

동영상 부호화를 위한 반자동 영상분할 기법

김민호, 김대희, 호요성
 광주과학기술원 정보통신공학과
 광주광역시 광산구 쌍암동 572 번지

Semi-Automatic Segmentation for Video Coding

Minho Kim, Daehee Kim and Yosung Ho
 Kwangju Institute of Science and Technology (K-JIST)
 572 Ssang-am Dong Kwang-san Gu, Kwang-ju, Korea
 E-mail: minokim@geguri.kjist.ac.kr

요약

본 논문에서는 이중 레이블 구조를 이용하는 반자동 영상분할 기법을 제안한다. 반자동 영상분할 기법은 크게 화면내 영상분할과 화면간 영상분할의 두 단계로 이루어진다. 화면내 영상분할에서는 사용자에 의한 인지 정보와 영상내의 밝기 정보를 이용하는데, 이러한 두 가지 정보를 효과적으로 융합하여 영상을 분할하기 위해 제안된 이중 레이블 구조를 이용한다. 이중 레이블 구조에서는 우선, 사용자에 의한 인지 정보를 이용하여 전체 레이블 마스크를 얻고, 다음으로 영상내의 밝기 정보를 이용하여 국부 레이블 마스크를 만든다. 이때 형태학적 분할 기법 (Morphological Segmentation Technique)을 이용한다. 이렇게 얻어진 두 가지 레이블 마스크를 융합하여 화면내 영상분할을 완성한다. 화면간 영상분할에서는 화면내 영상분할에서 정의된 비디오 객체에 대한 움직임을 추적하여 영상을 분할한다.

1. 서론

MPEG-4 동영상 부호화 표준은 영상을 객체 중심으로 부호화한다. 객체기반 부호화에서는 동영상 비디오의 각각의 화면을 VOP (Video Object Plane)의 형태로 표현한다. 이러한 VOP를 얻기 위해서는 영상분할이 이루어져야 하는데, 이를 위해 몇몇 자동 영상분할 기술들이 개발되었다. 가장 대표적인 방법들로서 변화검출 (Change Detection) 기법, 움직임 예측 기반 영상분할 기법, 형태학적 영상분할 기법, 시공간 정보를 이용하는 영상분할 기법 등이 있다.

변화 검출 기법은 연속하는 두 영상의 차 영상을 이용한 통계적 가설 검증 기법이다. 이 기법은 Global Thresholding 방법을 이용하고 있기 때문에 본질적으로 국부 정보를 이용하지 못한다. 이를 보완하기 위해 MRF (Markov Random Field)를 이용하기도 한다 [1]. 하지만 이 방법은 벗겨진 배경 (Uncovered Background)과 덮이진 배경 (Background to be Covered)을 구별하지 못

한다. 움직임 예측 기반 영상분할 기법은 화소 밝기의 시공간 미분값을 이용한 Mapping Parameter의 예측을 통하여 이러한 문제를 해결할 수 있다 [2].

위와 같이 연속되는 두 영상의 차이값만을 이용하는 기법들은 비디오 객체의 기하학적 특성을 효율적으로 다루지 못하지만, 형태학적 영상분할 기법은 이것이 가능하다. 형태학적 영상분할 기법은 형태학적 도구를 이용하여 영상을 단순화시킨 다음, Watershed 알고리즘으로 영상을 분할한다 [3, 4]. 그러나 형태학적 영상분할은 영상의 공간적 정보만을 이용하고 있을 뿐만 아니라, 분할된 영역 자체로는 하나의 비디오 객체를 나타내지 못하는 객체 대응의 문제점이 있다.

지금까지의 영상분할 기법들은 시간축 정보와 공간적 정보중에서 어느 하나만을 이용하므로 비디오 객체의 경계를 정확히 찾아내지 못하고 심지어는 비디오 객체와 배경이 병합되는 결과를 얻을 수도 있다. 이러한 이유로 최근 시간축 정보와 공간적 정보를 함께 이용하여 좀 더 나은 분할 결과를 얻고자 하는 시도들이 있었다 [6]. 하지만 두 가지 정보 모두를 이용하는 방법 조차도 이러한 문제를 완전히 해결하지 못했다. 왜냐하면 비디오 객체의 정의가 본질적으로 주관적인데다가, 자동 영상분할 기법은 단지 영상의 밝기 정보만을 이용했기 때문이다.

