

논문 2011-48SP-1-5

윤곽선 정보를 이용한 동영상에서의 객체 추출

(Video Object Extraction Using Contour Information)

김재광*, 이재호*, 김창익**

(Jaekwang Kim, Jaeho Lee, and Changick Kim)

요약

본 논문에서는 객체의 윤곽선 정보에 기반한 수정된 그래프컷(Graph-cut) 알고리즘을 이용하여 동영상에서 효율적으로 객체를 추출하는 방법을 제안한다. 이를 위해 먼저, 첫 프레임에서 자동 추출 알고리즘을 이용하거나 사용자와의 상호작용을 통해 영상에서 객체를 분리한다. 객체의 형태 정보를 상속시키기 위해 이전 프레임에서 추출된 객체 윤곽선의 움직임을 예측한다. 예측된 윤곽선을 기준으로 블록 단위 히스토그램 역투영(Block-based Histogram Back-projection) 알고리즘을 수행하여 다음 프레임의 각 픽셀에 대한 객체와 배경의 컬러 모델을 형성한다. 또한, 윤곽선을 중심으로 전체 영상에 대한 로그함수 기반의 거리 변환 지도(Distance Transform Map)를 생성하고 인접 픽셀간의 연결(link)의 확률을 결정한다. 생성된 컬러 모델과 거리 변환 지도를 이용하여 그래프를 형성하고 에너지를 정의하며, 이를 최소화하는 과정을 통해 객체를 추출한다. 다양한 영상들에 대한 실험 결과를 통해서 기존의 객체 추출 방법보다 제안하는 방법이 객체를 보다 정확하게 추출함을 확인할 수 있다.

Abstract

In this paper, we present a method for extracting video objects efficiently by using the modified graph cut algorithm based on contour information. First, we extract objects at the first frame by an automatic object extraction algorithm or the user interaction. To estimate the objects' contours at the current frame, motion information of objects' contour in the previous frame is analyzed. Block-based histogram back-projection is conducted along the estimated contour point. Each color model of objects and background can be generated from back-projection images. The probabilities of links between neighboring pixels are decided by the logarithmic based distance transform map obtained from the estimated contour image. Energy of the graph is defined by predefined color models and logarithmic distance transform map. Finally, the object is extracted by minimizing the energy. Experimental results of various test images show that our algorithm works more accurately than other methods.

Keywords : Video Object Segmentation, Graph Cut, Histogram Back-projection, Distance Transform

I. 서론

동영상 기반의 시스템과 이의 정보전달이 보편화되면서, 컴퓨터 비전 분야에서도 하나의 프레임이 아닌 연속적인 프레임에 대한 연계성을 보장하는 편집 및 검

색, 데이터 처리 기술이 중요하게 연구되고 있다. 대용량 멀티미디어 데이터를 다룰 때에는 압축, 검색(Retrieval) 및 색인(Indexing) 모델을 형성하기 위해 콘텐츠 기반의 기능(content-based functionality)을 제공하는 것이 필수적이기 때문에, 동일한 장면을 구성하는 프레임들에 대한 콘텐츠 기반 표현은 중요한 연구분야로 자리매김해 왔다^[1-2]. 한 장면 내에서는 프레임간의 컬러와 텍스처, 움직임 등 전역적인 특징들이 유사하기 때문에, 장면은 비디오 부호화, 검색 등 멀티미디어 신호처리 분야에서 하나의 큰 단위로서의 역할을 한다^[1]. 그러므로 장면의 콘텐츠 기반 표현을 위한 정확한 객체 추출 기법은 다양한 동영상 응용분야의 성능을 좌우하

* 학생회원, ** 정회원-교신저자, 한국과학기술원 전기 및 전자공학부

(School of Electrical Engineering, Korea Advanced Institute of Science and Technology)

※ 본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었음 (NIPA-2010-(C1090-1011-0003))

접수일자: 2010년7월16일, 수정완료일2010년9월30일

는 핵심적인 기술로 연구되어 왔다. 그러나 객체의 정밀한 추출은 복잡한 배경 영역 및 객체와 배경 사이의 모호한 구분 등으로 인해 도전적인 과제로 평가되고 있다.

그간 동영상에서의 객체 추출은 크게 두 가지 방향으로 연구가 진행되어 왔다. 물체의 추적(tracking) 알고리즘을 이용하는 것과 기존의 객체 추출 알고리즘을 영상의 특성에 기초해 변형하여 사용하는 것이다. 물체의 추적 알고리즘은 사용되는 특징의 종류에 따라 점 추적(Point Tracking), 커널 추적(Kernel Tracking), 실루엣 추적(Silhouette Tracking)으로 나누어진다. 점 추적과 커널 추적은 객체의 위치를 파악하는 것이 주목적이기 때문에 객체 형태의 정확한 추출을 위해서는 추가적인 알고리즘이 필요하게 된다^[3~4]. 실루엣 추적 방식에서는 이전 프레임에서 추출된 객체의 형태를 이용해 실루엣을 정의하고, 다음 프레임에서 이를 매칭(matching)하거나 윤곽선을 추적함으로써 객체를 추출해 낸다. [5]에서는 배경을 제외시켜 얻은 후보 객체들의 실루엣을 이전 프레임의 실루엣과 비교하면서 가장 유사한 객체를 선택한다. [6]에서는 윤곽선 근처 에지(edge)를 확장하여 모델링(modeling)된 에너지 함수가 최소가 되는 윤곽선을 찾아 추적함으로써 객체를 추출한다. 그러나 객체 추적 알고리즘은 객체가 가려지는 (occluded) 경우에 대한 판단과 복구에 어려움이 있고, 객체의 움직임이 큰 경우에도 취약하다는 단점이 있다.

정지 영상에서의 객체 추출 알고리즘들을 동영상에 적용시키는 방법도 연구되어 왔다. 그래프 컷 알고리즘을 동영상에 적용하기 위해 3차원 그래프를 형성하여 각 프레임간의 노드(node)의 관계가 포함된 에너지를 최소화하는 방법이 있다^[7~8]. 기존의 정지영상에서의 그래프를 시간에 관한 차원으로 확대함으로써, 공간 차원에서의 노드 간의 관계 뿐만 아니라 시간 차원에서의 노드 간의 관계를 함께 고려하여 에너지 함수를 구성한다. 이 때, 워터셰드(watershed) 알고리즘을 적용시켜 영상을 의미 있는 영역들의 집합으로 나타냄으로써 계산량을 감소시킨다^[7]. 그러나 이 같은 3차원의 그래프 컷을 이용한 방법은 미리 분할된 의미 있는 영역이 잘못되었을 경우 인접 영역에서의 오류를 유발하기도 한다. 객체의 윤곽선을 이용하는 방법으로는 에지에 기반을 둔 액티브 컨투어 모델(active contour model)이 많이 사용된다. [9]에서는 변형된 스네이크(snake) 알고리즘인 V스네이크(Vsnake) 알고리즘을 정의하고 이전 프레임에서 추출된 객체를 기초로 다음 프레임의 윤곽선

을 예측하여 객체를 추출한다. V스네이크는 프레임 사이의 윤곽선의 차이를 이용하여 에너지 함수를 생성하고 이를 최소화하여 객체를 추출한다. 또한 레벨셋(Level Set) 알고리즘을 이용하여 윤곽선을 찾아 객체를 추출하는 방법도 있다^[10~11]. 레벨셋 알고리즘에서 높은 계산량을 가지는 기존의 편미분방정식(partial differential equation)의 해(solution)를 찾아 커브를 확장(evolution)해 나가던 방식을 사용하지 않고, 새로 정의된 두 번의 순환과정으로 계산속도를 높이기도 한다^[10]. 하지만 실제 자연 영상에서 에지는 객체의 윤곽선 뿐만 아니라 배경 영역에서도 많이 발생하므로, 윤곽선 주변에서 발생하는 에지에 의해 성능이 크게 영향을 받게 되는 단점이 있다. 즉, 영상의 배경이 복잡할수록 정확한 객체 추출이 어려워진다.

