

Markov Random Fields 를 이용한 움직이는 객체 추출 및 추적

장세일*, 황선규, 김희율
한양대학교 전자통신전파공학과

Moving Object Segmentation and Tracking Using Markov Random Fields

Se-il Jang*, Sun-kyoo Hwang, Whoi-yul Kim

Department of Electrical and Computer Engineering, Hanyang University

E-mail : seiljang@vision.hanyang.ac.kr

Abstract

기존의 객체 추출 및 추적 기법은 외형 변화가 없는 객체를 대상으로 하거나 배경이 고정된 영상만을 고려하였다. 본 논문에서는 영역의 색상과 움직임 정보, 그리고 인접한 영역의 상관 관계를 고려한 Markov Random Field (MRF) 모델을 제안한다. MRF 모델은 영상의 시간적·공간적 상관성을 기반으로 최적의 레이블 셋을 계산함으로써 보다 정확하게 객체를 추출 및 추적할 수 있다. 또한, 블록 기반 움직임 추출 알고리즘인 Diamond Search (DS)를 분할된 영역에 적용하여 빠르게 영역의 움직임과 전역 움직임을 추정하였다. 실험 결과 제안한 방법이 객체의 외형 변화와 카메라 움직임이 있는 동영상에서 빠른 속도로 정확하게 객체를 추출 및 추적하는 것을 확인하였다.

I. 서론

최근 인터넷과 영상 장비의 보편화로 인하여 영상 데이터의 사용이 급격하게 증가하고 있다. 영상 데이터를 효율적으로 관리하고 편리하게 사용하기 위한 방법으로 객체 중심의 내용기반 검색 기법이 활발히 연구되고 있다[1]. 객체 중심의 내용기반 검색 기법은 정지 영상 및 동영상에서 의미 있는 객체의 특징을 추출하여 이를 기반으로 영상 데이터 베이스를 검색하는 기법이다. 이러한 객체 중심의 내용기반 검색 기법에서 객체 추출은 필수적인 요소이다. 또한, 보안 시스템, 가상현실, 영상 데이터 분석, 그리고 핫스팟 등의 여러 응용분야에서 움직이는 객체 또는 의미 있는 정지 객체를 추출 및 추적하는 기술이 요구되고 있다.

일반적으로 동영상 내의 객체는 모양, 크기, 회전 등의 많은 외형 변화를 가지므로 객체와 배경을 분리하

고 연속된 프레임에서 객체를 추적하는데 많은 문제점을 야기시킨다. 최근에는 이러한 객체의 외형 변화에 강인하게 객체를 추출 및 추적하는 기법이 많이 연구되고 있다[2-4]. 그러나 기존의 연구된 방법들은 객체의 변화나 카메라의 움직임이 있는 경우 부정확한 객체 추출 결과를 나타낸다[2]. 또한, 정확한 객체를 추출하는 방법들은 많은 계산 시간을 요구한다[3][4].

본 논문에서는 영역 정보의 시간적·공간적 상관성에 기반한 Markov Random Field (MRF) 모델을 정의하여 객체의 외형 변화와 카메라 움직임에 강인하게 객체를 추출 및 추적하는 방법을 제안한다. 또한, Diamond Search (DS)알고리즘을 이용하여 영역의 움직임과 전역 움직임을 추정하고, Highest Confidence First (HCF)알고리즘을 사용하여 MRF 모델의 최적 레이블 셋을 계산함으로써 객체 추출 및 추적 방법에서 요구되는 계산 시간을 최소화 하였다.

II. 영상 분할

영상 분할(Image Segmentation)이란 영상의 밝기, 색상, 질감 등의 특징을 고려하여 유사한 속성을 갖는 영역 단위로 영상을 표현하는 것을 의미한다[5]. 일반적으로 영상 내에 존재하는 특정 객체를 인식하거나 분석하기 위해서는 그 객체의 영역만을 정확하게 추출할 필요가 있다.

본 논문에서는 자연 영상을 영역 단위로 표현하기 위한 방법으로 rainfalloing 방식의 watershed 알고리즘을 사용하였다[5]. Watershed 알고리즘은 영상의 그래디언트 (gradient) 정보를 이용하여 영역을 구분하는 방법이다. 일반적으로 watershed 알고리즘은 영상의 잡음으로 인하여 수 많은 영역으로 나누어지는 과분할 현상을

발생시킨다. 본 논문에서는 이러한 과분할 현상을 제거하기 위해 anisotropic diffusion 필터[6]를 사용하여 영상을 부드럽게 함으로써 영상의 잡음을 제거하고, Region Adjacent Graph (RAG)를 이용하여 과분할된 영역을 병합함으로써 객체와 배경을 적절한 개수의 영역으로 표현하였다[7].