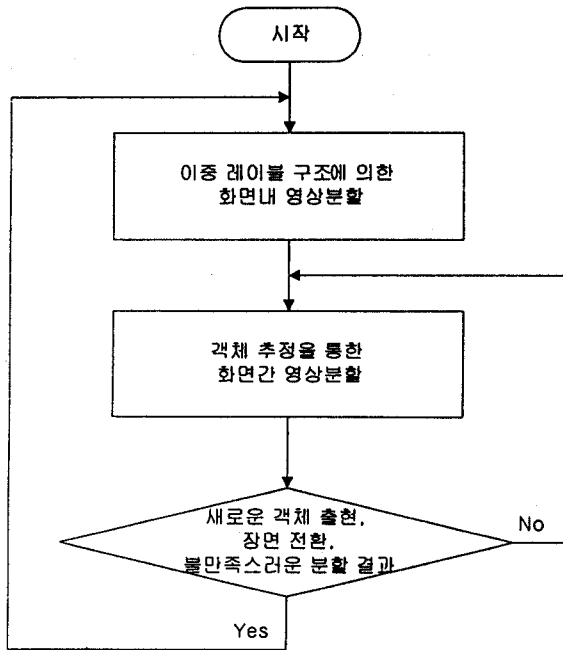
본 논문에서는 이러한 문제점들을 해결하기 위해 이중 레이블 구조를 이용하는 반자동 영상분할 기법을 제안한다. 반자동 영상분할 기법의 핵심은 자동 영상분할 기법에 제공된 인지 정보를 효과적으로 이용하는 부분과 분할된 객체를 추적하는 부분이라 할 수 있다. 인지 정보의 이용은 본 논문에서 제안한 이중 레이블 구조를 통해 이루어질 수 있고, 분할된 비디오 객체의 추적은 비디오 객체의 움직임 예측에 의한 영상분할 기법 [7]을 이용할 수 있다.

반자동 영상분할 알고리즘은 크게 화면내 영상분할과 화면간 영상분할로 구성된다. 화면내 영상분할은 Global Labeling, Local Labeling, 전체 레이블 마스크와 국부 레이블 마스크의 융합으로 이루어진다. 화면간 영상분할에서는 분할된 비디오 객체의 움직임 추정을 통해 영상분할을 완성한다.

본 논문에서 제안한 이중 레이블 구조를 이용하는 반자동 영상분할 알고리즘을 이용하면 기존의 자동 영상분할에서 인지 정보의 부재로 인해 발생하는 객체 대응의 문제점을 해결할 수 있었고, HVS (Human Visual System)의 입장에서 볼 때에도 자동 영상분할 기법보다 훨씬 나은 결과를 얻을 수 있다.

2. 반자동 영상분할 알고리즘

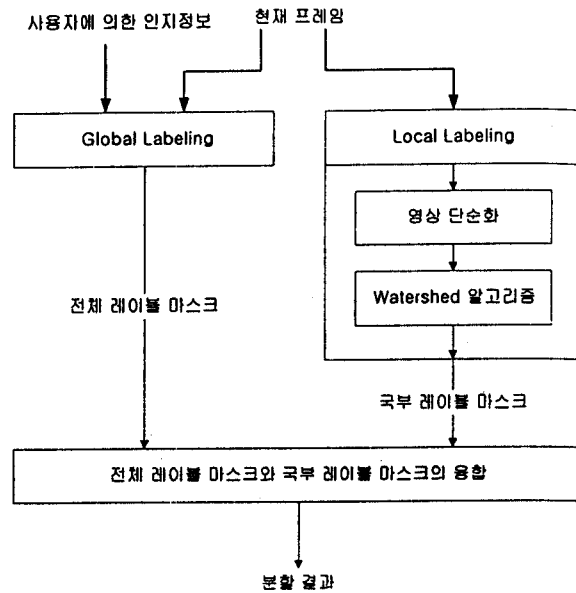
<그림 1>에서 보는 바와 같이 반자동 영상분할은 화면내 영상분할과 화면간 영상분할의 두 부분으로 이루어져 있다. 첫번째 프레임에 대해, 화면내 영상분할을 통해 관심있는 비디오 객체를 배경으로부터 분리한다. 이때 사용자의 개입을 통해 인지 정보가 화면내 영상분할 알고리즘에 제공된다. 다음으로 후속되는 프레임들에 대해서 세가지 사건 - 1) 관심있는 새로운 비디오 객체 출현, 2) 장면 전환 (Scene Change), 3) 불만족스러운 분할 결과 - 이 발생하기 전까지는 화면간 영상분할을 반복한다. 만약 위에서 언급한 사건들이 발생한다면 화면내 영상분할을 다시 시작한다.



