

약한 지도 학습의 다중 랜덤워크 기반 동영상 객체 분할

허민혁, 임경선, 김한울, 고영준, 김창수

고려대학교

{mhheo, kslim, hanulkim, yjkoh, cskim}@mcl.korea.ac.kr

Weekly Supervised Video Object Segmentation based on Multiple Random Walker

Minhyeok Heo Kyungsun Lim Han-Ul Kim Yeong Jun Koh Chang-Su Kim
Korea University

요 약

본 논문에서는 간단한 사용자 입력과 다중 랜덤 워크(multiple random walker) 기법을 기반으로 동영상 내의 주요 객체를 분할하는 알고리즘을 제안한다. 우선 동영상의 첫 프레임에서 점 형태의 사용자의 입력을 받아 대략적인 객체와 배경의 위치를 얻고, Lab 색상의 측지거리를 이용하여 객체와 배경의 중요도 지도를 얻는다. 다음으로 영상을 슈퍼 픽셀 단위로 분할하고, 다중 랜덤 워크 기법을 적용하여 객체 분할을 수행한다. 랜덤 워크 기법 적용 시, 중요도 지도를 각 랜덤 워커의 초기 분포로 설정하고, 노드간 색상과 움직임 차이를 이용하여 전이 행렬을 계산한다. 마지막으로 결과를 정련한 뒤, 다음 프레임으로 분할 결과를 전파하여 시간적 일관성을 유지한다. 실험을 통하여 제안 기법이 기존 기법에 비하여 우수한 객체 분할 성능을 보임을 확인한다.

1. 서론

동영상 객체 분할 기법은 입력 동영상을 객체와 배경으로 분리하는 기법이다. 동영상 내의 주요 객체를 추출함으로써 영상의 특징을 정밀하게 표현할 수 있기 때문에 객체 검출, 추적 등의 고차원 컴퓨터 비전 알고리즘의 성능 향상을 위하여 활용된다. 정밀한 객체 분할을 위하여 다양한 기법들이 개발되었는데 그 중 다수의 랜덤 워커들의 반발 작용을 이용한 다중 랜덤 워크 기법 [1, 2]도 우수한 성능을 보이고 있다.

한편 다양한 객체가 포함된 동영상의 경우, 사전 정보 없이는 원하는 객체를 추출하는데 어려움이 있다. 이러한 문제를 해결하기 위하여 동영상 내에서 움직임이 큰 객체나 지속적으로 등장하는 객체를 찾아서 분할하는 기법들이 제안되었지만 동영상 전체를 활용하기 때문에 실시간 동작이 불가능하다는 한계가 있다. 동영상의 첫 프레임에서 박스 형태의 사용자 입력을 받는 알고리즘[2]의 경우에는 다양한 객체의 모양을 제대로 반영할 수 없다는 문제가 있다.

본 논문에서는 점 형태의 간단한 사용자 입력과 다중 랜덤 워크 기법을 이용해 동영상에서 이용자가 원하는 주요 객체를 분할하는 기법을 제안한다. 첫 프레임의 사용자 입력으로 얻어진 객체와 배경의 중요도 지도와 다중 랜덤워크 기법을 이용하여 주요 객체를 분할한다. 객체 분할 결과를 픽셀 단위로 정련한 뒤 다음 프레임으로 전파하여 분할 결과의 시간적 일관성을 유지한다.

본 논문의 구성은 다음과 같다. 2 절에서는 제안하는

기법에 대해 살펴보고, 3 절에서는 제안 기법의 성능을 실험을 통해 기존 알고리즘과 비교한다. 마지막으로, 4 절에서는 본 논문의 결론을 맺는다.

2. 제안하는 기법

2.1 중요도 지도

입력 동영상의 첫 프레임에서 점 형태의 사용자의 입력을 이용하여 객체와 배경 각각의 중요도 지도를 구한다. 먼저, SLIC 알고리즘[3]을 이용하여 동영상의 각 프레임을 슈퍼 픽셀로 분할하고, 슈퍼 픽셀을 하나의 노드로 하는 그래프를 구성한다. 노드 간의 유사도는 아래와 같이 계산한다.

$$w_{ij} = \exp\left(-\frac{(1-\lambda)d_{Lab}(i,j) + \lambda d_{Flow}(i,j)}{\sigma^2}\right) \quad (1)$$

여기서 d_{Lab}, d_{Flow} 는 각각 Lab 색상, 광흐름의 유클리드 거리를 의미하고 λ, σ^2 는 스케일 파라미터로 0.4, 0.005로 실험하였다.