III. 영역의 움직임 이용한 객체 추출

일반적으로 동영상에서의 움직임 추출은 움직임 고려 대상에 따라 픽셀 기반, 블록 기반, 그리고 영역 기반 움직임 추출 방법으로 분류할 수 있다. 본 논문에서는 블록 기반 움직임 추출 방법에서 많은 연구가 이루어진 Diamond Search (DS) 알고리즘[8]을 영역 기반 움직임 추출 방법에 응용하여 영역의 움직임을 추출하고, 추출된 움직임을 이용하여 전역 움직임을 추정하였다. 또한, 추정된 움직임을 바탕으로 객체 영역과 배경 영역을 판단하고 초기 설정된 영역을 추적하여 움직이는 객체 및 사용자가 원하는 객체를 추출 및 추적하였다.

DS 알고리즘은 그림 1과 같은 두 개의 탐색 패턴의 각 탐색점에 대한 Mean Absolute Difference (MAD)를 계산하여 영상의 움직임 벡터 (v_x, v_y) 을 추출하는 방법이다. 자세한 방법은 아래와 같이 4 단계로 구성된다.

- 1 단계: $(0, 0)$ 을 중심점으로 정의한다.
- 2 단계: 그림 1(a)와 같은 large 탐색 패턴의 9 개의 탐색점에 대한 MAD를 계산한다.
- 3 단계: 중심점을 최소의 MAD를 갖는 점으로 이동시킨다. 만약 중심점에서 최소의 MAD를 갖는다면 4 단계로 이동하고, 그렇지 않으면 2 단계를 반복 수행한다.
- 4 단계: 그림 1(b)와 같은 small 탐색 패턴의 5 개의 탐색점에 대한 MAD를 계산한다. 여기서 최소의 MAD를 갖는 점의 위치를 구하고자 하는 움직임 벡터 (v_x, v_y) 로 판단한다.

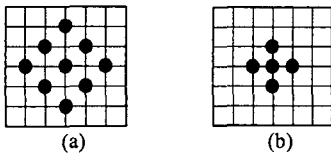


그림 1. Diamond Search 패턴
(a) Large 탐색 패턴 (b) Small 탐색 패턴

본 논문에서는 그림 2와 같이 DS 알고리즘을 분할된 영역에 적용하여 영역의 움직임을 추출하였다. 즉,

분할된 영역을 탐색 패턴의 탐색점을 따라 이동시켜 MAD를 계산하고, 영역의 움직임 벡터를 추출하였다. MAD는 식 (1)과 같이 YUV 색상 공간에서 각 성분에 대한 절대 거리의 합으로 계산하였다. 여기서 N_R 은 영역 내의 픽셀 수를 의미한다.

$$MAD(v_x, v_y) = \frac{1}{N_R} \sum_{r=1}^{N_R} |I'(r_x, r_y) - I'^{-1}(r_x + v_x, r_y + v_y)| \quad (1)$$

$$|I' - I'^{-1}| = |I'_Y - I'^{-1}_Y| + |I'_U - I'^{-1}_U| + |I'_V - I'^{-1}_V|$$

또한, 객체 및 배경의 움직임이 시간적·공간적 상관성을 갖는다는 특성[8]을 기반으로 주변 영역과 이전 프레임의 움직임 정보를 이용하여 움직임을 보정함으로써 움직임 추정의 정확성을 향상시켰다.

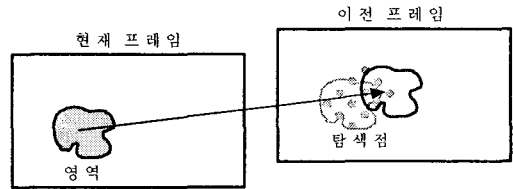


그림 2. 영역 기반 움직임 추출

영역 단위로 추출된 움직임을 기반으로 영상에 가장 많이 분포하는 움직임을 전역 움직임으로 정의하고, 전역 움직임과 영역 움직임 벡터가 다른 영역을 현재 프레임에서의 움직임 영역으로 판단하였다. 또한, 식 (2)와 같이 각 영역에 움직임 벡터를 적용하여 이전 프레임에서 객체로 판단된 영역의 넓이를 계산하고, 식 (3)과 같이 문턱값($Th=0.5$)에 따라 객체 영역과 배경 영역을 판단하여 객체를 추출 및 추적하였다. 식 (2)에서 $Area(r)$ 은 영역의 넓이를 의미하고, $Select(r)$ 은 객체로 판단된 영역의 넓이를 나타낸다.

