

# 시공간 정보를 이용한 동영상 객체 분할 기법

오 혁, \*최환수, \*\*정동석  
인하대학교 공과대학 전자공학과

## Video Object Segmentation Method Using Spatio-Temporal Information

Hyuk Oh, \*Hwan-Soo Choi, \*\*Dong-Seok Jeong

Department of Electronic Engineering, Inha University

E-mail : g1991198@inhavision.inha.ac.kr, \*chs2000@dreamwiz.com, \*\*dsjeong@dragon.inha.ac.kr

### 요 약

영상으로부터 의미있는 객체를 영역화하기 위하여, 움직임에 의한 시간적 정보를 이용하거나, 형태학적(Morphological) 기법과 같이 공간적 정보를 이용하는 방법이 있다. 그러나, 단지 시간적 정보나 공간적 정보만을 이용하는 방법은 그 한계를 가지고 있으며, 본 논문에서는 시공간 정보를 이용하여 분할하는 방법을 채택하였다. 시간적 분할에서는, 두 프레임에서 움직임 정보를 찾아내었던 기존 방법을 보완하여 연속되는 세 프레임을 사용하도록 하였다. 이렇게 하면 움직임이 미세한 영상에 대해서도 객체를 분리해 낼 가능성을 높일 수 있게 된다. 공간적 분할에서는, Watershed 알고리즘을 이용하는 형태학적 분할(Morphological Segmentation)[1][2]을 하게 되는데, 전처리 과정의 단일적도 경사(Monoscale Gradient) 대신 다중적도 경사(Multiscale Gradient)[3][4]를 사용하여 미세한 경사는 누그러뜨리고 에지 부분의 경사만을 강조하게 하였다. 또한 개선된 Watershed 알고리즘을 제안하여 기존의 Watershed 알고리즘의 과분할 문제를 보완하였다.

### I. 서 론

MPEG-4 동영상 부호화 표준[1]의 가장 큰 특징은 객체 기반 부호화 방식을 채택하고 있다는 점이다. 영상을 객체기반으로 부호화하기 위해서는 우선 각각의 화면을 VOP(Video Object Plane)의 형태로 나타내야 한다. 이를 위해서는, 배경으로부터 움직이는 객체인 VOP를 분리해내어야 하는데, 이러한 기술을 영상분할(Image Segmentation)이라고 한다. 영상분할을 할 때에는 시간적 정보를 이용하거나, 형태학적(Morphological) 기법과 같은 공간적 정보를 이용하는데, 시간적 정보만을 이용하게 되는 경우 움직임 객체의 존재를 확인

할 수 있지만 확인된 객체의 정확한 경계를 찾을 수 없으며, 공간적 정보만을 이용하는 알고리즘에서는 그 경계를 보존할 수 있는 반면 움직임 객체의 존재를 파악할 수 없다. 그래서 MPEG-4 표준의 부기(annex)에서는 시공간 정보를 이용하는 방식을 소개하고 있으며, 본 논문에서도 시간적, 공간적으로 각각 분할을 한 후에 두 결과를 조합하는 방법을 채택하였다.

시간적 영상분할에서는, 연속하는 프레임들에서 움직임 정보를 찾아내어 변화 검출 마스크(Change Detection Mask, CDM)를 얻어낸다.[5][6] 기존에는 연속하는 두 프레임만으로 CDM을 얻어내었는데, 움직임이 매우 작은 영상에서는 움직임을 놓치는 경우가 많으므로, 한 프레임 더 이전의 것도 같이 사용할 것을 제안하였다. 즉, 본 논문에서는 이전 CDM와 현재 CDM를 OR연산함으로써 최종 CDM을 얻도록 하고 있다.

공간적 영상분할에서는, 작은 경사값에도 과분할(Oversegmentation)을 만들었던 기존의 Watershed 알고리즘[4]을 보완하는 개선된 Watershed 알고리즘을 제안하였다. 이를 이용하면, 별다른 후처리 없이도 과분할을 상당히 줄일 수 있는 장점이 있다. 또한 Watershed의 전처리 과정으로 다중적도 형태학적 경사(Multiscale Morphological Gradient)를 사용하므로써 Watershed 알고리즘이 과분할을 일으키는 것을 최대한 억제하였다.

