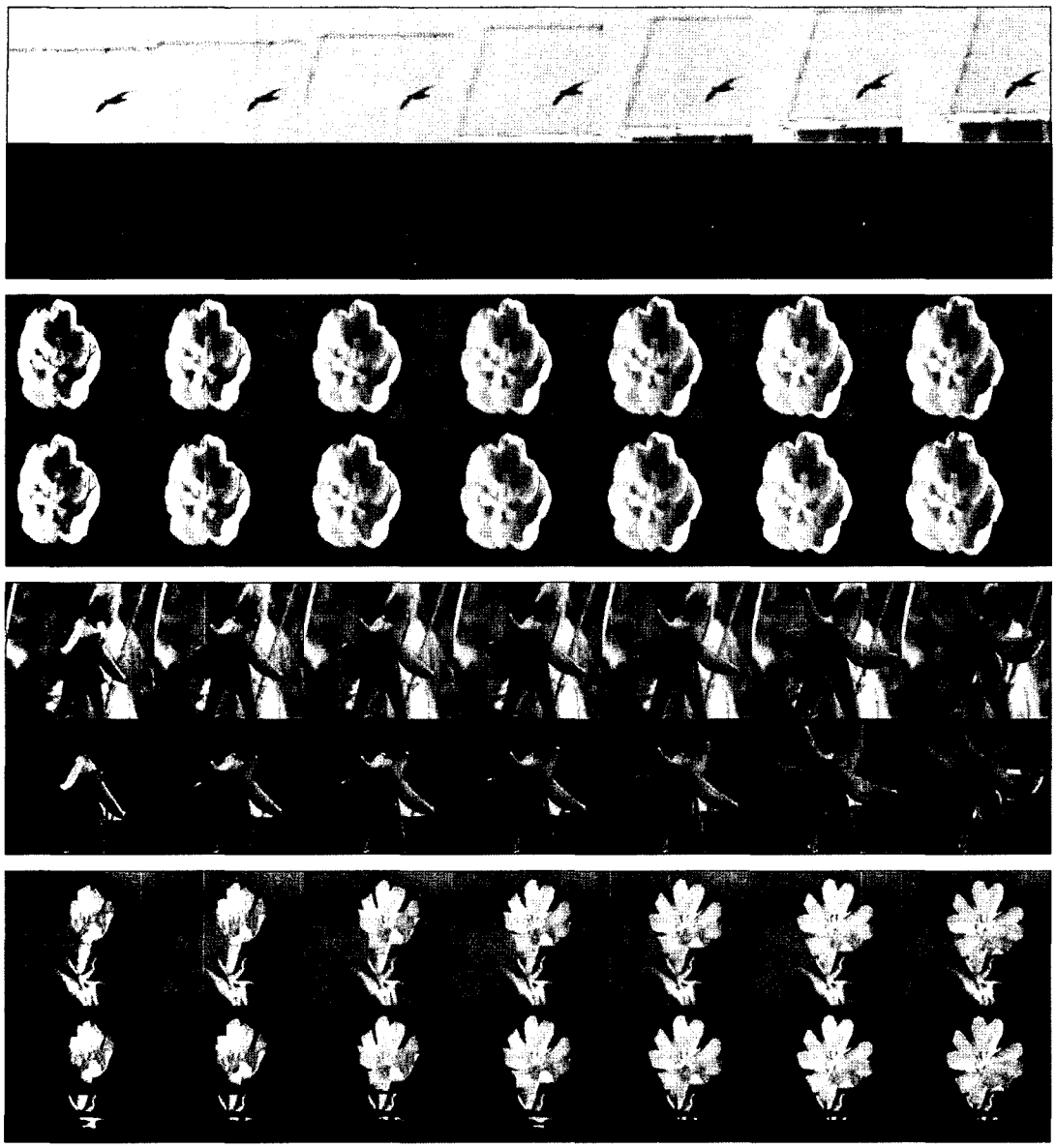


그림 9 실제 촬영 영상으로부터 추출한 비디오 객체

분야에 적용하는 것이다. 실시간 응용을 위해 코드와 알고리즘의 최적화를 통해 현재 계산 복잡도를 줄여 나아가는 것 또한 계속되고 있는 연구의 한 부분이다.

참 고 문 헌

[1] C. Kim, "Segmenting a Low Depth-of-Field Image Using Morphological Filters and Region Merging," *IEEE Tr. on Image Processing*, vol. 14, issue 10, pp. 1503-1511, Oct. 2005.

[2] J.Z. Wang, J. Li, R.M. Gray, and G. Wiederhold, "Unsupervised multiresolution segmentation for images with low depth of field," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no.1, pp. 85-90, Jan. 2001.