$$S(r) = \frac{Select'^{-1}(\vec{r})}{Area'(r)} \quad (2)$$

$$R_{Select}(r) = \begin{cases} foreground, & S(r) > Th \\ Background, & S(r) < Th \end{cases} \quad (3)$$

IV. 제안한 MRF 모델

일반적인 동영상은 객체의 외형이 변화하거나 카메라의 움직임에 의해 배경이 변화한다. 이러한 동영상에서의 변화는 부정확한 객체 추출 결과를 나타내는 원인이 된다. 본 논문에서는 객체의 외형 변화와 카메라 움직임에 강인하게 객체를 추출 및 추적하기 위한 방법으로 영상의 시간적·공간적 상관성에 기반한

MRF 모델을 제안한다.

MRF 는 미래의 상태는 현재의 상태에만 의존하고 과거의 상태와는 독립적이라는 Markov 이론을 2 차원 또는 3 차원으로 확장한 개념이다. 즉, 2 차원 평면에서 한 점 (x, y) 와 이웃하는 점을 (x_N, y_N) 라고 정의할 때, 현재의 상태 $S(x, y)$ 는 주변의 상태 $S(x_N, y_N)$ 에만 의존한다는 것을 의미한다[9].

본 논문에서는 이전 단계에서 계산된 영역 단위 색상 정보와 움직임 정보의 시간적·공간적 상관성을 기반으로 동영상에서 객체의 외형 변화와 카메라 움직임에 강인하게 객체를 추출 및 추적하기 위해 식 (4)과 같은 에너지 함수 모델을 제안한다. 식 (4)에서 m 은 영역의 개수, d 는 이전 단계에서 부여한 영역의 레이블 셋, 그리고 f 는 영역에 부여 가능한 레이블 셋을 나타낸다.

$$U(f|d) = \sum_{i=1}^m \left[V^M(f_i|d_i) + \sum_{j=N_i, ec} V^S(f_i|f_j) + V^T(f_i|f_{i-1}) \right] \quad (4)$$

식 (4)에서 포텐셜 함수 $V^M(f_i|d_i)$ 는 i 번째 영역의 이전 단계의 선택된 정보와 영역의 움직임을 이용하여 식 (5)와 계산된다. 여기서 위 첨자 M 은 영역의 움직임 정보를 이용한 포텐셜 함수임을 표시한다. 즉, 전역 움직임과 동일한 움직임을 갖는 영역은 배경 영역으로 다른 움직임을 갖는 영역은 객체 영역으로 판단할 수 있도록 에너지를 조정한다.

$$V^M(f_i|d_i) = \begin{cases} -\alpha D_i^{GMV} N_i, & (f_i = \text{foreground} \& \mathbf{MV}_i \neq \mathbf{GMV}) \text{or} \\ & (f_i = \text{background} \& \mathbf{MV}_i = \mathbf{GMV}) \\ \alpha D_i^{GMV} N_i, & (f_i = \text{foreground} \& \mathbf{MV}_i = \mathbf{GMV}) \text{or} \\ & (f_i = \text{background} \& \mathbf{MV}_i \neq \mathbf{GMV}) \end{cases} \quad (5)$$

식 (5)에서 N_i 는 영역의 픽셀 개수이고, \mathbf{MV}_i 는 i 번째 영역의 움직임 벡터, \mathbf{GMV} 는 전역 움직임 벡터를 나타낸다. D_i^{GMV} 는 식 (6)와 같이 영역의 움직임과 전역 움직임의 절대 거리의 합으로 계산된다.

$$D_i^{GMV} = |\mathbf{GMV}^x - \mathbf{MV}_i^x| + |\mathbf{GMV}^y - \mathbf{MV}_i^y| \quad (6)$$

식 (4)의 두 번째 항, 포텐셜 함수 $V^S(f_i|f_j)$ 는 영역의 공간적 상관성을 부여하고, 주변 영역의 현재 영역에 대한 영향을 나타낸다. N_i 는 i 번째 영역과 이웃한 영역의 위치를 나타낸다. $V^S(f_i|f_j)$ 는 식 (7)와 같이 색상, 움직임, 경계선의 길이에 의해 결정된다.