### II. 시간적 영상분할

시간적 영상분할에서는 통계학적 가설검증 기법(Statistical Hypothesis Testing)[6]을 사용하였다. 즉, 연속되는 두 프레임에서 차영상을 구하고 관찰 윈도우(Observation Window) 안팎의 분산(Variance)을 조사하여 가설을 검증해 가는 방법이다. 관찰 윈도우 밖의 분산을  $\sigma_1^2$ , 관찰 윈도우 안의 분산을  $\sigma_2^2$ 라 하고, 초

기가정(Null Hypothesis,  $H_0$ )으로 두 분산이 같다고 세운다. 만약, 초기 가정을 만족하지 못한다면, 다른 가정 ( $H_1$ )을 검사하게 된다.

$$\begin{aligned} H_0: \sigma_1^2 &= \sigma_2^2 \\ H_1: \sigma_1^2 &< \sigma_2^2 \end{aligned} \quad (1)$$

표본수가  $n_1, n_2$ 인 분산의 추정치를  $S_1^2, S_2^2$ 라 한다면,  $S_1^2/\sigma_1^2$ 와  $S_2^2/\sigma_2^2$ 는 각각  $n_1-1$ 과  $n_2-1$ 의 자유도를 갖는  $\chi^2$ 분포( $\chi^2$  distribution)가 된다. 검사통계(Test Statistic)  $V$ 는

$$V = \frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2} \quad (2)$$

$n_1-1$ 과  $n_2-1$ 의 자유도를 갖는  $F$  분포( $F$  distribution)가 될 것이다. 초기 가정에 의하면  $\sigma_1^2 = \sigma_2^2$  이므로  $V = S_2^2/S_1^2$  이 되고,  $V$ 를 구함에 있어서 실제 분산은 필요없게 된다. 초기 가정하에서는  $S_1^2$ 와  $S_2^2$ 가 비슷한 값으로 되어  $V = S_2^2/S_1^2 \approx 1$  이 된다. 만약, 이 비가 1보다 상당히 크게 된다면, 관찰 윈도우가 움직임 영역을 지나고 있다고 가정할 수 있다. 이렇게 검사 통계  $V$ 에 대하여 (1)식에서 세운 가정을 검사(Testing)하므로써 변화 검출 마스크를 얻게 되는데, CDM은 움직임 영역인 전경(Foreground)과 움직임이 없는 배경(Background)으로 구성된 이진 영상(Binary Image)이다.

연속하는 두 프레임에 대해 CDM을 구해보면, 움직임이 작은 영상들에 대해서는 그림 1의 (b), (c)처럼 실제 전경으로 판단되어야 할 부분들을 놓치는 경우가 많다. 그래서 이를 보완하기 위해 이전 CDM을 함께 참조한다. 현재 프레임  $F_k$ 와 이전 프레임  $F_{k-1}$ 에서 얻어진 현재 CDM을  $CDM_k$ 라 하고,  $F_{k-1}$ 과 그 이전 프레임  $F_{k-2}$ 에서 얻어진 이전 CDM을  $CDM_{k-1}$ 이라 한다면, 변화 검출 마스크는 두 CDM의 OR연산으로 얻게 된다. 즉,

$$CDM = (CDM_k \text{ or } CDM_{k-1}). \quad (3)$$

이렇게 이전의 두 프레임까지 참조하면 움직임이 미세한 경우에도 그림 1의 (d)처럼 움직임을 놓치는 부분을 줄일 수 있게 되어 더 좋은 분할결과를 얻게 된다.

### III. 공간적 영상분할

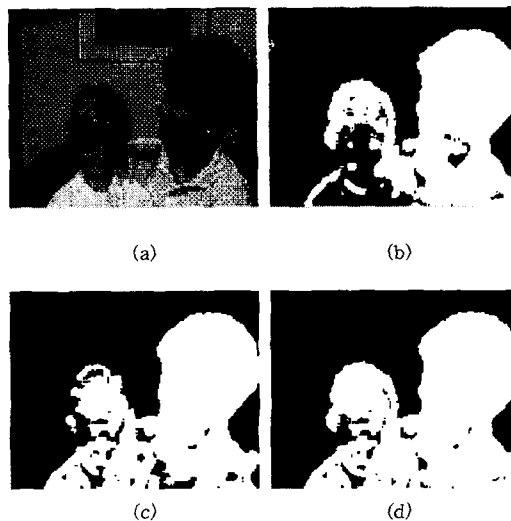


그림 1. 시간적 영상분할 결과. (a) 원영상 Mother & Daughter 33번째 프레임, (b)  $CDM_{33}$ , (c)  $CDM_{32}$ , (d) 최종 시간적 영상분할 결과 ( $CDM = (CDM_{33} \text{ or } CDM_{32})$ ).