본 논문에서는 객체의 윤곽선 정보를 활용하도록 그래프 컷 알고리즘을 수정하여 연속된 프레임에서 안정적으로 객체를 추출하는 방법을 제안한다. 그래프 컷 알고리즘은 주변의 컬러 정보에 민감하게 반응하여 컬러의 변화가 큰 영역을 경계로 객체를 추출하기 때문에 배경과 객체 사이에 컬러의 변화가 작은 경우 객체의 형태를 온전히 보전하기 어렵다. 제안하는 방법에서는 먼저 키 프레임(key frame)의 객체를 자동 추출 알고리즘^[12]이나 사용자와의 상호작용으로 추출한다^[13]. 나머지 프레임들에 대해서는 윤곽선 정보를 기초로 한 컬러 모델과 형태 지도를 생성한 후, 그래프 컷 알고리즘을 통해 객체를 추출한다. 먼저 형태 정보의 상속을 위해 이전 프레임에서 추출된 객체의 윤곽선 정보로부터 움직임을 추적하여 현재 프레임의 윤곽선 지도를 형성한다. 기존의 그래프 컷은 전역적인(global) 컬러 모델의 형성으로 인한 객체 주변의 컬러 정보에 민감하게 반응하는데 이를 보완하기 위해 본 논문에서는 윤곽선 정보를 기초로 지역적인(local) 컬러 모델을 형성한다. 예측된 윤곽선을 중심으로 블록 단위의 히스토그램 역투영을 통해서 객체와 배경에 컬러 모델을 형성한다. 윤곽선 주변의 컬러 정보만을 사용하기 때문에, 불필요한 컬러 정보로 인한 오류를 줄일 수 있게 된다. 또한, 윤곽선 지도를 기준으로 거리 변환하여 형태 지도를 형성하고, 예측된 윤곽선에 가까운 링크(link)의 확률값을 작게 설정하여 다음 프레임에서 컷(cut)될 확률을 높인다. 마지막으로 에너지 함수를 제안하고 에너지를 최소화하는 영역을 객체로 추출한다.

제안하는 알고리즘은 움직임 추적과 거리 변환을 통

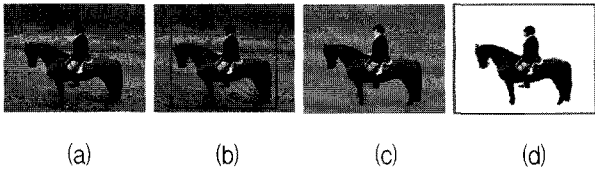


그림 1. 그라브컷(GrabCut)을 이용한 객체 추출. (a) 입력영상 (b) 사용자 입력 (c) 추출된 결과에 대한 사용자 후처리 (d) 최종 추출 결과
 Fig. 1. Object extraction by using GrabCut. (a) input image (b) user interaction (c) post processing of the initial result by user (d) final result.

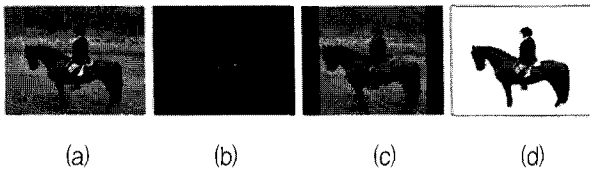


그림 2. 자동 객체 추출. (a) 입력 영상, (b) 특이점 추출 결과, (c) 특이점을 이용한 시드(seed) 부여, (d) 객체 추출 결과
 Fig. 2. Automatic object extraction. (a) input image (b) result of saliency extraction (c) seed extraction by saliency (d) result of object extraction.

해 객체의 형태를 예측한다. 그렇기 때문에 기존의 방법과 달리 움직임이 많은 객체에 대해서도 효과적으로 추출해낸다. 또한, 지역적인 컬러 모델의 형성으로 불필요한 컬러 정보의 사용을 줄여 객체와 배경의 컬러가 유사한 경우에도 안정적인 성능을 보인다.

본 논문의 구성은 다음과 같다. 먼저 II장에서는 첫 프레임에서 객체를 추출하기 위한 기존의 방법들에 대해 간략히 설명한다. III장에서는 객체의 윤곽선 정보를 상속하기 위한 방법에 대해 설명하고 제안하는 에너지 최소화 함수를 정의한다. IV장에서는 다양한 실험 영상을 통해 제안하는 알고리즘의 강건함과 효율성을 보인다. 마지막으로 V장에서 결론 및 향후 과제에 대해 논의한다.

II. 첫 프레임에서의 객체 추출

장면의 첫 프레임에서는 기존의 정지 영상에서 객체를 추출하는 방법을 이용하였으며, 이는 크게 사용자의 상호 작용에 기반한 방법과 자동으로 객체를 추출하는 방법의 2가지가 있다.

주어진 영상으로부터 그래프를 형성하기 위해서는 각 픽셀에서 배경 영역과 객체 영역에 대한 사전 정보가 필요하다. 그라브컷(GrabCut)^[14]에서는 사용자의 입

력에 의거하여 객체와 배경 영역에 대한 사전 정보를 획득한다. 사용자의 입력을 바탕으로 각 영역에 대한 시드(seed)를 형성한 후, 이를 가우시안 혼합 모델(Gaussian Mixture Model)로 나타내어 각 픽셀에서의 객체와 배경에 대한 사전 확률을 정의할 수 있다. 이렇게 획득된 사전 정보와 노드 간의 관계를 바탕으로 객체를 추출한다. 또한 사용자의 후처리를 통하여 정밀한 형태의 객체를 얻을 수 있다. 그림 1에 그라브컷을 이용한 객체 추출의 예를 나타내었다.

자동으로 객체를 추출하기 위해서는 입력 영상에서 자동으로 시드를 추출하는 과정이 선행되어야 한다. 이를 위하여 영상에서 특이점(saliency)을 이용하는 방법^[12]이 연구되었다. [12]에서는 스펙트럼 여분(spectral residual)^[15]을 이용하여 영상에서 관심이 있는 영역을 자동으로 찾은 후, 이를 바탕으로 시드를 부여하고 그래프 컷의 영역 특성 항(region property term)의 결정을 위해 가우시안 혼합 모델을 형성한다. 이 때, 특이점을 통해 얻어진 시드는 불완전하므로 마르코프 체인 룰(markov chain rule)을 이용하여 시드로 결정된 픽셀들이 그래프 컷 결과에 따라 상태(state)를 바꿀 수 있게 한다. 상태의 변화가 생기지 않는 시점까지 반복적으로 가우시안 혼합 모델을 학습(learning)시키며 객체의 추출을 진행한다. 이러한 과정을 이용하여 주어진 영상에서 관심이 있는 객체 영역을 자동으로 추출한 결과를 그림 2에 나타내었다.