$$V^S(f_i|f_j) = \begin{cases} -\beta D_{ij}^{Color} D_{ij}^{MV} B_{ij}, & f_i = f_j \\ \beta D_{ij}^{Color} D_{ij}^{MV} B_{ij}, & f_i \neq f_j \end{cases} \quad (7)$$

식 (7)에서 D_{ij}^{Color} 는 두 영역의 평균 색상 차로 식 (8)과 같이 정의 되고 D_{ij}^{MV} 는 두 영역의 움직임 벡터의 차로 식 (9)과 같이 정의 된다. B_{ij} 는 두 영역 간의 경계선의 길이를 나타낸다. 식 (9)에서 가중치 w_v , w_l , w_r 는 각각 1:2:2 의 비율로 정의하여 밝기 정보보다 색상 정보에 더 많은 가중치를 부여함으로써 조명 변화로 인한 잡음의 영향을 최소화 하였다.

$$D_{ij}^{Color} = w_v |I_{i_v} - I_{j_v}| + w_l |I_{i_l} - I_{j_l}| + w_r |I_{i_r} - I_{j_r}| \quad (8)$$

$$D_{ij}^{MV} = |\mathbf{MV}_i^x - \mathbf{MV}_j^x| + |\mathbf{MV}_i^y - \mathbf{MV}_j^y| \quad (9)$$

식 (4)의 마지막 항, 포텐셜 함수 $V^T(f_i|f_{i-1})$ 는 시간적 상관성을 부여하고, 이전 프레임에서의 해당 영역의 정보를 이용하여 계산된다. 식 (10)과 같이 영역의 색상, 움직임, 크기 등의 정보에 의해서 결정된다.

$$V^T(f_i|f_{i-1}) = \begin{cases} -\gamma D_{i,i-1}^{Color} D_{i,i-1}^{MV} N_i, & f_i = f_{i-1} \\ \gamma D_{i,i-1}^{Color} D_{i,i-1}^{MV} N_i, & f_i \neq f_{i-1} \end{cases} \quad (10)$$

여기에서 에너지 함수 $U(f|d)$ 를 최소화하여 객체와 배경을 결정하는 최적 레이블 셋 f' 을 얻기 위해 Highest Confidence First (HCF) 알고리즘[9]을 사용하였다. 그림 3은 영역의 움직임을 이용하여 객체와 배경을 판단한 결과와 MRF 를 이용한 객체 추출 결과를 비교하여 보여주고 있다. MRF 를 이용한 결과는 영상에서 잡음에 의한 영향을 최소화하여 강인하게 객체를 추출함을 확인할 수 있다.

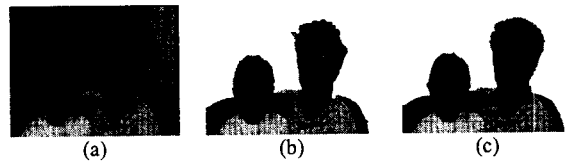


그림 3. MRF 를 이용한 객체 추출 결과
(a) 원 영상 (b) 영역의 움직임을 이용한 객체와 배경 판단 결과 (c) MRF 를 이용한 객체 추출 결과

V. 실험 결과

제안된 방법의 성능을 평가하기 위해 객체의 외형 변화 또는 카메라 움직임이 있는 다양한 동영상에 대하여 실험을 수행하였다. 실험 환경은 Pentium 4 1.8GHz, RAM 512M 의 IBM 호환 PC 상에서 수행하였으며, OS 는 Window XP, 개발툴은 Microsoft™ Visual C++ 6.0 을 사용하였다.

그림 4는 객체의 외형 변화가 있는 동영상에서 실험한 결과를 보여주고 있다. 먼저 첫 프레임에서 사용자

가 원하는 객체를 지정하고, 제안된 방법을 이용하여 객체를 추출 및 추적하였다. 그림 4에 보인 바와 같이 객체의 외형 변화에 강인하게 객체를 추출 및 추적을 확인하였다. 그림 5에서는 객체의 외형 변화와 카메라 움직임이 있는 동영상에 대한 실험 결과를 나타내고 있다. 카메라 움직임과 객체의 외형 변화가 동시에 일어나는 동영상에서 프레임당 평균 0.2 초의 우수한 속도로 객체를 정확하게 추출 및 추적함으로써 제안한 MRF를 이용한 방법의 성능을 확인하였다.

객체 영역에 의해 가려져 있던 배경 영역이 나타나는 경우, 그림 5의 마지막 결과와 같이 객체 주변에서 약간의 잡음이 나타나는 문제점이 있다. 이러한 문제점은 영역 병합 단계에서 객체 영역이 주변의 배경 영역과 색상이 유사하여 잘못 병합이 이루어진 경우로 판단할 수 있다.