그림 1(a)와 같이 사용자로부터 객체와 배경이 존재하는 부분에 점 형태의 입력을 받는다. 주어진 객체와 배경 입력 픽셀로부터 Lab 색상의 측지 거리를 계산하는 알고리즘 [4]을 이용하여 객체와 배경의 중요도 지도를 구한다(그림 1(b), (c)).

2.2 다중 랜덤워크 기반 객체 분할

본 논문에서는 [1]에서 제안된 다중 랜덤워크 기반의 객체 분할 프레임 워크를 따른다. 객체와 배경을 나타내는 두 명의 랜덤 워커를 기용하고, 매 반복마다 k 랜덤 워커의 확률분포

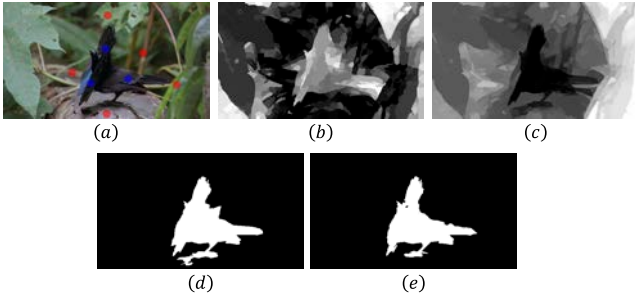


그림 1. 제안 기법의 각 과정별 결과: (a) 입력 영상, (b) 객체 중요도 지도, (c) 배경 중요도 지도, (d) 객체 분할 결과, (e) 결과 정련

$\mathbf{p}_k^t = [p_{1,k}^t, \dots, p_{N,k}^t]^T$ 다음과 같이 계산한다.

$$\mathbf{p}_k^{t+1} = (1 - \epsilon - \gamma)\mathbf{A}\mathbf{p}_k^t + \epsilon\mathbf{r}_k^t + \gamma\mathbf{c}_k \quad (2)$$

\mathbf{A} 는 전이 행렬로써 노드 간의 유사도 행렬 $\mathbf{W} = [w_{ij}]$ 를 열방향으로 정규화하여 얻는다. \mathbf{r}^t 는 일정 확률로 각 위커를 정해진 위치로 이동시키는 재시작 분포이다. 본 논문에서는 [1] 알고리즘에서 제안한 반발 재시작 분포를 이용한다. \mathbf{c} 는 사용자가 입력한 점을 포함하는 슈퍼 픽셀에 대해서만 1의 값을 부여한 뒤 정규화하여 사용자가 입력한 부분에 높은 확률을 부여하는 항이다. 객체와 배경 랜덤 위커의 초기 분포 $\mathbf{p}_f^0, \mathbf{p}_b^0$ 은 각각 앞서 얻어진 객체, 배경 중요도 지도를 슈퍼 픽셀 단위로 평균을 취하고 정규화하여 구한다. ϵ, γ 는 0.05로 실험하였다.

반복을 수행하여 각 랜덤 위커들이 정상 상태에 도달하였을 때, 각 노드에서 객체 랜덤 위커의 확률이 배경 랜덤 위커의 확률보다 높은 경우, 객체라고 판단하여 객체 분할 결과(그림 1(d))를 얻는다.

2.3 결과 정련 및 전파

슈퍼 픽셀 단위의 결과를 픽셀 단위로 정련한다. 먼저 경계 추정 알고리즘인 HED [5]를 이용하여 영상의 경계를 추정한다. 경계와 인접한 픽셀에서의 분할 결과를 인접한 주변 픽셀의 결과의 가중합으로 계산하고, 문턱 값을 적용하여 정련된 결과(그림 1(e))를 얻는다. 인접 픽셀간 가중치는 (1)과 동일하게 색상과 움직임의 유클리드 거리를 이용해 구한다.