위와 같은 방법들을 이용해 첫 프레임에서의 객체를 추출하는 과정에는 각각의 장단점이 있다. 그라브컷을 이용할 경우에는 객체 영역을 정교하게 추출해 낼 수 있지만, 매 장면마다 사용자가 직접 시드를 설정해주어야 하는 번거로움이 있다. 반면 특이점을 이용할 경우에는 모든 장면에 대해 자동으로 객체를 추출할 수 있지만, 특이점의 정확도에 따라 추출되는 객체의 정확성이 좌우되기 때문에 원하지 않는 결과가 나타날 수 있다. 응용목적에 따라 첫 프레임에서의 객체 추출 방법은 선택적으로 결정될 수 있다.

III. 동영상에서의 객체 추출

먼저 동영상의 t 번째 프레임을 P_t 와 같이 나타낸다. 또한, 각 해당 프레임에서 추출된 객체의 영역을 O_t 로 배경의 영역을 B_t 로 표시한다. 즉,

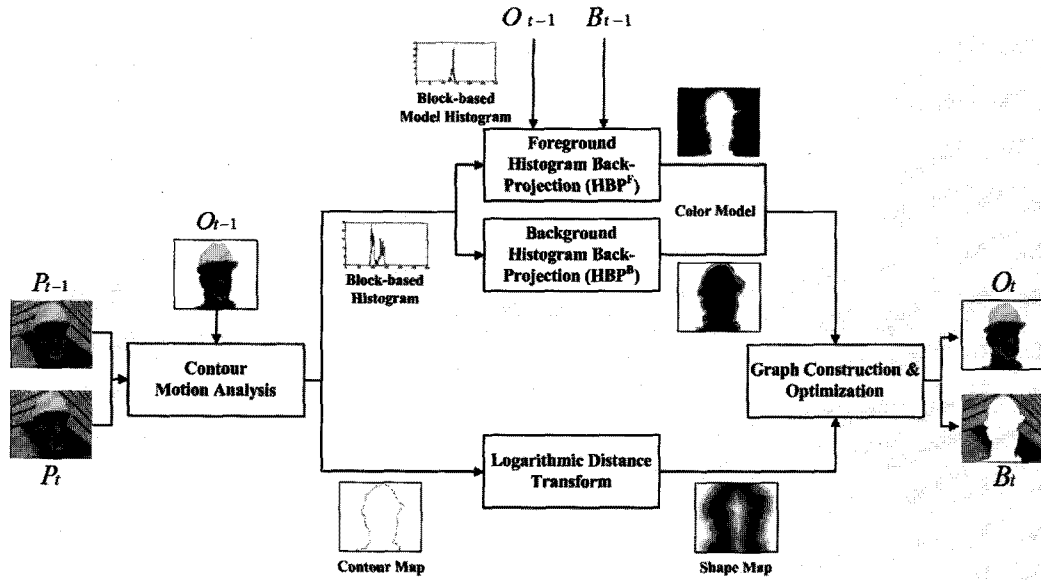


그림 3. 제안하는 시스템의 블록 다이어그램
Fig. 3. Block diagram of the proposed system.

$$P_t = O_t \cup B_t, \quad O_t \cap B_t = \emptyset \quad (1)$$

그림 3에 제안하는 알고리즘에 대한 블록 다이어그램이 있다. 현재 프레임(P_t)에 대한 객체(O_t) 추출은 이전 프레임(P_{t-1})의 안정적인 결과(O_{t-1}, B_{t-1})를 최대한 활용할 수 있게 구성된다. 이전 프레임과 추출 결과를 이용하여 현재 프레임의 윤곽선 지도 C_t 를 추정한다. 윤곽선을 중심으로 컬러모델을 형성하고 현재 프레임의 모든 픽셀의 사전 확률을 결정한다. 또한, 윤곽선 지도를 거리 변환하여 형태 지도 S_t 를 형성하고, 윤곽선 부근에 가까운 그래프 컷의 인접 픽셀간의 연결(n-link)일수록 가중치를 부여하여 에너지의 최소화에 의해 잘려지게 한다.

1. 윤곽선 지도의 형성

일반적으로 프레임 사이의 객체의 형태는 크게 변하지 않고 높은 연관성(correlation)을 유지한다. 그러므로 이전 프레임에서 추출된 객체의 형태는 현재 프레임의 객체 추출의 중요한 사전 정보(prior knowledge)로 활용될 수 있다. 객체의 형태정보의 상속을 위해 다음과 같은 과정을 거친다.

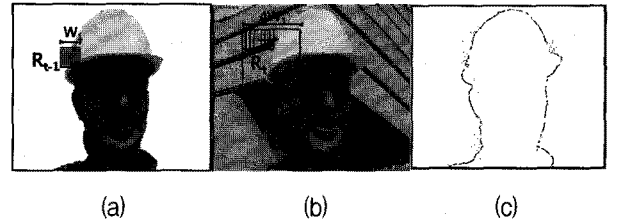


그림 4. 윤곽선 지도의 형성. (a) 이전 프레임 결과 (b) 현재 프레임에서 윤곽의 움직임 추적 (c) 예측된 윤곽선 지도
Fig. 4. Contour map construction. (a) result of prior frame (b) contour motion analysis on current frame (c) contour map.

그림 4에서 보면 이전 프레임에서 구해진 객체의 윤곽선 부분을 따라서 $W_1 \times W_1$ 크기의 블록을 이용하여 현재 프레임에서 움직임을 추정(Block-based Motion Estimation)한다. 이때, 움직임을 추정하는 검색 범위(search range)는 $4W_1 \times 4W_1$ 를 사용한다. 검색 범위 내에서 식 (2)와 같이 SAD_{\min} 을 가지는 블록을 찾아낸다.

R_{t-1} 은 이전 프레임에서 예측하고자 하는 기준 블록이고, R_t 는 검색 범위 내의 블록들이다. 검색 범위 내에서 최소값의 SAD 를 가지는 블록을 예측하면 이

$$SAD(x, y, r, c) = \sum_{i=0}^{W_1} \sum_{j=0}^{W_1} |R_{t-1}(x+i, y+j) - R_t((x+r)+i, (y+c)+j)| \quad (2)$$

전 프레임의 윤곽선 좌표 (x, y) 는 예측된 블록의 중심 좌표인 (r, c) 로 예측된다. 그림 4의 (c)와 같이 예측된 현재 프레임의 윤곽선 지도 C_t 는 예측된 $(r, c) \in C_t$ 의 모임으로 불연속적으로(discrete) 형성된다.

2. 컬러 모델의 형성

그래프 컷을 이용한 객체 추출 알고리즘에서는 컬러 정보를 이용하여 객체와 배경에 대한 확률밀도함수(probability density function)를 정의하고 픽셀의 객체와 배경 터미널과의 연결(t-link)의 에너지로 확장한다^[13]. 본 논문에서는 이 같은 객체와 배경의 확률밀도함수, $P(p|Obj)$ 와 $P(p|Bkg)$ 를 각각 객체와 배경에 대한 컬러 모델이라고 정의한다. 기존의 그래프 컷을 이용한 객체 추출 알고리즘은 컬러 모델을 형성하기 위해 R, G, B 3개의 채널을 이용한 3차원 컬러 공간의 가우시안 혼합 모델을 이용하였다^[14~16]. 그러나 특징 벡터들의 컬러 정보만을 이용한 가우시안 혼합 모델의 형성은 영상을 컬러의 변화를 경계로 객체와 배경으로 이분화할 가능성이 크기 때문에 온전한 객체 추출이 어렵

다. 그래서 이전 연구에서는 위치 정보를 포함하여 가우시안 혼합 모델의 차원을 확장시킴으로써 유사한 컬러라도 위치에 따라 다른 컬러로 인식되게 하는 방법을 제안하였다^[17]. 그러나 이 같은 방법 역시 컬러 모델이 전역적으로 형성되기 때문에 객체와 인접 배경의 컬러가 유사할 경우 성능이 저하된다. 그러므로 본 논문에서는 예측된 윤곽선 주변을 따라 블록 단위의 객체와 배경에 대한 지역적인 컬러 모델을 형성한다.