<그림 1> 반자동 영상분할 알고리즘

2.1 화면내 영상분할

화면내 영상분할은 사용자의 개입을 통해 주어진 인지 정보와 영상에서 주어진 밝기 정보를 이용하여 <그림 2>에 주어진 알고리즘을 이용하여 영상을 분할한다. 화면내 영상분할은 Global Labeling, Local Labeling, 전체 레이블 마스크와 국부 레이블 마스크의 융합으로 구성된다.



<그림 2> 화면내 영상분할

우선 Global Labeling에서는 사용자에 의한 인지 정보를 GUI (Graphic User Interface)와 같은 도구를 이용하여 알고리즘에 추가한다. 좀 더 구체적으로 설명하면, GUI 도구를 이용하여 우리가 찾고자 하는 비디오 객체의 실제 경계 주위에 불확정 영역을 설정하여 전체 레이블 마스크 (Globally Labeled Mask)를 얻는다. 이렇게 얻어진 전체 레이블 마스크는 배경, 비디오 객체, 불확정 영역으로 이루어진다.

Local Labeling에서는 영상의 밝기 정보를 이용하여 국부 레이블 마스크 (Locally Labeled Mask)를 얻는다. 이때 형태학적 영상분할 기법을 이용하며, 이것은 일반적으로 영상 단순화 (Simplification) 과정과 Watershed 알고리즘으로 이루어져 있다.

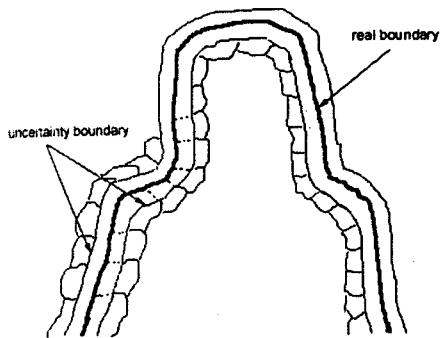
영상 단순화 과정은 복구를 통한 형태학적 Opening 과 Closing (Morphological Opening/Closing by Reconstruction)을 이용한다. 복구를 통한 형태학적 Opening 은 구조 요소 (Structuring Element)보다 작은 밝은 영역을 없애주고, 복구를 통한 형태학적 Closing 은 구조 요소보다 작은 어두운 영역을 없애준다. 단순한 Opening 이나 Closing 이 경계 정보를 손실을 초래하는 것과는 달리, 복구를 통한 형태학적 Opening 과 Closing 은 경계 정보를 보존시켜 준다.

Watershed 알고리즘은 형태학적 경사(Morphological Gradient Operator)를 이용하여 얻은 영상을 균일한 영역들로 나누어 준다. 경사화된 영상에서 침수 실험 (Immersion Simulation)을 이용한다. 이 방법은 각각의 지역 최소값을 기반으로 영역을 분할한다 [5].

형태학적 영상분할 기법은 영상을 비디오 객체나 배경과 같은 의미 있는 객체와는 관계없이 단지 밝기 정보가 비슷한 여러 개의 작은 영역들로 나눈다. 따라서 Local Labeling에서 나오는 결과를 국부 레이블 마스크라 하며, 마지막으로 전체 레이블 마스크와 국부 레이블 마스크를 융합하여 최종 마스크를 얻는다.

2.2 두 레이블 마스크의 융합

전체 레이블 마스크와 국부 레이블 마스크의 융합 과정을 <그림 3>에 설명하였다.



<그림 3> 전체와 국부 레이블 마스크의 융합

전체 레이블 마스크는 비디오 객체 영역, 배경, 불확정 영역으로 구성된다. 이 마스크는 불확정 영역을 포함하고 있지만, 분할된 영역과 비디오 객체가 일대일 대응성을 가지고 있다. 이에 반해 국부 레이블 마스크는 각각의 분할된 영역이 비디오 객체와의 일대일 대응성을 갖지 못하지만, 분할된 영역들이 영상의 지형학적 특성을 포함하고 있기 때문에 비디오 객체의 경계 영역의 정보를 보존한다. 전체 레이블 마스크와 국부 레이블 마스크와의 융합은 각각의 마스크가 가지고 있는 장점인 전체 레이블 마스크의 분할된 영역과 비디오 객체의 일대일 대응성과 국부 레이블 마스크의 경계 정보를 이용한다.

<그림 3>은 전체 레이블 마스크 위에 국부 레이블 마스크를 겹쳐 놓은 것이다. 그림에서 굵은 선(Real Boundary)은 실제 비디오 객체의 경계 영역이고, 불확정 경계(Uncertainty Boundary)라고 표시된 선 사이가 불확정 영역이다. 국부 레이블 마스크의 영역들 중에서 비디오 객체 또는 배경과 불확정 영역 사이에 걸쳐 있는 국부 영역이 존재하게 된다.

