

근거리 전경 물체를 위한 파노라마 알고리즘

이순영 †심재영 이상욱

서울대학교 대학원 전기공학부 뉴미디어통신공동연구소

†울산과학기술대학교 전기전자컴퓨터공학부

sylee@ipl.snu.ac.kr jysim@unist.ac.kr sanguk@ipl.snu.ac.kr

A panorama algorithm for close foreground objects

Soon-Young Lee †Jae-Young Sim Sang-Uk Lee

School of Electrical Engineering and INMC, Seoul National University, Korea

†School of Electrical and Computer Engineering, UNIST, Korea

요약

본 논문에서는 멀티뷰 환경에서 촬영된 영상을 이용하여 카메라와 가까이 위치한 전경 물체를 중심으로 뷰가 확장된 영상을 생성하는 파노라마 알고리즘을 제안하였다. 먼저 전경 물체는 사용자가 선택한 중심 물체를 기준으로 하여 평행 이동 관계로 전경 파노라마를 생성한다. 배경은 특징점 기반의 파노라마 기법을 이용하여 초기 결과를 얻고, 멀티뷰 영상의 시차에 기인한 가려진(occluded) 영역을 제한 영역으로 설정한 후 최적화 과정을 이용하여 시각적으로 자연스러운 배경 파노라마를 얻는다. 마지막으로 배경 파노라마와 전경 파노라마와 크기를 동일하게 하고 배경 파노라마와 전경 파노라마를 합성하여 결과 파노라마를 구성한다. 모의 실험결과 제안 알고리즘은 원본 영상의 전경 물체들을 모두 포함하면서도 배경 부분은 시각적으로 자연스러운 파노라마 결과를 도출함을 확인하였다.

1. 서론

영상 파노라마 기법은 한정된 뷰를 제공하는 싱글뷰 기반의 카메라로 촬영한 일반 사진의 한계를 극복하기 위해 특별한 하드웨어 없이 뷰를 확장하여 실제감을 높이는 기법이다 [1-3]. 초기의 영상 파노라마 기법은 원본 영상들 사이에서 호모그래피(homography)를 추정하고 원본 영상들을 공통 평면에 평면 변환(planar transform)을 통해 투영하는 기법이 널리 사용되었다 [1,2]. 세부적으로는 전체 픽셀을 참고하여 호모그래피를 추정하거나 [1], 영상 내에서 시각적으로 중요한 특징점(feature point)을 추출하여 구한 호모그래피로 전체 원본 영상을 투영하는 기법이 연구되었다 [2]. 하지만 이들 기법은 투영된 결과 영상이 같은 깊이(depth)를 가진 평면으로 가정하기 때문에 전경과 배경이 다른 깊이를 가지는 경우는 왜곡된 결과를 야기한다. Cohen *et al.* [3]은 원본 영상을 여러 개의 층(layer)로 구분한 후 층별로 독립적으로 계산하여 서로 다른 깊이를 갖는 영상에 대한 파노라마 기법을 제안하였다. 하지만 이 기법은 깊이 정보에 기반하고 있으므로, 깊이 예측의 정확도에 따라 성능이 크게 좌우되는 단점을 지닌다.

본 논문에서는 멀티뷰 환경에서 카메라와 근거리에서 위치한 전경 물체를 포함한 영상을 대상으로 하여, 전경 물체를 모두 포함하도록 확장된 뷰를 제공하는 파노라마 기법을 제안한다. 먼저 근경에 위치한 전경 물체를 평행 이동 관계로 추정하고, 사용자가 지정한 중심 물체를 기준으로 하여 각 뷰에서의 물체를 배타적으로 합성하여 전경 파노라마를 구성한다. 배경 파노라마는 원본 영상을 특징점을 이용한 평면 변환 파노라마 기법으로 초기화한 후, 결과 영상의 각 픽셀에 대응하는 원본 영상의 픽셀을 최적화 과정을 통해 구한다. 멀티뷰 환경의 시차로

인해 전경 물체에 가려진 배경 영역이 다른 뷰에서 관찰되는 경우 이를 최적화 과정에 이용함으로써 배경 파노라마의 공간적 연속성이 유지된다. 마지막으로 배경 파노라마의 크기를 전경 파노라마와 일치시키기 위해 틈(seam)을 삽입하고, 두 파노라마를 합성하여 결과 파노라마를 얻는다.

2. 제안 알고리즘

본 논문은 멀티뷰 환경에서 뷰가 수평적으로 2개로 구성된 경우를 가정한다. 서로 다른 뷰의 원본 영상 중 왼쪽 영상을 S_1 , 오른쪽 영상을 S_2 라 할 때, 우선 물체들을 세그멘테이션 기법을 이용하여 분리한다. 그리고 사용자가 모든 뷰에서 공통으로 관찰되는 물체 중 하나를 중심 물체 Ω_1^p , Ω_2^p 로 선택한다. 중심 물체 Ω_2^p 에서 Ω_1^p 로 대응하는 평행 이동 관계를 블록 기반 매칭 기법으로 구하여 이를 t 로 나타낸다 [4]. 전경 파노라마 P_f 는 Ω_1^p 과 Ω_2^p 를 중심으로 왼쪽에 있는 물체와, t 만큼 이동시킨 Ω_2^p 의 오른쪽에 있는 물체를 합집합하여 구성한다. 따라서 이 전경 파노라마는 근거리에서 있는 전경 물체를 모두 포함한다. 또한 중심 물체로 카메라와 가까이 위치한 물체를 선택하기 때문에 시차(disparity)가 크고 이로 인해 P_f 의 가로 길이가 확장되는 효과를 얻는다.