히스토그램 역투영을 위해서는 역투영할 목표 영상의 히스토그램 $I(i)$ 와 모델 히스토그램 $M(i)$ 이 필요하게 된다^[18]. 먼저 객체의 히스토그램 역투영을 수행하는 경우, 그림 5와 같이 목표 영상의 히스토그램은 예측된 윤곽선이 중심인 $W_2 \times W_2$ 사이즈의 블록 HB_t 의 히스토그램에 해당하고, 객체의 모델 히스토그램은 예측되기 전 윤곽선을 중심으로 한 같은 크기 블록 HB_{t-1} 중 객체에 해당하는 영역의 히스토그램이다. 배경의 경우는 이와 유사하게 HB_{t-1} 중 배경에 해당하는 영역의 히스토그램이다. 윤곽선을 따라 블록을 계속 형성해 나가며 생기는 계산량을 줄이기 위해 3차원의 컬러 히

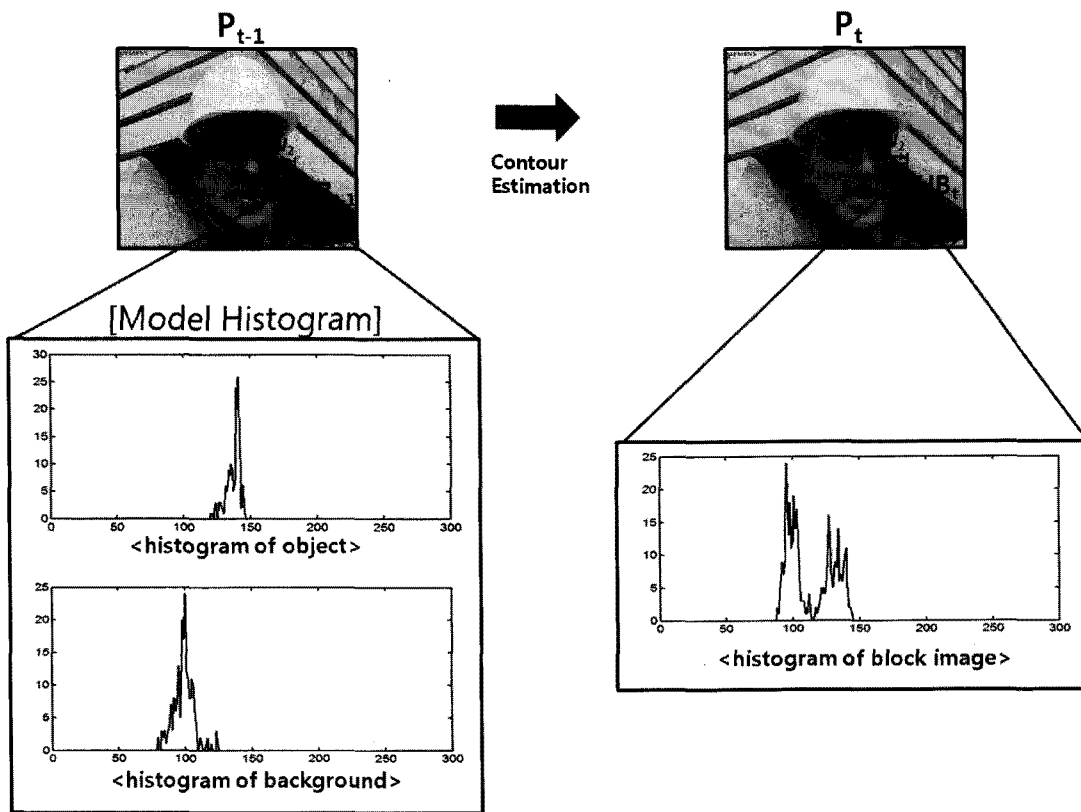


그림 5. 블록 단위의 히스토그램 역투영시 필요한 히스토그램들의 취득과정
 Fig. 5. Process of histogram acquisition for block-based histogram back-projection.

스토그램을 사용하지 않고, 영상의 밝기(intensity)값만을 사용한 1차원의 히스토그램을 사용하였다. 목표 영상의 히스토그램과 모델 히스토그램을 이용하여 식 (3)과 같이 비율 히스토그램 HBP 를 얻는다.

$$HBP(i) = \min\left[\frac{M(i)}{I(i)}, 1\right], \quad (3)$$

여기에서 i 는 밝기값을 나타낸다. 이렇게 얻어진 HBP 를 윤곽선지도의 해당 블록에 역투영하게 된다.

각 블록 내의 픽셀들은 그림 6의 (a)와 같이 불연속적인 윤곽선을 따라 0에서 1사이의 HBP 의 값을 가지게 된다.

식 (4)와 같이 겹쳐지는 영역을 고려하여 누적되는 값들의 평균값을 취함으로써 컬러 모델을 정의한다.

$$P(p|Obj) = \frac{1}{n} \sum_{j=1}^n HBP_j^F(i_p), \quad p \in R_c \quad (4)$$

$$P(p|Bkg) = \frac{1}{n} \sum_{j=1}^n HBP_j^B(i_p), \quad p \in R_c$$

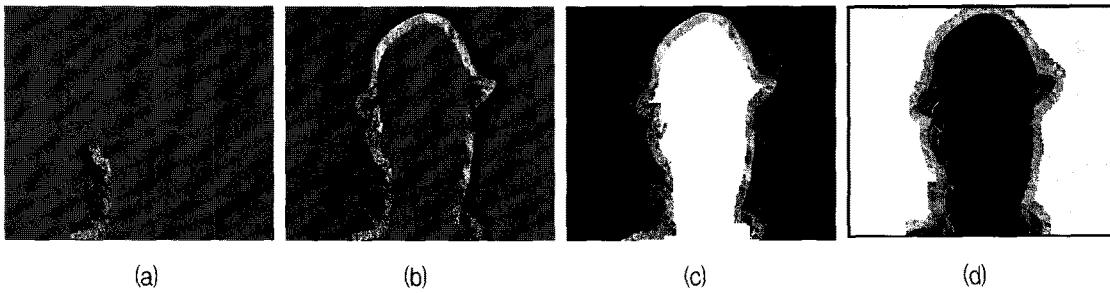


그림 6. 컬러 모델 형성 과정. 확률값이 1에 가까울수록 밝기값을 가장 밝게 하여 표현하였다. (a) 블록 단위 히스토그램 역투영 진행(객체의 경우만 명시) (b) 블록 단위 히스토그램 역투영을 통한 윤곽선 부근 컬러 모델 형성(객체의 경우만 명시) (c) 최종적인 객체의 컬러 모델 (d) 최종적인 배경의 컬러 모델

Fig. 6. Process of constructing the color models. The high intensity value is represented as the high probability. (a) block-based histogram back-projection(only object case.) (b) color model construction close-by contour(only object case.) (c) object color model (d) background color model.