공간적 영상분할에서는 형태학적 필터(Morphological Filter)와 Watershed 알고리즘을 이용하는 형태학적 분할을 사용하였다[1][2]. 이 방법은 크게 세 단계를 거치게 되는데, 영상의 단순화와 형태학적 경사(Morphological Gradient), Watershed 알고리즘을 수행하게 된다.

#### A. 영상의 단순화

영상의 단순화에는 재구성에 의한 열림과 닫힘(Opening by Reconstruction & Closing by Reconstruction)을 많이 사용한다. 재구성에 의한 열림과 닫힘은 영상을 평탄화 시켜주는 동시에 영역들의 경계를 유지시켜 주는 장점이 있으며, 각각

$$\begin{aligned} \gamma^{(rec)} &(\gamma_n(F_k), F_k) \\ \phi^{(rec)} &(\phi_n(F_k), F_k) \end{aligned} \quad (4)$$

로 정의된다. 여기서  $\gamma_n$ 은 형태학적 열림이고  $\phi_n$ 은 형태학적 닫힘이다. 형태소(Structuring Element)의 크기는  $(2n+1) \times (2n+1)$ 이다.

#### B. 다중척도 형태학적 경사

기존에는 일반적으로

$$G_1(f) = \delta_1(f) - \epsilon_1(f) \quad (5)$$

로 정의되는 단일척도 형태학적 경사를 사용하였으나,

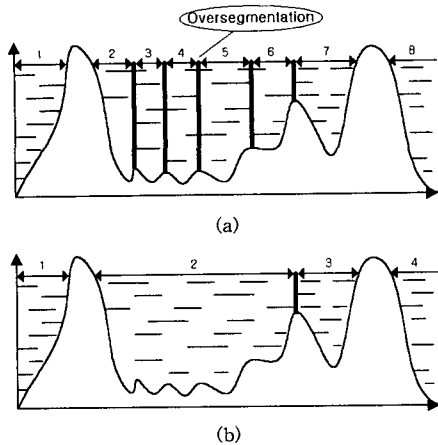


그림 2. Watershed 알고리즘의 비교. (a) 일반적인 Watershed 알고리즘, (b) 제안한 Modified Watershed 알고리즘.

미세한 경사들을 많이 남겨두어 과분할을 일으키는 원인이 되었다. (5)식에서  $\delta_n$ ,  $\epsilon_n$ 는 각각 형태학적 팽창(Dilation), 부식(Erosion)이다.

본 논문에서는 다중척도 형태학적 경사[5][6]를 사용하므로써, 단일척도 경사에서 남겨놓은 미세한 경사들을 누그러뜨리고 에지부분의 경사들만을 강조하여 과분할을 최대한 줄이도록 하였다.

$$MG_n(f) = \frac{1}{n} \sum_{i=1}^n [ \epsilon_{i-1} ( WTH_i ( G_i(f) ) ) ], \quad 1 \leq i \leq n. \quad (6)$$

여기에서 WTH은 White Top-Hat 연산이고, 실험에서는  $n=2$ 인 2중척도(2 scale)를 사용하였다.

### C. 개선된 Watershed 알고리즘

일반적인 Watershed 알고리즘에서는, 영상의 낮은 값에서부터 단계적으로 침수(Immersion)를 해 나갈 때, 다른 집수영역(Catchment Basin)과 만나면 무조건 가상댐(Watershed Line)을 세운다[2]. 그러나 이렇게 되면, 그림 2의 (a)와 같이 낮은 경사들에 대해서도 서로 다른 영역으로 레이블(Label)이 되므로 과분할을 일으키게 된다. 그러나, 그림 2의 (b)처럼 낮은 경사들에서는 댐을 세우지 않는다면, 과분할을 상당히 줄일 수 있을 것이다. 본 논문에서는 이러한 개념에 착안하여, 침수를 해 나갈 때 다른 집수영역과 만나면 먼저 집수영역의 깊이를 조사하고 문턱값 이하이면 같은 영역으로 포함시켜 침수를 계속해 나가는 방식의 개선된 Watershed 알고리즘을 제안하였다.