배경 파노라마 P_b 는 원본 영상 S_1 , S_2 에서 전경 물체를 제외하고 남은 홀(hole)영역을 시각적으로 자연스럽게 복원하면서 동시에 원본 영상의 뷰를 확장하여 얻는다. 제안 알고리즘은 배경 파노라마 P_b 의 각 픽셀에 원본 영상의 픽셀을 대응시키는 이산 조합 문제(discrete combinatorial problem)으로 설계하여 P_b 를 구한다.

S_1 , S_2 에서 물체를 제거한 순수한 배경 영상을 각각 B_1 , B_2 라고 할 때, 먼저 두 배경을 특징점 기반의 평면 변환을 구해 공통 평면으로 투영하고 이를 각각 B_1' , B_2' 라 정

의한다 [2]. B'_1, B'_2 는 제한 영역(constraint area) Φ 를 제공하여 이산 조합 문제에서의 해를 결정한다. 이 Φ 는 배경 파노라마 P_b 가 반드시 포함해야 하는 영역으로, B'_1 의 가장 왼쪽 열의 픽셀과 B'_2 의 가장 오른쪽 열의 픽셀, 그리고 한 뷰에서 물체가 있던 영역이 다른 뷰에서 관찰되는 영역이다. 또한 B'_1, B'_2 는 P_b 의 픽셀값을 가지고 올소스(source) 영상으로 사용되는데, 간결한 표현을 위해 B'_1, B'_2 를 수평으로 나란히 붙인 영상을 소스 영상 B' 이라 정의한다.

배경 파노라마 P_b 의 위치 \mathbf{p} 에서의 픽셀값은 다음과 같이 B' 에서 변위 벡터 $\mathbf{d}(\mathbf{p})$ 만큼 떨어진 픽셀값으로 나타낸다.

$$P_b(\mathbf{p}) = B'(\mathbf{p} + \mathbf{d}(\mathbf{p})). \quad (1)$$

여기서 위치 \mathbf{p} 를 노드(node), 변위 벡터 $\mathbf{d}(\mathbf{p})$ 를 레이블(label)로 볼 수 있고, 이는 전통적인 그래프 레이블링 문제(graph labeling)이다 [5]. 모든 위치 \mathbf{p} 에 대해 최적의 $\mathbf{d}(\mathbf{p})$ 를 결정하는 것은 다음의 에너지를 최소로 하는 해를 구하는 것으로 풀 수 있다.

$$E(\mathbf{d}) = \sum_{\mathbf{p}} E_d(\mathbf{d}(\mathbf{p})) + \lambda \sum_{\mathbf{q} \in N(\mathbf{p})} E_s(\mathbf{d}(\mathbf{p}), \mathbf{d}(\mathbf{q})), \quad (2)$$

여기서 E_d 는 제한 조건에 대한 단항 에너지 항(unitary energy term)이고 E_s 는 시각적 일관성(visual consistency)을 나타내는 이항 에너지 항(pair-wise energy term)이다. 또한 \mathbf{p} 와 \mathbf{q} 는 위치 벡터, $N(\mathbf{p})$ 는 \mathbf{p} 의 4방향 이웃 픽셀의 집합, λ 는 가중 상수이다.

E_d 는 제한 영역 Φ 에 관련된 에너지 항으로 다음과 같다.

$$E_d(\mathbf{d}(\mathbf{p})) = \begin{cases} 0, & \mathbf{p} \in \Phi \text{ and } \mathbf{p} + \mathbf{d}(\mathbf{p}) = \mathbf{u}, \\ \infty, & \mathbf{p} \in \Phi \text{ and } \mathbf{p} + \mathbf{d}(\mathbf{p}) \neq \mathbf{u}, \\ 0, & \mathbf{p} \notin \Phi, \end{cases} \quad (3)$$

여기서 \mathbf{u} 는 Φ 에 있는 \mathbf{p} 에 대응하는 B' 에서의 위치이다. E_s 는 P_b 에서의 시각적 연속성을 나타내는 항으로써, 픽셀의 RGB 색상값의 차이의 합을 이용하여 다음과 같이 표현된다 [6].

$$\begin{aligned} \sum_{\mathbf{q} \in N(\mathbf{p})} E_s(\mathbf{d}(\mathbf{p}), \mathbf{d}(\mathbf{q})) \\ = \sum_{\mathbf{p}} \sum_i |P_b(\mathbf{p} + \mathbf{e}_i) - B'(\mathbf{p} + \mathbf{d}(\mathbf{p}) + \mathbf{e}_i)|^2, \end{aligned} \quad (4)$$

여기서 \mathbf{e}_i 는 2차원에서 4방향의 단위 벡터를 나타낸다.

