

시간적으로 증가하는 가중치 기반의 파노라마 영상 구성

이종재, 장석우, 최형일
송실대학교 컴퓨터학부

Construction of Panorama Image Based on Temporally Incremental Weight

Joong-Jae Lee, Seok-Woo Jang, Hyung-Il Choi
School of Computing, Soongsil University

요 약

파노라마 영상은 카메라의 동작을 고려하여 공간적으로 영상을 정렬한 영상으로서 효과적으로 비디오 데이터를 표현할 수 있는 방법이다. 그러나 기존의 파노라마 영상 구성 기법들은 영상을 정렬할 때 오버랩(overlap)되는 화소값들을 단순히 미디언 필터링하기 때문에 시간의 경과에 따른 밝기값의 동적인 변화를 수용하지 못한다. 따라서 본 논문에서는 이런 문제점을 해결하기 위해 시간적으로 증가하는 가중치를 사용하여 영상을 통합함으로써 최근의 영상 정보를 충분히 반영하는 파노라마 영상 구성 방법을 제안한다.

1. 서 론

우리가 일상생활에서 흔히 접할 수 있는 비디오 데이터는 그 중요성이 강조되기 시작하면서 표현 방법에 있어서도 다양한 연구가 이루어지고 있다. 이 중 파노라마 영상은 주어진 입력 영상 사이의 카메라 동작을 추출하고 보정한 뒤 하나의 기준 좌표계에 시공간적으로 정렬한 영상으로서 비디오 데이터에 대한 효과적인 표현 방법 중의 하나라고 할 수 있다. 이런 파노라마에 대한 연구가 활발히 진행되고 있는 이유는 파노라마 영상이 다음과 같은 장점들을 가지고 있기 때문이다[1].

첫째, 파노라마 영상은 연속된 영상 프레임을 하나의 큰 뷰(view)를 갖는 영상으로 통합해서 일종의 스냅샷 효과를 얻을 수 있다. 즉, 카메라의 제한된 뷰를 사람의 눈과 같은 보다 큰 뷰로 영상을 재구성할 수 있게 되므로 영상 검색에 있어서 매우 유리하다.

둘째, 연속된 영상 프레임을 일정한 기준에 의해 오버랩(overlap)하여 파노라마를 구성하기 때문에 저장해야 할 데이터의 양을 상당히 줄일 수 있다.

이러한 파노라마를 구성하는 과정은 다음과 같이 크게 세 부분으로 나뉘어진다[1][4]. 즉, 입력 영상 프레임들 기준 영상 프레임에 정렬(alignment)하는 부분, 정렬된 영상 프레임들 파노라마 영상으로 통합(integration)하는 부분, 그리고 파노라마 영상과 각각의 입력 영상 프레임 사이의 나머지(residual)를 계산하는 부분이 필요하다.

위와 같은 일련의 과정을 통해 만들어진 파노라마 영상이 효과적인 비디오 데이터 표현 방법이 되기 위해서는 무엇보다도 정확한 파노라마 영상 획득이 요구된다. 이런 관점에서 정확한 카메라의 동작 파라미터 추출이 필수적이며, 이를 통한 입력 영

상에 대한 카메라 동작 보정, 즉 파노라마를 구성하는 연속된 영상 프레임의 정렬이 파노라마 영상 구성의 기본이 된다. 또한, 정렬이 완료된 입력 영상 자체에 카메라 동작을 통해 발생하는 왜곡 현상이 포함되어 있을 수 있기 때문에 입력 영상을 통합하기 전에 영상 변형(warping) 기법을 활용해 파노라마 영상의 정확성을 높일 수 있다[3].

파노라마를 구성하는 기존 연구에서 지적되는 문제점은 정확한 카메라 동작 파라미터를 추출하기 위해서 사용되는 비선형 최적화 기법에 너무 많은 시간이 소요된다는 것과 고해상도 입력 영상에 대해 발생하는 지역적 최소화(Local minima)문제이다 [1]. 또한 파노라마 구성 과정의 하나인 영상통합시 기존의 연구에서는 대부분 미디언 필터링을 이용하고 있는데[2][3] 이 방법은 시간의 경과에 따라 동적으로 변하는 밝기 값의 변화를 고려하지 않기 때문에 최근의 정보를 반영하지 못하는 문제점을 가지고 있다.

본 논문에서는 위와 같은 문제점을 해결하기 위해 블록 정합 알고리즘(Block Matching Algorithm)을 사용해 시간 복잡도를 줄이는 한편 영상 통합시 최근의 영상일수록 가중치를 더 많이 부여하는 방법을 사용하여 최근의 정보를 반영할 수 있도록 했다.

그림 1은 입력 영상으로부터 파노라마 영상을 구성하는 과정을 도식적으로 나타낸 전체 개요도이다. 그림 1과 같이 파노라마 영상을 구성하는 과정은 크게 입력영상을 정렬하는 부분과 파노라마 영상을 구성하는 두 부분으로 나뉘어 진다.

본 논문의 구성은 다음과 같다. 2장에서 블록 정합 알고리즘을 이용한 카메라의 동작 파라미터 추출에 대해 설명하고, 3장에서는 카메라 동작을 보정한 입력 영상을 통한 파노라마 구성에

대해 설명한다. 그리고 4장에서는 실험 결과에 대해서 기술하며 5장에서는 본 논문의 결론을 맺는다.