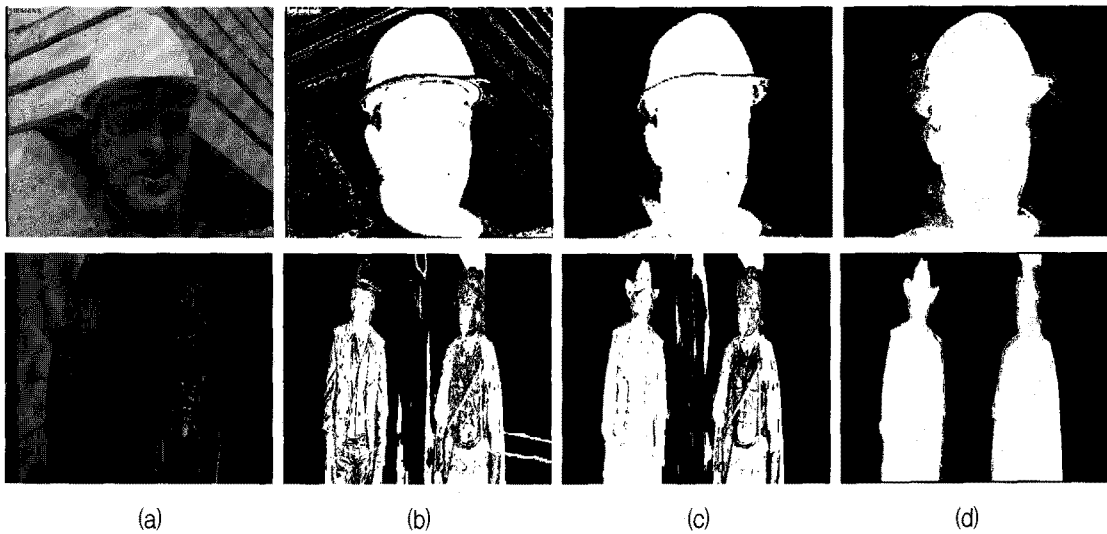


그림 7. 타 컬러 모델들과의 비교. 위에서부터 Foreman 영상과 cowboy and knight 영상. (a), 원본 영상, (b) 컬러 정보만을 이용한 가우시안 혼합 모델을 사용한 결과 (c) 컬러와 위치 정보를 이용한 가우시안 혼합 모델을 사용한 결과 (d) 제안하는 컬러 모델을 사용한 결과

Fig. 7. Comparison with other color models. from top to bottom, Foreman and cowboy & knight image. (a) original image(current frame) (b) result in which the regions are decided to object or background by color GMM (c) by 5 domain(color and location) GMM (d) by the proposed color model.

여기서 HBP^F 와 HBP^B 는 각각 객체와 배경 비율 히스토그램을 의미하며, n 은 p 가 속하는 영역의 개수, 즉, 움직임 추적된 블록의 개수를 의미한다. 그림 6의 (b)와 같이 윤곽선 주변의 영역 R_c 들은 객체와 배경에 대한 확률값을 가진다. 나머지 영역에 대해서는 식 (5)와 같이 픽셀단위로 이전 프레임의 객체와 배경임을 확인하여 객체와 배경에 대한 우도(likelihood)를 가진다.

$$P(p|Obj) = \begin{cases} 1, & \text{if } p \in O_{t-1} \text{ and } p \notin R_c \\ 0, & \text{if } p \in B_{t-1} \text{ and } p \notin R_c \end{cases} \quad (5)$$

$$P(p|Bkg) = \begin{cases} 0, & \text{if } p \in O_{t-1} \text{ and } p \notin R_c \\ 1, & \text{if } p \in B_{t-1} \text{ and } p \notin R_c \end{cases}$$

이렇게 픽셀단위로 얻어진 각각의 값들은 그림 6의 (c), (d)와 같은 컬러 모델을 형성한다. 형성된 컬러 모델만으로도 객체와 배경에 대한 구분이 이루어지는 것을 확인할 수 있다.

컬러모델의 성능을 확인하기 위해 이전 프레임의 결과를 3차원의 컬러 도메인을 가지는 가우시안 혼합 모델을 이용하여 모델링하는 경우^[14~16]와 컬러와 위치정보를 포함한 5차원의 가우시안 혼합 모델을 사용하는 경우^[17]를 비교하였다. 5개의 가우시안 분포가 혼합된 결과를 그림 7에 나타내었다. 각 영상들은 이전 프레임의 객체와 배경 영역을 기준으로 모델링 되었고, 객체일 확률값과 배경일 확률값을 비교하여 이진화 하였다. 위치정보를 포함한 가우시안 혼합 모델의 경우, 객체와 배경에 대한 구분이 어느 정도 이루어지지만, 배경과 객체의 컬러가 유사한 경우에는 성능이 떨어지는 것을 알 수 있다. 본 논문에서 제안하는 방법으로는 객체와 배경의 윤곽선 부근 내에서의 밝기 정보만을 사용하여 컬러 모델을 형성하기 때문에, 유사한 컬러에서도 객체와 배경의 차이를 찾아내는 것이 가능하다. 형성된 컬러모델의 확률값은 에너지 최소화 함수에 반영된다.

3. 형태 지도의 형성

3.1에서 형성한 불연속적인 윤곽선 지도는 객체의 전체적인 형태에 대한 확률값을 정의하기에 부적합하다. 그러므로 식 (6)과 같은 거리 변환(Distance Transform)^[19] 알고리즘을 이용하여 각 픽셀에 대한 연속적인(continuous) 확률값 D_p 를 정의하고 이는 그림 8의 (b)와 같이 나타난다.

$$D_p = \min_{c \in C_t} D(p, c), \quad c \in C_t \quad (6)$$

여기에서 $D(p, c)$ 는 픽셀 p 와 c 사이의 거리이다. 윤곽선의 D_p 는 0의 값을 가지게 되고, 윤곽선으로부터의 거리에 따라 각 픽셀은 Chessboard Distance를 적용하여 값을 가지게 된다^[20]. [21]에서는 거리 변환을 통해 얻은 연속적인 확률값을 그래프 컷의 영역 특성 항(region property term)을 설정하는 데 컬러 모델과 함께 사용하였다. 그러나 거리 변환을 통한 확률값은 점진적인(gradual) 변화를 가지기 때문에 급격히 변하는 컬러모델과 함께 사용할 경우, 형태보다는 컬러에 더욱 민감하게 객체의 추출 윤곽이 결정되게 된다. 그러므로 우리는 식 (7)과 같이 로그 함수에 기반한 거리변환을 적용하여 형태 지도 $s_p \in S_t$ 를 완성한다.

$$s_p = \ln(D_p + 1) \quad (7)$$

여기에서 로그 스케일을 취함으로써, 예측된 윤곽선에 근접하는 픽셀일수록 윤곽선이 될 상대적인 확률값을 높인다. 이렇게 완성된 형태 지도는 그림 8의 (c)처럼 나타나며, 이는 경계 특성 항(boundary property term)의 확률값에 적용되어 에너지 최소화 함수에 반영된다.