정련된 객체 분할 결과를 광흐름을 이용하여 다음 프레임으로 와핑(warping)한다. 와핑된 결과를 객체 랜덤 위커의 초기 분포로 설정하고, 영상의 경계 영역을 배경 랜덤 위커의 초기 분포로 설정한 뒤 앞선 과정을 반복하여 동영상에서 시간적 일관성이 있는 객체 분할 결과를 얻는다.

3. 실험 결과

제안 기법의 객체 분할 성능 측정을 위해 상용화된 데이터셋인 SegTrack v2에 대하여 제안 기법을 적용하고 결과를 관찰하였다. 또한, 본 실험에서는 기존 기법과의 비교를 위해 비교 기법으로 동영상 객체 분할 분야의 최신 기법인 Ramakanth [6]의 기법을 사용하였다. 실험의 공정성을 위하여 [6] 알고리즘은 첫 프레임에서 ground truth를 사용자 입력으로 설정하였다.

표 1과 그림 2는 비교 기법과 제안 기법의 정량적, 정성적 결과를 도시한다. 결과에서 확인할 수 있듯이, 제안

표 1. 비교 기법과의 정량적 성능 비교.

	제안 기법	비교 기법 [6]
IOU	66.0	49.7

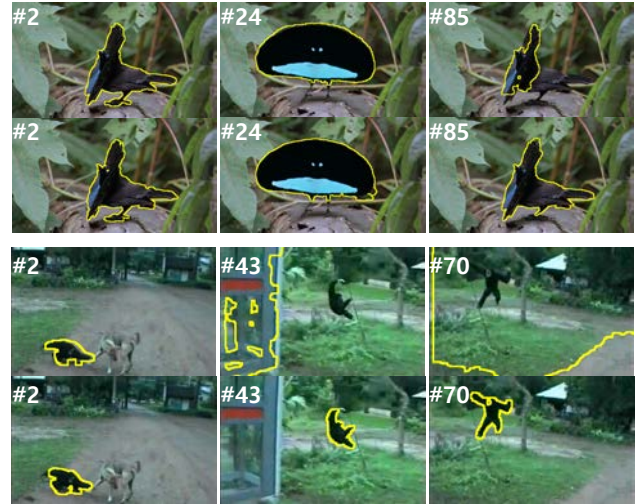


그림 2. 비교 기법과의 정성적 결과 비교. (위) 비교 기법, (아래) 제안 기법

기법이 기존 기법 대비 우수한 객체 분할 결과를 보인다.

4. 결론

본 논문에서는 다중 랜덤워크 기법을 이용하여 동영상 내의 주요 객체를 분할하는 기법을 제안하였다. 우선 점 형태의 간단한 사용자 입력을 통하여 중요도 지도를 계산한다. 다음으로, 다중 랜덤워크 기법을 이용하여 객체를 분할한 뒤, 결과를 정련하고 다음 프레임으로 전파하여 동영상에서의 객체 분할을 수행한다. 실험을 통해 제안 기법이 기존 기법 대비 우수한 성능을 보임을 확인하였다.

5. 참고 문헌

- [1] C. Lee, W.-D. Jang, J.-Y. Sim, and C.-S. Kim, "Multiple random walkers and their application to image cosegmentation," in *Proc. IEEE ICCV*, 2015.
- [2] M. Heo, W.-D. Jang, and C.-S. Kim, "Video object segmentation using multiple random walkers with GMM restart rule," in *Proc. IEEE APSIPA*, 2016.
- [3] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," in *IEEE TPAMI*, 2012.
- [4] Y. Wei, F. Wen, W. Zhu, and J. Sun, "Geodesic saliency using background priors," in *Proc. IEEE ECCV*, 2012.
- [5] S. Xie and Z. Tu, "Holistically-nested edge detection," in *Proc. IEEE ICCV*, 2015.
- [6] S. A. Ramakanth, and R. V. Babu, "Seamseg: Video object segmentation using patch seams," in *Proc. IEEE CVPR*, 2014.