<그림 3>에서 점선으로 된 부분과 비디오 객체 내의 실선으로 된 영역은 실제 하나의 국부 영역으로써 두 영역을 걸쳐 있는 부분이다. 이 그림에서 보는 바와 같이, 불확정 영역은 이러한 영역들로 구성되어 있다고 볼 수 있는데, 국부 레이블 마스크의 지형학적 정보를 이용하여 전체 레이블 마스크의 불확정 영역을 제거해 나간다.

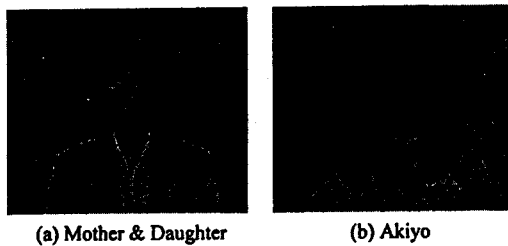
역으로 말하면, 이러한 과정을 통해 비디오 객체 영역과 배경 영역을 각각 성장시켜 실제의 경계 영역을 찾는다 볼 수 있다. 이 과정에서 국부 영역이 두개의 영역에 걸쳐 있지 못하고 불확정 영역에 완전히 포함되는 경우가 발생할 수 있는데, 이 경우 주위 국부 영역들과의 유사성, 구체적으로 영역의 밝기의 평균을 이용하여 주위 영역으로 병합시켜서 이런 영역을 제거할 수 있다. 위와 같은 융합 과정을 통해 화면간 영상분할을 완성하게 된다.

2.3 화면간 영상분할

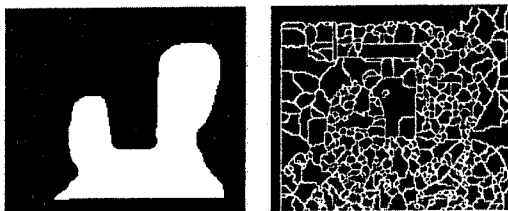
화면간 영상분할은 화면내 영상분할에서 얻어진 비디오 객체의 추적을 통해 영상을 분할한다. 우선 비디오 객체의 움직임 예측을 통해 이전 프레임에서 얻어진 비디오 객체를 현재 프레임에 움직임을 보상하여 투영한다. 그런 다음 현재 프레임의 밝기 정보를 이용하여 움직임이 보상된 투영에서 발생할지 모르는 오류를 보정한다. 즉, 현재 프레임의 경계 정보를 이용하여 비디오 객체의 정확한 경계를 찾는 것이다 [7].

3. 실험 및 결과

본 논문에서 제안된 이중 레이블 구조를 이용하는 반자동 영상분할 기법의 성능을 알아보기 위해 176x144 크기의 QCIF 형태의 Mother & Daughter 영상과 Akiyo 영상을 사용하여 실험하였다. <그림 4>는 실험에서 사용된 Mother & Daughter와 Akiyo의 원영상을 보여주고 있다.



<그림 4> 원영상



(a) 전체 레이블 마스크 (b) 국부 레이블 마스크



(c) 최종 마스크

<그림 5> 이중 레이블 구조

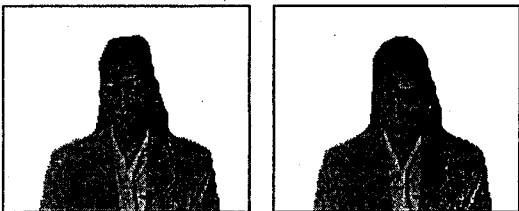
<그림 5>에는 반자동 영상분할 기법의 첫번째 과정인 화면내 영상분할 알고리즘의 처리 과정을 보였다. 사용자의 개입에 의해 제공된 인지정보를 이용하여 <그림 5> (a)와 같은 전체 레이블 마스크를 얻게 된다. <그림 5> (a)에서 알 수 있듯이, 전체 레이블 마스크는 비디오 객체 영역, 배경, 불확정 영역으로 구

성되어 있다. <그림 5> (b)는 Local Labeling 과정을 통해서 얻은 국부 레이블 마스크이다. 국부 레이블 마스크는 각각의 분할된 영역이 비디오 객체와의 일대일 대응성을 갖지 못함을 볼 수 있다. 하지만 각각의 분할된 영역에는 영상의 지형학적 특성이 고려되어 있기 때문에, 비디오 객체의 경계 정보를 보존하고 있음을 알 수 있다.

