

# 생물학 기반의 확률적 모션 추정 모델을 이용한 영상 분할

## Image Segmentation Using Bio-Inspired Probabilistic Motion Estimation Model

이세형, 박영빈, 서일홍

한양대학교 전자컴퓨터통신공학부

Se Hyung Lee, Young Bin Park, Il Hong Suh

Department of electronics and computer engineering, Hanyang University

### Abstract

In this paper we present a model for image segmentation taking into consideration motion and color. We divide the image into segments in first stage to resolve the problem of two or more object having similar color values being in one segment. Additionally a textured object divided into segments initially is merged with one segment by using motion information. For reliability and accuracy, we developed a bio-inspired motion estimation model based on the Bayesian formulation. In this model, motion is estimated by probabilistic processing of the pixel and region layers. Several experiments were performed to evaluate the performance of our model. The results of these experiments showed that the proposed methodology is effective for image segmentation.

### I. 서 론

영상 분할은 컴퓨터 비전 분야에 있어 중요한 연구 분야 중 하나이다. 만일 영상 내에서 객체 단위 혹은 다른 의미 있는 개체로 분할을 할 수 있다면, 객체 인식이나 객체 추적뿐 아니라, 영상 압축, 3D복원 등 다양한 분야에 도움을 줄 수 있다. 따라서 이러한 영상의 분할을 위한 다양한 연구들이 진행되어 왔고, 그 결과 여러 방법들이 제안되었다.

특히 Graph-cut을 이용한 energy-minimization 방법<sup>[3][4]</sup>이나 Mean-shift<sup>[1]</sup>를 통한 특성 분석을 이용한 방법들이 좋은 성능을 보여주었으며 이 외에도<sup>[2]</sup>와 같은 다양한 방법들이 제안되었다. 이러한 색상 특성 기반의 분할 기법들은 복잡한 무늬를 가지는 객체의 경우 하나의 객체가 여러 개의 영역들로 분할되는 경향이 크다. 또한

비슷한 색상 특성을 가지는 여러 객체들이 인접해 있을 경우 하나의 영역으로 합쳐지게 된다.

이 논문에서는 색상 정보와 시점이 조금 다른 두 장의 영상을 이용하여 모션 차이를 이용하여 객체 분할에 활용하고자 한다. Global 매칭<sup>[6]</sup>, Local 매칭<sup>[7]</sup>, 생물학 기반의 모델<sup>[5]</sup> 등 다양한 방법들 중 기본적으로 생물학 기반 모델로부터 영향을 받아 feed-forward 및 feed-back 모델을 Bayesian formulation을 이용하여 설계하였다. 이를 이용하여 모션을 추정할 때 발생하기 쉬운 aperture-problem을 해결하였다.

### II. 영상 분할 과정

#### 1. 초기 과분할

본 논문에서는 우리는 픽셀의 색상이 주변 픽셀들과의 평균 및 분산이 정해진 임계값 이하 일 경우 균질하다고 판단하였다.

$$\sigma(p_x, p_y, n) = \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} \frac{I(p_x+i, p_y+j)^2}{n^2} - \left( \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} \frac{I(p_x+i, p_y+j)}{n^2} \right)^2 \quad (1)$$

$$\mu(p_x, p_y, n) = \sum_{i=-n/2}^{n/2} \sum_{j=-n/2}^{n/2} \frac{I(p_x+i, p_y+j)}{n^2} \quad (2)$$

여기서  $p_x, p_y$ 는 x, y 축에 대한 픽셀의 좌표이고,  $n$ 은 윈도우의 크기를 뜻한다. (1), (2)를 기반으로 픽셀이 균질한지를 아래와 같이 판단한다.

$$(p_x, p_y) \text{ is homogeneous} \quad (3)$$

$$\text{if } \sigma(p_x, p_y, n_b) - \sigma(p_x, p_y, n_s) \leq Th_\sigma$$

$$\mu(p_x, p_y, n_b) - \mu(p_x, p_y, n_s) \leq Th_\mu$$

여기서  $Th_\sigma$ 와  $Th_\mu$ 는 분산과 평균에 대한 임계값을 뜻하고  $n_b, n_s$ 는 큰 크기의 윈도우와 작은 크기의 윈도우를 의미한다. 만약 어떤 픽셀이 (3)을 만족한다면 그 픽셀은 균질하다 판단한다.

우리는 균질이라 검출된 픽셀들을 바탕으로 초기 균질 영역을 만들었다. 이 후, 임계값  $Th_\sigma$ 와  $Th_\mu$ 을 기하급수적으로 증가시켜 새로운 균질 영역들을 찾고, 새롭게 검출된 균질 픽셀들을 이전에 검출된 균질 영역들에 추가시켰다. 이렇게 검출된 균질 영역들을 seed로 이용하여 Region growing 방법으로 초기 이미지 과분할을 완성하였다.