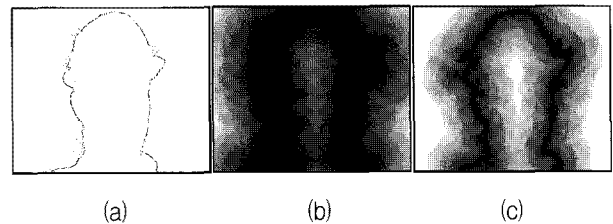


그림 8. 형태 정보의 상속. (a) 예측된 윤곽선 지도 (b) 거리 변환 후의 윤곽선 지도의 변화 (c) 현재 프레임의 형태 지도

Fig. 8. Coherence of the object's shape. (a) estimated contour map (b) transformed contour map after distance transform (c) shape map for the current frame.

4. 에너지 최소화 함수

앞에서 형성된 컬러 모델과 형태 지도는 현재 프레임의 그래프를 형성하는데 영향을 주게 된다. 그래프 $G = \langle V, E \rangle$ 는 노드(Node, V)와 에지(edge, E)로 이루어진다^[13]. 노드들은 프레임 내의 모든 픽셀들과 터미널이라 불리는 S, T 노드를 포함한다. S 터미널은 객체를 대표하고, T 터미널은 배경을 대표하며 각 픽셀은 터미널과의 연결을 통해 객체와 배경에 대한 사전 확률값을 결정하게 된다. 또한, 에지들은 인접노드들과의 연결인 $n-link$ 와 터미널과의 연결인 $t-link$ 에

$$E(G) = \sum_{p \in P_t} E_c(p) + \alpha \times \sum_{(p,q) \in N} E_s(p,q) + \beta \times \sum_{(p,q) \in N} E_n(p,q) \quad (9)$$

의한 에지들로 정의된다. 그래프 컷 알고리즘은 n -link와 t -link에 가중치 w_e 를 부여하고, 에지의 가중치의 합을 최소화하는 $C \subset E$ 를 찾아서 두 터미널을 경계로 노드들을 분할하게 된다.

$$|C| = \sum_{e \in C} w_e \quad (8)$$

우리는 3.2와 3.3의 과정을 거쳐 얻은 컬러, 형태에 관한 정보들을 에지의 가중치를 부여하는데 적용하고 최적의 분할을 찾는다. 그래프의 에너지를 최소화 하여 객체 영역을 추출하기 위해서 에너지를 식 (9)과 같이 정의한다.

그래프를 이루는 에너지는 링크에 부여된 3가지 에너지의 합으로 정의된다. $E_c(p)$ 는 각 픽셀이 컬러모델을 기초로 가지는 사전 확률을 통해서 정의된다. $E_s(p,q)$ 는 앞서 설명한 형태 지도를 통해 얻는 픽셀들의 확률값이 링크의 에너지로 확장되어 얻어진다. $E_n(p,q)$ 는 기존의 그래프 컷 알고리즘의 경계 특성 항 (boundary property term)에 해당한다^[13]. α 와 β 는 $E_s(p,q)$ 와 $E_n(p,q)$ 에 대한 가중치 상수(weighting factor)를 나타내며 본 논문에서는 실험을 통해 각각 50과 10을 사용하였다.

$E_c(p)$ 는 식 (10)과 같이 두 터미널에 대한 각각의 에너지로 이루어진다.

$$E_c(p) = E_c(p, S) + E_c(p, T) \quad (10)$$

여기에서 두 에너지는 컬러 모델을 통해 얻은 각 픽셀의 객체와 배경에 대한 우도로부터 이끌어낸다.

$$E_c(p, S) = -\ln P(p|Bkg) \quad (11)$$

$$E_c(p, T) = -\ln P(p|Obj)$$

각 에너지는 상반되는 터미널일 확률의 역 로그 (Negative log-likelihood) 함수로 모델링 된다. 이로 인해 객체일 확률이 클수록 $E_c(p, T)$ 의 값이 작아지게 되고 그래프의 에지들의 에너지 최소화시 해당하는 에지가 잘려나갈 가능성이 커지므로, S 터미널과 연결될 가능성이 높아진다.

$E_s(p,q)$ 는 형태 지도에서 가지는 인접 픽셀들의 윤

곽선이 될 확률값 s_p, s_q 의 합으로 이루어진다.

$$E_s(p,q) = s_p + s_q, \quad p, q \in N \quad (12)$$

인접픽셀들의 확률값을 에지의 에너지로 정의함으로써 윤곽선이 될 확률이 큰 픽셀들의 연결은 작은 에너지를 가지게 된다. 즉, 에지가 윤곽선이 되는 것은 픽셀들의 유사성이 작다는 것과 같은 효과를 주게 되고, 세 번째 항의 $E_n(p,q)$ 와 같은 역할을 한다.

$E_n(p,q)$ 는 인접 픽셀과의 컬러, 거리 차이를 계산하여 얻어진다.

$$E_n(p,q) = \frac{1}{D(p,q)} \exp\left(-\frac{(Z_p - Z_q)^2}{2\sigma^2}\right), \quad p, q \in N \quad (13)$$

$D(p,q)$ 는 인접 픽셀 p, q 의 거리이고, 8개의 인접 픽셀 중, 대각선에 위치하는 인접픽셀이 4-인접픽셀 (4-neighbors)보다 큰 값을 가지며 유사도가 적게 된다. Z_p 는 픽셀 p 의 컬러 정보를 나타내며 컬러정보의 차이를 통해 유사도를 판단한다. 그러므로 거리와 컬러에서 유사한 픽셀일수록 높은 에너지를 가지며 같이 분류될 가능성이 커진다.

이렇게 3개의 항으로 형성된 에너지를 최소화하기 위해 우리는 그래프 컷 알고리즘을 적용하였다. 모든 에지, E 에 대해서 생성된 에너지들을 바탕으로 에너지의 합이 최소화되는 C 를 찾음으로써 객체를 배경으로부터 추출한다. 추출 후 객체 주변 배경 영역에서 발생하는 랜덤 잡음에 의한 영향을 줄이기 위해 제거 (opening)연산 후 채움(closing) 연산을 수행하였다.

IV. 실험 결과

제안하는 시스템의 성능을 판단하기 위해 MPEG-4 테스트 영상 및 다양한 시퀀스를 이용하여 실험을 진행하였다. 실험에 사용한 모든 시퀀스의 첫 프레임은 그래프 컷 알고리즘^[13]을 통해 사용자의 상호작용을 통해 추출되었다. 먼저 사용된 MPEG-4 테스트 영상으로는 Akiyo, Claire, Foreman, Grandma, Mother-daughter, Miss America, News, Suzie 시퀀스를 사용하였다. 실