전체 레이블 마스크와 국부 레이블 마스크와의 융합은 각각의 마스크가 가지고 있는 장점인 전체 레이블 마스크의 분할된 영역과 비디오 객체의 일대일 대응성과 국부 레이블 마스크의 경계 정보를 이용한다. 이렇게 얻어진 최종 마스크가 <그림 5> (c)에 주어진다. 이 마스크를 이용하여 얻어진 영상의 전경이 <그림 6> (a)에 주어진다. <그림 6> (a)에서 보듯이, 제안된 이중 레이블 구조가 인지 정보와 밝기 정보를 효율적으로 이용하고 있음을 보여주고 있다. <그림 6> (b)의 전경은 화면내 영상분할에서 얻어진 비디오 객체의 추적, 즉, 화면간 영상분할 알고리즘을 적용하여 얻어진 결과의 99 번째 프레임이다.



(a) 화면내 영상분할 (b) 화면간 영상분할 (99th Frame)
<그림 6> Mother & Daughter 영상을 이용한 결과



(a) 화면내 영상분할 (b) 화면간 영상분할 (47th Frame)
<그림 7> Akiyo 영상을 이용한 결과

<그림 7> (a)는 Akiyo 영상에 화면내 영상분할을 적용하여 얻은 결과이고, <그림 7> (b)는 화면내 영상분할에서 얻어진 결과를 이용하여 후속되는 프레임들에 대해 화면간 영상분할 알고리즘을 적용하여 얻어진 결과의 47 번째 프레임이다. <그림 7> (a)의 결과를 보면, 사람의 귀의 일부분과 머리카락의 일부분이 잘려져 나가는 문제점을 보여주고 있다. 이것은 배경과 영상 사이의 화소값의 유사함으로 인해, Local Labeling 과정에서 잘못된 국부 레이블 마스크를 생성했기 때문이다. 다시 말해 경계의 불확실성으로 인해 국부 레이블 마스크에서 배경과 비디오 객체 영역이 병합되어 버리는 문제를 초래했다. 이러한 문제점은 좀 더 정확한 Local Labeling 을 통해 해결될 수 있다.

4. 결론

본 논문에서는 이중 레이블 구조를 이용하는 반자동 영상분할 기법을 제안하였다. 반자동 영상분할 알고리즘은 화면내 영상분할과 화면간 영상분할로 이루어진다. 화면내 영상분할에서는 사용자의 개입을 통해 주어지는 인지 정보와 영상에서 주어지는 밝기 정보를 효율적으로 융합하기 위해 제안된 이중 레이블 구조를 이용하였다. 전체 레이블 마스크는 인지 정보를 이용하여 얻었고, 국부 레이블 마스크는 밝기 정보를 이용하는 형태학적 영상분할 알고리즘을 이용하여 얻었다. 이 두 가지 마스크를 융합하여 분할된 비디오 객체를 얻을 수 있었다. 이 과정에서 이중 레이블 구조의 성능을 확인할 수 있었다. 그리고 화면간 영상분할을 통해 후속되는 프레임들에 대해 영상을 분할할 수 있었다. 이러한 반자동 영상분할 알고리즘은 MPEG-4 와 같은 객체 기반의 영상처리 분야에서 폭넓게 이용될 수 있을 것이다.

참고문헌

- [1] T. Aach and A. Kaup, "Statistical Model-based Change Detection in Moving Video", *Signal Processing*, 31, pp. 165-180, 1993
- [2] M. Hötter and R. Thoma, "Image Segmentation Based on Object Oriented Mapping Parameter Estimation", *Signal Processing*, 15, pp. 315-334, 1988
- [3] F. Meyer and S. Beucher, "Morphological Segmentation", *J. Visual Communication and Image Representation*, vol. 1, no. 1, pp. 21-46, September 1990
- [4] P. Salembier, "Morphological Multiscale Segmentation for Image Coding", *Signal Processing*, 38, pp. 359-386, 1994
- [5] L. Vincent and P. Soille, "Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, no. 6, June 1991
- [6] J. G. Choi, S. W. Lee, and S. D. Kim, "Spatio-Temporal Video Segmentation Using a Joint Similarity Measure", *IEEE Trans. Circuits and Systems for Video Technology*, vol. 7, no. 2, April 1997
- [7] J.G. Choi and et al., "A User-Assisted Segmentation Method for Video Object Plane Generation", *ITC-CSCC'98*, pp. 7-10, July 1998.