[3] J. Pan, S. Li, and Y. Zhang, "Automatic extraction of moving object using multiple features and multiple frames," in *Proc. of IEEE International Symposium on Circuits and Systems*, vol. 1, pp. 36-39, May. 2000

[4] C. Gu and M.C. Lee, "Semiautomatic Segmentation and Tracking of Semantic Video Objects," *IEEE*

Trans. Circuits Syst. Video Technol. VOL 8, NO. 5, Sept. 1998.

[5] M. Kass, A. Witkin, and D. Terzopoulos, "Snake: active contour model," in *Proc. of First International Conference on Computer Vision*, pp. 259-269, 1987

[6] P.J. Besl and R.C. Jain, "Segmentation through variable-order surface fitting," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 10, pp. 167-192, March 1988.

[7] L.M. Lifshitz and S.M. Pizer, "A multiresolution hierarchical approach to image segmentation based on intensity extrema," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 12, pp. 529-540, June 1990.

[8] D. Comaniciu, P. Meer, "Robust Analysis of Feature Spaces: Color Image Segmentation," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition* (CVPR'97), San Juan, Puerto Rico, 750-755, 1997.

[9] K. Aizawa, A. Kubota, K. Kodama, "Implicit 3D Approach to Image Generation: Object-Based Visual Effects by Linear Processing of Multiple Differently Focused Images," in *Proc. 10th International Workshop on Theoretical Foundations of Computer Vision*, Vol. 2032, pp. 226-237, Dagstuhl Castle, Germany, March 2000.

[10] C. Kim and J.-N. Hwang, "Video Object Extraction for Object-Oriented Applications," *Journal of VLSI Signal Processing - Systems for Signal, Image, and Video Technology*, Special Issue on Multimedia Signal Processing, vol. 29, no.1/2, pp. 7-21, August 2001.

[11] Ju Guo, J. Kim, and C.-C. Jaykuo, "Fast and Accurate Moving Object Extraction Technique for MPEG-4 Object-Based Video Coding," in *Proc. SPIE*, vol. 3653, pp. 1210-1221, 1999.

[12] M. Kim, J.G. Choi, D. Kim, H. Lee, M.H. Lee, and Y. Ho, "A VOP Generation Tool: Automatic Segmentation of Moving Objects in Image Sequences Based on Spatio-Temporal Information," *IEEE Trans. Circuits Syst. Video Technology*, vol. 9, no. 8, 1999.

[13] G. Borgefors, "Distance Transformations in Digital Images," *Computer Vision, Graphics, and Image Processing*, vol. 34, pp. 344-371, 1986.

[14] M. Bierling, "Displacement estimation by hierarchical blockmatching," in *Proc. SPIE Visual Commun. Image Processing*, VCIP'88, vol. 1001, pp. 942-951, Cambridge, MA, Nov. 1988.

[15] M. Wollborn and R. Mech, "Refined procedure for objective evaluation of video generation algorithms," Doc. ISO/IEC JTC1/SC29/WG11 M3448, March 1998.

[16] G. Gelle, M. Colas, G. Delaunay, "Higher Order Statistics for Detection and Classification of Faulty Fanbelts Using Acoustical Analysis," in *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics* (SPW-HOS '97), pp. 43-46, Banff, Canada, July 21-23, 1997.

[17] P. Salembier and M. Pardas, "Hierarchical Morphological segmentation for Image sequence Coding," *IEEE Transactions on Image Processing*, vol. 3, no. 5, pp. 639-651, Sept. 1994.

박 정 우

2003년 2월 성균관대학교 정보통신공학부(학사). 2006년 2월 한국정보통신대학교(ICU) 공학부(석사). 2006년 2월~현재 삼성전자 기술총괄 생산기술연구소 연구원. 관심분야는 HCI, 컴퓨터비전, 기계학습, 유비쿼터스 컴퓨팅

김 창 익

1989년 2월 연세대학교 전기공학과(학사). 1991년 2월 포항공과대학교(POSTECH) 전기전자공학과(석사). 1991년 1월~1997년 7월 SKC Ltd. R&D 센터 선임 연구원. 2000년 12월 워싱턴주립대학교 전기공학과(박사). 2000년 12월~2005년 1월 Senior Member of Technical Staff, Epson Palo Alto Laboratory, Epson R&D Inc. 2005년 2월~현재 한국정보통신대학교(ICU) 공학부 조교수. 지능형 비디오, 3D 비디오, Next Generation Video Communication Systems, Multimedia Signal Processing, Digital TV Broadcasting, Video Analysis