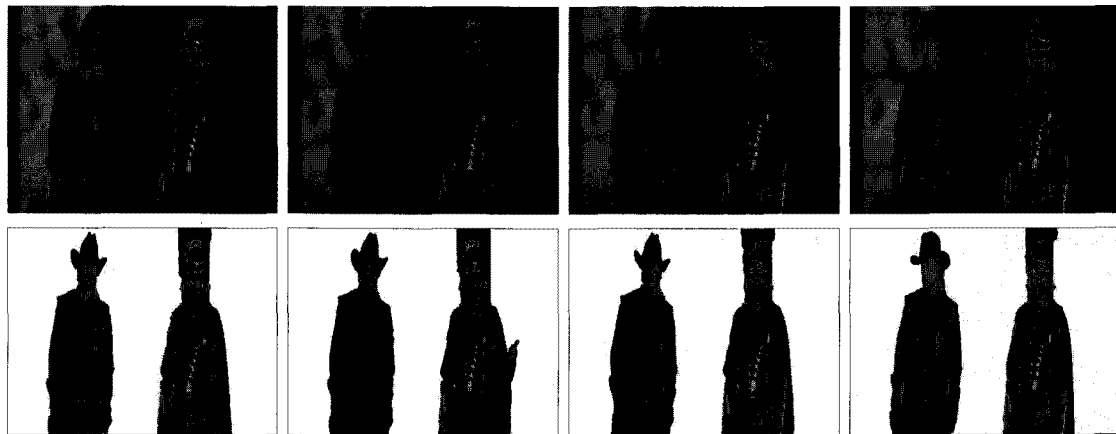
그림 9. MPEG-4 테스트 영상의 객체 추출 결과. 위에서부터 아래로 Akiyo, Claire, Foreman, Grandma, Miss America, Mother-daughter, News, Suzie 시퀀스. (a) 원본 영상(Frame : 20) (b) [21]에서 제안하는 알고리즘을 이용한 (a)의 결과 (c) 본 논문에서 제안하는 알고리즘을 이용한 (a)의 결과 (d) 원본 영상(Frame : 40) (e) [21]에서 제안하는 알고리즘을 이용한 (d)의 결과 (f) 본 논문에서 제안하는 알고리즘을 이용한 (d)의 결과

Fig. 9. Object extraction results for MPEG-4 test sequence, from top to bottom, Akiyo, Claire, Foreman, Grandma, Miss America, Mother-daughter, News, Suzie sequence. (a) original image(Frame : 20) (b) result of (a)(by the method of [21]) (c) result of (a)(by the proposed method) (d) original image(Frame : 40) (e) result of (d)(by the method of [21]) (f) result of (d)(by the proposed method).

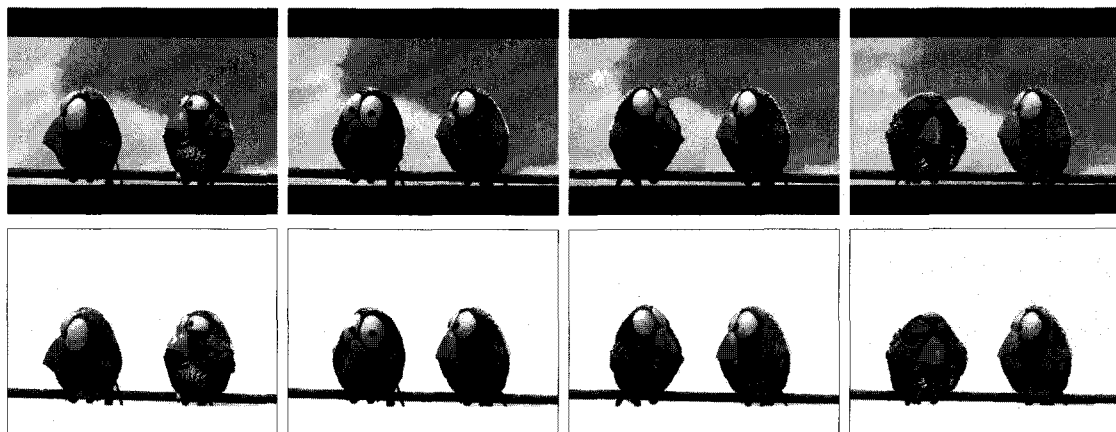
표 1. 본 논문에서 제안하는 알고리즘과 [21]에서 제안하는 알고리즘을 이용한 실험 결과들의 precision 과 recall의 평균값

Table 1. Mean value of the precision and recall(by the proposed method and the method of [21]).

Test Sequence	result of [21]		result of proposed system	
	precision(%)	recall(%)	precision(%)	recall(%)
Akiyo	99.81	95.60	99.38	96.11
Claire	99.58	97.88	98.61	98.19
Foreman	97.75	22.34	99.50	91.32
Grandma	93.97	94.04	99.35	96.34
Miss America	99.67	60.33	98.49	96.57
Mother-daughter	53.17	98.49	96.57	97.19
News	87.77	42.69	99.00	93.36
Suzie	75.90	98.72	99.70	91.36
Average	88.45	76.26	98.83	95.06



(a)



(b)

그림 10. 다양한 영상의 객체 추출 결과. (a) cowboy and knight(Frames : 22, 49, 103, 130)

(b) two birds(Frames : 2, 16, 32, 47)

Fig. 10. Object extraction result for various sequence. (a) cowboy and knight(Frames : 22, 49, 103, 130)

(b) two birds(Frames : 2, 16, 32, 47).

험에 사용한 영상들은 CIF(352×288) 크기의 영상들이다. 각 영상들은 객체와 인접한 배경의 컬러 대조(contrast)가 작아 객체 추출에 어려움이 따르며 객체는 비강체(nonrigid)의 움직임을 가지는 영상들이다.

그림 9에 본 논문에서 제안하는 알고리즘과 [21]에서 제안하는 알고리즘의 추출 결과를 나타내었다. 비교대상 방법이 Claire영상을 제외하고는 유사한 컬러를 가지는 배경에 쉽게 영향을 받으며 추출이 제대로 이루어지지 않는 것에 비해 제안하는 방법은 높은 성능을 보이며 객체를 추출하는 것을 확인할 수 있다. 특히, Mother-daughter, Suzie 시퀀스의 경우 비교대상 방법에서는 에러가 누적되어 결국 영상의 거의 모든 영역이 객체라고 판단되는 것과 달리 제안하는 방법에서는 객체와 배경의 윤곽을 정확히 구분하는 것을 알 수 있다. 또한, Foreman 시퀀스를 통해 제안하는 방법을 이용하여 움직임이 큰 객체에 대해서도 추출이 가능함을 확인할 수 있다.

비교 결과를 수치로 나타내어 정확성을 측정하기 위해 식 (14)과 같이 정의되는 *precision*과 *recall*을 사용하였다.

$$precision = \frac{N(Obj_{EX} \cap Obj_{GT})}{N(Obj_{EX})} \quad (14)$$

$$recall = \frac{N(Obj_{EX} \cap Obj_{GT})}{N(Obj_{GT})}$$

$N(\cdot)$ 는 영역의 픽셀 수를 의미하며, Obj_{EX} 는 제안하는 시스템에서 얻는 객체를 나타내고, Obj_{GT} 는 수동으로 추출(Ground Truth)한 객체를 나타낸다. 그림 9의 MPEG-4 테스트 영상을 사용하였으며 첫 프레임을 기준으로 각각 20 프레임까지의 *precision*과 *recall*을 구하였다. 표 1은 각 영상들의 *precision*과 *recall*의 평균값을 나타낸다. 제안하는 알고리즘에 의해 높은 *precision*과 *recall* 값을 얻는 것을 확인할 수 있다.

또한, 영화와 애니메이션 영상과 같이 객체의 움직임이 빠르고, 배경이 복잡하여도 객체를 정확하게 추출할 수 있다. 그림 10-(a)의 cowboy and knight(704×288) 시퀀스와 그림 10-(b)의 two birds(800×600) 시퀀스의 실험 결과를 통해 고화질의 영상에 대해서도 높은 성능을 가지고 객체를 추출함을 확인할 수 있다.