다음으로 배경 파노라마 P_b 를 전경 파노라마 P_f 와 합성하기 위해 두 파노라마의 크기를 동기화 한다. 일반적으로 멀티뷰 환경의 시차로 인해 P_b 보다 P_f 의 가로 길이가 더 크다. 따라서 틸삽입(seam insertion) 기법을 이용하여 시각적으로 자연스럽게 P_b 를 확장시켜 P_f 와 크기를 같게한다 [7]. 이 때, 제한 영역 Φ 은 원본 영상에서 기준으로 삼았던 부분이므로 틸을 삽입하지 않고 원래 영상 품질을 유지하도록 한다. 마지막으로 확장된 배경 파노라마 위에 전경 파노라마를 위치시켜 결과 파노라마를 도출한다.

3. 실험 결과

실험에 사용된 데이터는 "Ballroom" 멀티뷰 데이터이다. 해상도는 320x240이고 8개의 수평으로 배열된 뷰를 제공하는데, 그 중에서 그림 1(a), (b)와 같이 가장 좌우의 두 뷰를 이용하였다. 중심물체는 가장 근경에 있는 커플이고 $\mathbf{t} = (87, 0)$ 의 이동량을 가진다. 그림 1(c)에서 보듯

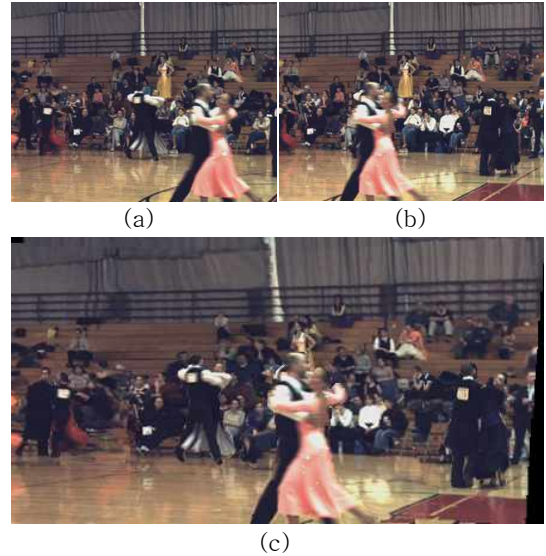


그림1. (a)(b) 원본 영상 S_1, S_2 , (c) 제안 알고리즘결과.

이, 전경 물체의 위치가 보존되면서도 배경이 확장되어 한 프레임에 전경 물체 모두를 포함할 수 있다. 전경 물체가 삭제되고 틸삽입으로 확장된 배경 파노라마는 시각적으로 자연스러운 결과를 보인다.

4. 결론

본 논문에서는 근경에 위치한 물체를 중심으로 뷰가 확장된 영상을 생성하는 파노라마 알고리즘을 제안하였다. 제안 알고리즘은 배경과 상이한 깊이를 가진 전경 물체들을 효과적으로 파노라마 하기 위해 전경과 배경을 분리하여 전경은 사용자가 선택한 중심물체를 기준으로 파노라마하고, 배경은 픽셀 단위의 이산 조합 문제로 최적화 과정을 통해 파노라마 하였다. 실험 결과에서 원본 영상의 전경 물체를 한 프레임에 포함하면서도 시각적으로 자연스러운 파노라마 결과를 얻었다.

제안 알고리즘은 기존 기법들에서 다루기 힘들었던 전경과 배경의 깊이 차이가 큰 영상을 대상으로 파노라마를 수행한 것에 의미가 있다. 향후 시간축의 정보를 고려하여 비디오 데이터로의 확장, 사용자의 개입을 최소화하는 인터페이스 개발이 필요할 것이다.

참고문헌

- [1] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, vol. 16, no. 2, pp. 22–30, Mar. 1996.
- [2] David G. Lowe, "Distinctive image features from scale invariant key points," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Jan. 2004.
- [3] M. F. Cohen K. C. Zheng, and R. Szeliski, "Layered depth panoramas," in *Proc. IEEE CVPR*, pp. 1009–1016, Oct. 2007.
- [4] Y. Wang, J. Ostermann, and Y.-Q. Zhang, *Video Processing and Communications*, Prentice Hall, 2002, pp. 152–155.
- [5] O. Veksler Y. Boykov, and R. Zabih, "Markov random fields with efficient approximations," in *Proc. IEEE CVPR*, pp. 648–655, June 1998.
- [6] Y. Pritch, E. Kav-Venaki, and S. Peleg, "Shift-map image editing," in *Proc. ICCV*, Sept. 2009, pp. 151–158.
- [7] S. Avidan and A. Shamir, "Seam carving for content-aware image resizing," *ACM Trans. Graph.*, vol. 26, no. 3, pp. 10(1)–10(8), July 2007.