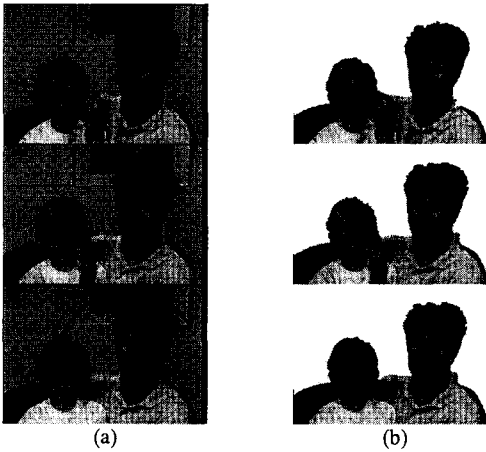


그림 4. 객체의 외형 변화가 있는 동영상
(a) 원본 동영상(M&A) (b) 객체 추출 및 추적 결과



(a) (b)

그림 5. 객체의 외형 변화와 카메라 움직임이 있는 동영상
(a) 원본 동영상 (b) 객체 추출 및 추적 결과

VI. 결론

본 논문에서는 블록 기반 움직임 추출 알고리즘인 Diamond Search (DS)를 분할된 영역에 적용하여 영역의 움직임 벡터를 추출하고, 추출된 움직임을 이용하여 객체 영역과 배경 영역을 판단한 결과에서 영역 정보의 시간적·공간적 상관성에 기반한 MRF 모델을 정의하여 객체 추출 및 추적의 정확성을 향상시키는 방법을 제안하였다. 또한, DS, HCF 등의 고속 알고리즘들을 사용하여 영역 기반 객체 추출 및 추적 방법에서 요구되는 계산 시간을 최소화하였다. 실험 결과 다양한 동영상에서 프레임당 평균 0.2 초의 빠른 속도로 강인하게 객체를 추출 및 추적함으로써 제안한 방법의 우수성을 확인하였다.

제안한 방법은 사용자가 초기 객체 영역을 설정하는 불편함이 있지만, 움직임이 거의 없는 의미 있는 객체를 정확하게 추출 및 추적하는 장점을 가지고 있다. 향후에는 영역의 움직임 정보를 이용하여 영역 병합의 문제점을 해결함으로써 객체 추출의 정확성을 향상시키는 연구가 진행되어야 할 것이다.

참고 문헌

- [1] S. Berretti, A. D. Bimbo and E. Vicario, "Efficient matching and indexing of graph models in content-based retrieval," *IEEE Trans. on PAMI*, vol. 23, no. 10, pp. 1089-1105, Oct 2001
- [2] J. Rehg and T. Kanade, "Model-Based Tracking of Self-Occluding Articulated Objects," *Proc. IEEE Conf. CVPR*, pp. 612-617, June 1995.
- [3] Y. Zhong, A. K. Jain and M. -P. Dubuisson-Jolly, "Object Tracking Using Deformable Templates," *IEEE Trans. on PAMI*, vol. 22, no. 5, pp. 544-549, May 2000.
- [4] Y. Tsai and A. Averbuch, "Automatic Segmentation of Moving Objects in Video Sequences: A Region Labeling Approach," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 12, no. 7, pp. 597-612, July 2002.
- [5] P. De Smet and D. De Vleeschauwer, "Performance and Scalability of a highly optimized rainfalling watershed algorithm," *Proc. Int. Conf. on Image Science, Systems and technology. CISST 98, Las Vegas, NV, USA*, July 1998, pp. 266-273.
- [6] P. Perona and J. Malik, "Scale Space and Edge Detection Using Anisotropic Diffusion," *IEEE Trans. PAMI*, Vol. 12, No. 7, pp. 629-639, July 1990.
- [7] 황선규, 권영진, 김희율, "영상 분할을 위한 효과적인 영역 병합 방법," *제 15 회 신호처리 합동학술대회 논문집*, pp. 73, 2002년 9월.
- [8] A. M. Tourapis, O. C. Au and M. L. Liou, "Predictive Motion Vector Field Adaptive Search Technique (PMVFAST) - Enhancing Block Based Motion Estimation," *Proc. on VCIP-2001*, San Jose, CA, January 2001.
- [9] S. Z. Li, *Makov Random Field Modeling in Image Analysis*, Springer, 2001.