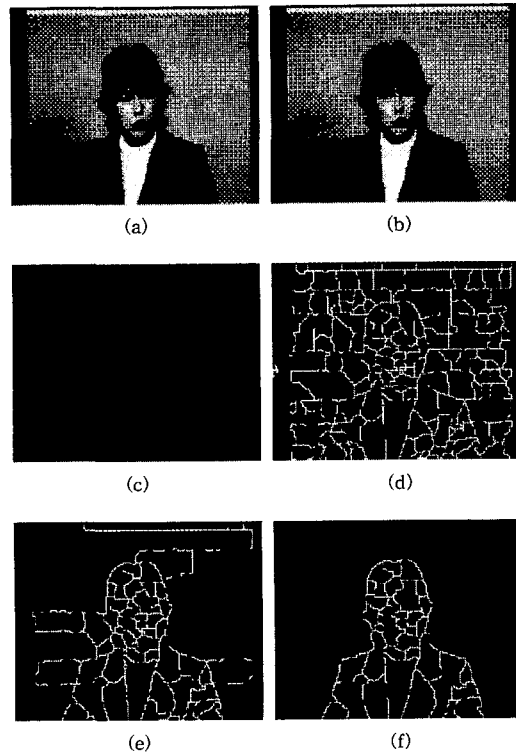


그림 3. 공간적 영상분할 결과. (a) 원영상 Claire 48번째 프레임, (b) 영상의 단순화, (c) 다중척도 형태학적 경사, (d) 일반적인 Watershed 알고리즘, (e) Modified Watershed 알고리즘(문턱치 1), (f) Modified Watershed 알고리즘(문턱치 2).

제안된 Watershed 알고리즘을 이용하면 영역 병합(Region Merging)과 같은 별다른 후처리 없이도 과분할을 줄일 수 있어서 상당히 좋은 분할결과를 얻을 수 있게 된다. 그림 3는 실제 실험결과를 보여준다. (e) 그림처럼 깊이가 1이상일 때만 댐을 세우도록 하여도 일반적인 Watershed 알고리즘을 사용했을 때보다 과분할을 상당히 줄일 수 있음을 볼 수 있다.

## IV. 시공간적 정보의 중첩

시간적 분할결과와 공간적 분할결과를 결합(Fusion)하므로써 최종 분할결과를 얻어낸다. 즉, Watershed 영상의 각 영역과 CDM을 중첩시켰을때, CDM의 전경영역과 일정 비율 이상 겹치면 객체부분으로 최종 판단하고, 그렇지 않으면 배경으로 판단하게 된다.