V. 결 론

본 논문에서는 객체의 윤곽선 정보를 이용한 동영상에서의 정확한 객체 추출 방법을 제안하였다. 움직임 추적으로 얻어진 객체의 윤곽선 정보는 컬러 모델의 형성과 형태 지도 형성, 두 가지 방향으로 이용된다. 기존의 대부분의 연구에서 사용하는 전역적인 컬러 모델의 불필요한 컬러 정보 사용으로 인한 오류를 줄이기 위해 블록 단위의 히스토그램 역투영을 통한 지역적인 컬러 모델을 만들고, 픽셀단위의 누적을 통해 객체와 배경에 대한 사전확률을 설정하였다. 또한 예측된 윤곽선 지도는 거리 변환을 통해 형태지도로 형성되어 인접픽셀간 연결의 확률값을 정의한다. 이때, 로그 함수를 사용하여 각 픽셀이 예측된 윤곽선과의 거리에 따라 다른 가중치를 부여받을 수 있도록 하여 추출의 오류를 줄였다. 이렇게 얻어진 각 픽셀의 컬러와 형태에 관한 확률값은 링크 에너지로 확장되어 그래프를 형성한다. 컬러, 형태, 인접 픽셀과의 유사성을 고려한 새로운 에너지함수를 제안하였고, 매 프레임 에너지의 최소화를 통해 객체를 추출하였다. 다양한 형태의 동영상 시퀀스에 대해서 실험한 결과 제안하는 시스템이 객체의 움직임이 많거나 배경이 복잡한 경우에도 객체를 정확하게 추출해 내는 것을 확인할 수 있었다. 따라서 본 논문이 제안하는 시스템은 3차원 변환을 위한 객체 추출과 같이 높은 성능을 요하는 콘텐츠 기반 동영상 응용분야에 유용하게 활용될 수 있을 것이다.

참 고 문 헌

- [1] G. Hayit, G. Jacob, and M. Arnoldo, "A Probabilistic Framework for Spatio-Temporal Video Representation & Indexing," 7th European Conference on Computer Vision-Part IV, vol. 2353, pp. 461-475, 2002.
- [2] L. Lijie and F. Guoliang, "Combined key-frame extraction and object-based video segmentation," IEEE Transactions on Circuits and Systems for Video Technology, vol. 15, pp. 869-884, 2005.
- [3] O. Javed, Z. Rasheed, K. Shafique, Mubarak Shah, "Tracking Across Multiple Cameras With Disjoint Views," IEEE International Conference on Computer Vision, vol. 2, pp. 952-957, 2003.
- [4] D. Comaniciu, V. Ramesh, P. Meer, "Kernel-Based Object Tracking," IEEE Transactions on Pattern Analysis and Machine

- Intelligence, vol. 25, no. 5, pp. 564-577, 2003.
- [5] J. Kang, I. Cohen, and G. Mediono, "Object reacquisition using geometric invariant appearance model," *IEEE International Conference on Pattern Recognition*, pp. 759-762, 2004.
- [6] A. Yilmaz, X. Li, and M. Shah, "Contour-Based Object Tracking with Occlusion Handling in Video Acquired Using Mobile Cameras," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 11, pp. 1531-1536, 2004.
- [7] Y. Li, J. Sun, and H.Y. Shum, "Video Object Cut and Paste," *ACM Transactions on Graphics*, vol. 24, no. 3, pp. 595-600, 2005.
- [8] B. Li, B. Yuan, and Y. Sun, "Moving Object Segmentation Using Dynamic 3D Graph Cuts and GMM," *IEEE International Conference on Signal Processing*, vol. 2, pp. 16-20, 2006.
- [9] S. Sun, D.R. Haynor, and Y. Kim, "Semiautomatic Video Object Segmentation Using VSnares," *IEEE Transactions on Circuit and System for Video Technology*, vol. 13, no. 1, pp. 75-82, 2003.
- [10] S. Yonggang and W. C. Karl, "A Real-Time Algorithm for the Approximation of Level-Set-Based Curve Evolution," *IEEE Transactions on Image Processing*, vol. 17, pp. 645-656, 2008.
- [11] P. Harper and R. B. Reilly, "Color based video segmentation using level sets," *IEEE International Conference on Image Processing*, vol. 3, pp. 480-483, 2000.
- [12] C. Jung, B. Kim, and C. Kim, "Automatic Segmentation of Salient Objects Using Iterative Reversible Graph Cut," will be appeared to *IEEE International Conference on Multimedia & Expo*, 2010.
- [13] Y. Y. Boykov and M. P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images," *IEEE International Conference on Computer Vision*, vol. 1, pp. 105-112, 2001.
- [14] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive Foreground Extraction using Iterated Graph Cuts," *ACM Transactions on Graphics*, vol. 23, no. 3, pp. 309-314, 2004.
- [15] X. Hou and L. Zhang, "Saliency detection: A Spectral Residual Approach," *IEEE International Conference on Computer Vision and Pattern Recognition*, pp. 1-8, 2007.
- [16] J.F. Talbot, X. Xu, "Implementing GrabCut," Brigham Young University, 2006.
- [17] Z. Liu, J. Cu, L. Shen, Z. Zhang, "Efficient Video Object Segmentation Based on Gaussian Mixture Model and Markov Random Field," *IEEE International Conference on Signal Processing*, pp. 1006-1009, 2008.
- [18] J. Lee, W. Lee, D. Jeong, "Object Tracking Method Using Back-Projection of Multiple Color Histogram Models," *IEEE International Symposium on Circuits and Systems*, vol. 2, pp. 668-671, 2003.
- [19] G. Borgefors, "Distance transformations in digital images," *Computer Vision, Graphics, and Image Processing*, vol. 34, pp. 344-371, 1986.
- [20] F. Y. C. Shih and O. R. Mitchell, "A mathematical morphology approach to Euclidean distance transformation," *IEEE Transactions on Image Processing*, vol. 1, pp. 197-204, 1992.
- [21] Z. Garrett and H. Saito, "Live Video Object Tracking and Segmentation Using Graph Cuts," *IEEE International Conference on Image Processing*, pp. 1576-1579, 2008.

— 저 자 소 개 —



김재광(학생회원)
2010년 연세대학교 전기전자
공학과 학사 졸업.
2010년~한국과학기술원 전기 및
전자공학과 석사과정.

<주관심분야 : 영상이해, 3D 영상처리, 패턴인식>



김창익(정회원)-교신저자
1989년 연세대학교 전기공학과
학사 졸업.
1991년 포항공과대학교 전자전기
공학과 석사 졸업.
1991년~1997년 SKC Ltd. R&D
센터 선임연구원

2000년 워싱턴주립대학교 전기공학과 박사 졸업.
2000년~2005년 Senior member of Technical
Staff, Epson Palo Alto Laboratory,
Epson R&D Inc.
2005년~2009년 한국정보통신대학교 공학부 부교수
2009년~한국과학기술원 전기 및 전자공학과
부교수

<주관심분야 : 3D 영상처리, 패턴인식, 영상이해>



이재호(학생회원)
2006년 부산대학교 전자전기통신
공학부 학사 졸업.
2007년~2009년 한국전자통신연구
원 실감방송시스템연구팀
위촉연구원.
2008년 한국정보통신대학교
공학부 석사 졸업.

2008년~한국과학기술원 전기 및 전자공학과
박사과정

<주관심분야 : 3D 영상처리, 의료영상처리, 패턴인식>