## 2. 재귀적 모션 추정

검출된 과분할 상태의 각 영역에 대하여 모션을 구하는데 있어, 픽셀 레벨에서의 단순한 픽셀 매칭 방법으로는 aperture-problem이 발생하기 쉽다. 이러한 문제를 해결하기 위해서 픽셀 레벨과 영역 레벨로 2개의 층으로 나누어 각 레벨에서의 모션 정보를 이용하여 베이지안 수식을 통해 확률적으로 모델링하고 최대 확률을 구하는 문제로 만들었다.

우리는 픽셀 레벨에서는 특정 크기의 윈도우를 설정하고 SAD(Sum of Absolute Differences)를 이용하여 픽셀간의 유사도를 측정하였다. 이 후, 유사도에 역수를 취하고 정규화를 해주어 픽셀이 매칭 될 확률을 구하였다.

$$p(e) = \frac{k}{SAD(e_x, e_y) + \varepsilon} \quad (4)$$

여기서  $\varepsilon=0.001$  이고,  $k$ 는 정규화 상수를 뜻하며,  $e$ 는 픽셀의 모션,  $e_x, e_y$ 은 각각 영상에서의 x축 및 y축에 대한 위치를 의미한다.

영역 레벨에서 각 영역에 속하는 픽셀들을 동시에 특정 모션으로 움직여 다음 영상과의 유사성을 비교였다.

$$p(s) = \frac{k}{\sum_{i=1}^N |I(e_x^i, e_y^i) - I(e_x^i + dx, e_y^i + dy) + \varepsilon|} \quad (5)$$

여기서  $s$ 는 영역의 모션을,  $N$ 은 영역에 속하는 픽셀의 개수를 뜻한다. 이 후, 우리는 영상의 과분할 된 모든 영역에 대하여 가장 큰 확률을 가지는 모션을 찾기 위하여 확률 모델을 수식화 하였다.

$$\hat{S}_t = \operatorname{argmax}_{s_t^1, s_t^2, \dots, s_t^n} P(s_t^1, s_t^2, \dots, s_t^n) \quad (6)$$

여기서  $\hat{S}_t$ 는 가장 높은 확률을 가지는 영역들의 모션 조합이다. 여기서 각 영역들이 서로 독립이라 가정하여 (6)을 다시 수식화 하게 되면

$$\hat{S}_t = \operatorname{argmax}_{s_t^1, s_t^2, \dots, s_t^n} P(s_t^1)P(s_t^2), \dots, P(s_t^n) \quad (7)$$

(7)은 각 영역들의 모션을 최대화하는 것을 의미한다. 그리고 각 영역들은 픽셀들로 이루어져 있으므로 영역의 모션이 주어졌을 때 각 픽셀의 모션을 최대화하도록 수식화하면

$$\operatorname{argmax}_{e_t} P(e_t | s_t) = \operatorname{argmax}_{e_t} P(s_t | e_t) P(e_t) \quad (8)$$

여기서  $s_t$ 는 영역의 모션이고,  $e_t$ 는 영역에 속해있는 픽셀의 모션이다. 픽셀 모션은 픽셀 레벨에서 최대 사후 확률(MAP)을 이용하여 결정한다. 우리는 초기 우도  $P(s_0 | e_0)$ 이 균등 분포를 가진다 가정하였다. 또한 사전 확률  $P(e_t)$ 은 (4)에 따라 계산된다.

우리는 각 픽셀의 이웃 픽셀들을 고려하여 확률을 계산하였다.

$$p(e_t) = \frac{P(e_t | neighbor(e_t))}{P(neighbor(e_t) | e_t) p(e_t)} = \frac{P(neighbor(e_t))}{P(neighbor(e_t))} \quad (9)$$

여기서  $neighbor(e_t)$ 는 픽셀  $e_t$ 의 이웃하는 상, 하, 좌, 우 픽셀의 모션을 의미한다. 우리는 이를 각각  $e_t^u, e_t^d, e_t^l, e_t^r$ 로 표시하였다.

$$P(e_t | \hat{s}_t) = k P(\hat{s}_t | e_t) P(e_t) \quad (10)$$

$$= k P(\hat{s}_t | e_t) \exp \{ \log P(e_t | \hat{e}_t^u) + \log P(e_t | \hat{e}_t^d) + \log P(e_t | \hat{e}_t^l) + \log P(e_t | \hat{e}_t^r) - 3 \log P(e_t) \}$$

여기서  $P(e_t | \hat{e}_t^u), P(e_t | \hat{e}_t^d), P(e_t | \hat{e}_t^l), P(e_t | \hat{e}_t^r)$ 는 이웃

한 픽셀의 모션을 바탕으로 가우시안 분포로 추정한다.