## V. 실험결과

본 논문에서는 배경이 단조롭거나 복잡한 경우, 움직

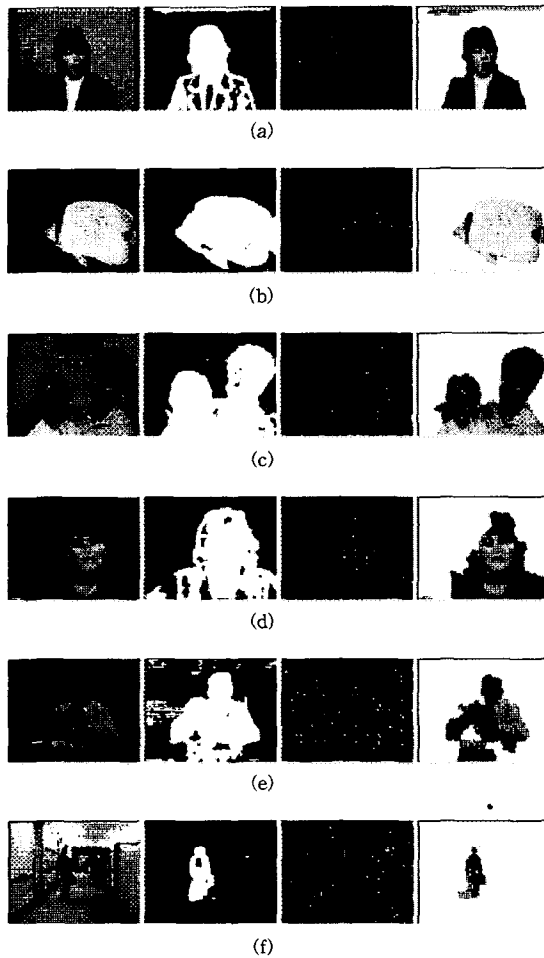


그림 4. 영상분할 결과. (a) Claire 141번째 프레임, (b) Bream2 162번째 프레임, (c) Mother & Daughter 252번째 프레임, (d) Miss America 102번째 프레임, (e) Salesman 17번째 프레임, (f) Hall Objects 52번째 프레임.

이는 객체가 작은 경우등 다양한 형태의 QCIF (176x144) 영상들에 대하여 실험을 해 보았다. 그림 4에서 보는 바와 같이 (a)와 (b), (c)의 경우처럼 배경이 단조로운 경우는 결과가 매우 좋았다. (d)의 경우 머리부분과 팔부분이 조금 유실되었는데, 이는 배경부분과 밝기값이 비슷하여 공간적 분할의 전처리 과정에서 경계를 잃어버렸기 때문이다. 시공간적 객체분할에서 가장 어려운 점은, 배경이 매우 복잡한 영상의 경우인데, 이는 공간적 분할에서 배경까지도 복잡하게 분할되어 시간적 분할결과와 잘못 결합되기 쉽기 때문이다. 그럼에도 불구하고, (e)는 비교적 움직인 객체를 양호하게 분할하였는데, 이는 본 논문에서 제안한 것처럼 공간적 분할에서 과분할을 최대한 억제하고, 시간적 분할

에서 움직임을 가급적 놓치지 않았기 때문이다. (f)처럼 영상에서 객체가 차지하는 영역이 작을 경우 공간적 분할의 성과가 매우 중요한데, 이는 배경이 단조롭지 않은 이상 배경과 객체부분이 구분없이 뒤섞일 가능성이 있기 때문이다. 본 논문에서 제안한 방법으로 공간적 분할의 결과, (f)에서 보는 바와 같이 움직이는 작은 객체를 대체로 잘 분리해 낼 수 있었다.

## VI. 결 론

본 논문에서는 시공간 정보를 이용하여 영상에서 움직이는 객체를 분리해 내었다. 시간적 분할을 할 때에는 이전 CDM을 참조함으로써 움직임을 놓치는 부분을 최소화하도록 보완했으며, 공간적 분할에서는 다중적도 경사로 전처리 과정을 개선하고, 개선된 Watershed 알고리즘을 제안함으로써 과분할을 최대한 억제하였다. 그러나, 움직임이 거의 없는 영상의 경우, 시간적 분할을 할 때에 여전히 움직임을 놓치는 부분이 있었고, 갑자기 움직임이 빨라질 경우 CDM이 실제 객체영역보다 훨씬 크게 전경으로 판단하게 되는 문제점이 있었다. 이러한 문제점들은 이전 분할영역을 참조하여 영역추적 (Region Tracking)을 함으로써 보완할 수 있을 것이다. 또한 컬러(Color)정보까지 함께 이용한다면 훨씬 더 좋은 결과를 얻을 수 있을 것 같다.

## 참고문헌

1. L. Vincent and P. Soille, "Watersheds in digital spaces: An efficient algorithm based on immersion simulations," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 13, no. 6, pp. 583-598, June 1991.
2. P. Salembier and M. Pardas, "Hierarchical morphological segmentation for image sequence coding," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 639-651, Sep. 1994.
3. D. Wang, "A multiscale gradient algorithm for image segmentation using watersheds," *Pattern Recognition*, vol. 3, no. 3, pp. 2043-2052, 1997.
4. P. Soille, *Morphological Image Analysis*, Springer, 1999.
5. ISO/IEC 14496-2, International Standard, first edition, 1st Dec. 1999.
6. M. Kim, J. G. Choi, D. Kim, H. Lee, M. H. Lee, C. Ahn, and Y.-S. Ho, "A VOP generation tool: Automatic segmentation of moving objects in image sequences based on spatio-temporal information," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1216-1226, Dec. 1999.