경계선 근처의 픽셀의 경우 유사도가 높은 후보들이 균등 지역의 픽셀에 비해 적게 나타나므로, 균등 지역에 비해 영역 레벨에서 픽셀의 모션을 취합했을 때, 균질 영역에 존재하는 픽셀들에 비하여 더 많은 영향을 미치게 된다. Feed-forward 모델은 아래와 같다.

$$P(E_i|s_i) \approx k \sum_{e_i \in s_i} P(e_i|\hat{s}_i) \quad (11)$$

여기서  $E_i$ 는 영역에 속하는 픽셀들의 집합의 모션이며,  $k$ 는 정규화 상수이다. 이렇게 구해진 확률은 영역 레벨에서의 우도로 이용된다.

$$\begin{aligned} \hat{s}_{t+1} &= \operatorname{argmax}_{s_t} P(s_t|E_t) \\ &= \operatorname{argmax}_{s_t} P(E_t|s_t)P(s_t) \end{aligned} \quad (12)$$

여기서 우도  $P(E_t|s_t)$ 은 (11)을 통해 얻어진 확률을 의미하며,  $P(s_t)$ 는 (5)에서 구해진 영역 단위의 모션을 뜻한다. 영역 모션은  $\hat{s}_{t+1}$ 로 갱신되고 다시 아래와 같은 feed-back 모델을 통해 초기에 균등 분포로 가정하였던  $P(\hat{s}_{t+1}|e_{t+1})$ 을 갱신한다.

$$P(\hat{s}_{t+1}|e_{t+1}) \approx \frac{P(e_{t+1}|\hat{s}_{t+1})P(\hat{s}_{t+1})}{P(e_{t+1})} \quad (13)$$

여기서  $P(e_{t+1}|\hat{s}_{t+1})$ 는 (12)에서 가장 높은 확률을 가지는 모션  $\hat{s}_{t+1}$ 를 바탕으로 가우시안 분포를 의미한다. 이러한 확률은 다시 (10)의 우도를 갱신한다. 반복 과정은 지정된 특정 횟수만큼 반복하거나 영역의 모션이 수렴하게 되면 종료된다. 이렇게 구해진 모션을 바탕으로 각 영역에서의 모션 크기를 구한다.

$$d = \sqrt{m_x^2 + m_y^2} \quad (14)$$

여기서  $m_x$ ,  $m_y$ 는 x축, y축에 대한 모션의 크기를 의미하며,  $d$ 는 이에 대한 유클리디안 거리를 의미한다.

### 3. 최종 분할

이전 과정에서 구해진 모션 정보와 색상정보를 이용하여 각 영역들을 다시 합친다.

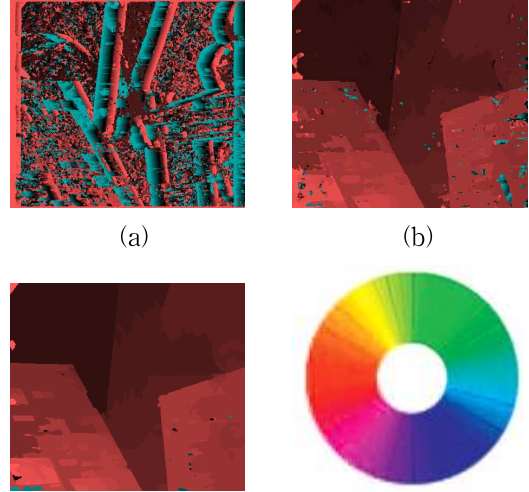


그림 1. 재귀적 모션 추정 결과, (a)초기모션, (b)1회 반복 후 모션, (c)4회 반복 후 모션, (d)모션의 방향

$$\begin{aligned} & \text{segment}_i \text{ and } \text{segment}_n \text{ are merged} \\ & \text{if } \begin{cases} |d_i - d_n| \leq Th_d \text{ or} \\ |c_i - c_n| \leq Th_c \end{cases} \end{aligned} \quad (15)$$

여기서  $d_i$ ,  $c_i$ 는  $i$ 번째의 영역의 색상 및 모션 정보를 의미하며  $d_n$ ,  $c_n$ 는  $i$ 번째의 영역에 인접해 있는 이웃영역의 색상 및 모션정보를 의미,  $Th_d$ ,  $Th_c$ 는 모션 및 색상에 대한 임계값을 뜻한다.

초기 과분할 상태에서 다수의 영역들로 분할이 되며, 동일한 객체에 속하는 영역들의 경우 모션이 서로 비슷하거나 또는 연속적이다. 따라서 우리는 임계값  $Th_d$ 을 1로 정하였다. 또한 색상에 대한 임계값  $Th_c$ 은 실험을 통해 비슷한 색상을 가지는 객체와 배경이 합쳐지지 않도록 충분히 작은 값으로 설정하였다.

## III. 실험결과

제안된 방법의 성능을 평가하기 위하여 다양한 환경에서 실험을 하였다. 첫 번째 그림에서는 영상의 모션을 추정하기 위해 사용한 반복적인 과정을 보여준다. 그림1에서 색상은 모션의 방향을 의미하고, 채도는 모션의 크기를 의미한다. 초기 단계에서 모션을 추정하게 되면, aperture-problem으로 인해 모션이 잘못 추정된 결과가 그림1(a)와 같이 나타난다. 이 후 재귀적 추정 과정을 통해 모션을 다시 추정하게 되면 그림1(b)와 같이 모션이 다시 추정되는 것을 알 수 있다. 이러한 과정을 반복함으로써 영역에 속한 픽셀들이 가장 최적이라 판단되는 모

선으로 수렴하게 되고 그림1(c)와 같은 결과가 나타났다.

그림2는 색상 정보만을 이용한 논문<sup>[2]</sup>의 결과 및 제안하는 방법의 결과를 보여준다. 논문<sup>[2]</sup>의 경우 그림2의 두 번째 줄의 결과와 같이 배경과 합쳐지거나 과분할 되는 것을 볼 수 있다. 제안하는 방법과 같이 초기 과분할 상태의 영역에서 모션을 구하여 영상을 분할한 결과 그림2의 세 번째 줄과 같이 나타났다. 그림2(a)와 같이 복잡한 무늬를 가진 객체의 경우 각 영역의 모션에 따라 다시 하나의 객체로 합쳐졌으며, 비슷한 색상을 가지는 그림2(b)의 경우 초기 과분할을 통해 다른 영역으로 나누어진 뒤, 각 영역의 모션이 다름으로 인하여 하나의 영역으로 합쳐지지 않았으며 객체와 배경으로 영상이 분할되었다.

#### IV. 결론

우리는 이 논문에서 영상을 분할하는 방법에 대하여 설명했다. 일반적인 색상만으로 영상을 분할 할 경우 발생하는 문제점으로 인하여 우리는 두 시점이 조금 다른 영상을 이용하여 두 영상 사이의 모션의 차이를 추정하여 영상을 분할하였다. 초기 과분할 상태로 만들기 위하여 균질 영역을 검출하였고, 각 영역별로 모션을 추정하기 위해 픽셀 레벨 및 영역 레벨로 나누고 레벨간의 feed-forward 및 feed-back을 통해 aperture-problem을 해결하였다. 그 결과 실험에서와 같이 복잡한 무늬를 가지는 객체 또는 비슷한 색상을 가지는 객체의 경우에도 올바르게 분할이 되는 것을 알 수 있다.

#### 감사의 글

이 논문(저서)는 2010년 교육과학기술부의 재원으로 한국연구재단의 지원을 받아 <실감교류 인체감응솔루션> 글로벌프린티어연구개발사업으로 수행된 연구임(한국연구재단-MIAXA003-2010-0029744)

교신저자: 서일홍, ihsuh@hanyang.ac.kr

#### 참고 문헌

[1] D. Comaniciu and P. Meer, "Mean Shift: A robust approach toward feature space analysis,"



(a)



(b)



(c)



(d)

그림 2. 영상 분할 비교. 왼쪽부터 입력 영상, 논문<sup>[2]</sup>결과, 본 논문의 결과

IEEE Transactions on pattern analysis and machine intelligence, pp, 603-619, 2002.

- [2] P. Felzenszwalb and D. Huttenlocher, "Efficient graph-based image segmentation," International journal of computer vision, vol. 59, no. 2, pp. 167-181, 2004.
- [3] Y. Boykov and M. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in nd images," International Conference on Computer Vision, vol.1, pp. 105-112, 2001.
- [4] N. Xu, R. Bansal, and N. Ahuja, "Object segmentation using graph cuts based active contours," IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, 2003.
- [5] P. Bayerl and H. Neumann, "A fast biologically inspired algorithm for recurrent motion estimation," IEEE Transactions on pattern analysis and machine intelligence, pp. 246-260, 2007.
- [6] M. Werlberger, T. Pock and H. Bischof, "Motion estimation with non-local total variation regularization," IEEE Conference on Computer Vision and Pattern Recognition, pp. 2464-2471, 2010.
- [7] K. J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," IEEE Transactions on pattern analysis and machine intelligence, vol. 28, no. 4, pp. 650-656, 2006.