파라미터를 구하는데 드는 시간이 너무 많이 걸린다는 점과 고 해상도의 입력 영상에 대해 지역적 최소화(local minima) 문제에 빠질수 있다는 문제점을 안고 있다. 본 논문에서는 이런 문제점을 해결하기 위해 블록 정합 알고리즘을 사용한다.

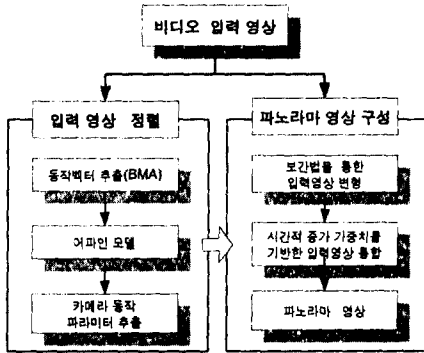


그림 1. 파노라마 영상 구성에 대한 전체 개요도

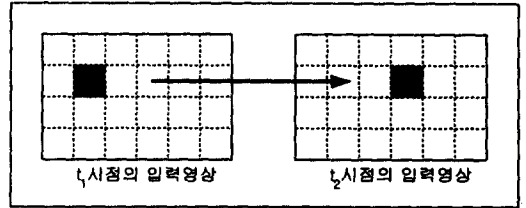


그림 3. 블록 정합

2. 카메라의 동작 파라미터 추출

정확한 파노라마 구성을 위해서는 연속된 입력 영상이 기준 좌표계에 대해서 정확하게 정렬되어야 한다. 입력 영상 자체에 카메라의 동작이 포함되지 않았을 경우, 즉 배경이 움직이지 않을 경우에는 문제가 되지 않지만 카메라 동작이 포함된 입력 영상에 대해서는 카메라 동작 파라미터를 정확하게 추출한 뒤 그에 따른 보정 작업이 필요하다.

카메라 동작 파라미터를 추출해 내기 위해 사용되는 동작 모델에는 2차원 파라미터 동작 모델과 복잡한 3차원 동작 모델이 사용되고 있다[1]. 본 논문에서는 2차원 동작 모델 중 비교적 간단하면서도 카메라 동작을 잘 설명할 수 있는 어파인 모델을 사용하기로 한다[5].

이 알고리즘은 그림 3과 같이 입력 영상을 일정한 크기의 블록으로 분할한 다음 각각의 블록에 대해 다음 시점의 입력 영상에서 대응되는 블록의 위치를 찾아내는 방법으로서 앞서 지적한 문제점들을 보완할 수 있다.

카메라의 동작 파라미터는 블록 정합 알고리즘을 통해 구한 동작 벡터들을 식(1)과 같이 정의된 어파인 모델을 적용하여 구할 수 있다.

3. 파노라마 영상 구성

파노라마 영상을 구성하기 위해서는 카메라의 동작을 보정한 후 영상들을 하나의 기준 좌표계에 정렬해야 한다. 또한 이 때 카메라로 인해 발생하는 왜곡현상 역시 고려되어야 만 좀 더 정확한 파노라마 영상을 획득할 수 있다. 이 점을 고려해 본 논문에서는 역방향 양선형 보간법을 이용한 입력영상 변형방법을 사용한다.

이 방법은 카메라 동작을 보정할 때 입력영상의 한 화소가 한 개 이상의 화소에 매핑되거나 정확하게 격자에 매핑되지 않는 문제점을 해결하기 위해 그림 4와 같이 주위에서 가장 가까운 네 개의 화소값으로 보간을 해서 영상을 변형하는 방법이다.

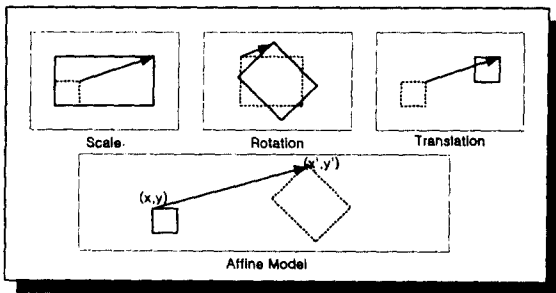


그림 2. 어파인 동작 모델

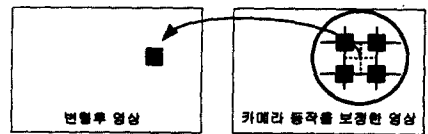


그림 4. 역방향 양선형 보간법을 이용한 영상변형

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_5 \\ a_6 \end{pmatrix} \quad (1)$$

그림 2은 어파인 동작모델을 보여주고 있다. 이 모델은 카메라의 동작인 확대/축소, 회전, 이동을 동시에 기술할 수 있는 모델이다.

동작 모델이 설정된 뒤에는 입력 영상에 대한 동작 파라미터를 구하게 된다. 기존의 연구에서는 카메라 동작 파라미터 추출 문제를 두 입력 영상간의 밝기 값의 차의 제곱의 합을 최소화시키는 최적화 문제로 해석하고 있다[1][2][3]. 그러나 이 방법은

입력영상이 정렬된 후의 다음 단계는 정렬된 입력영상들을 하나의 파노라마 영상으로 통합하는 것이다. 기존의 방법에서는 공간적인 보간을 수행하여 영상을 정렬하고, 정렬된 영상들을 결합 때 시간적인 미디언(median) 필터링을 수행한다. 그러나 이 방법은 입력영상의 동적인 밝기값의 변화를 수용하지 못하는 문제점을 안고 있다. 본 논문에서는 이 문제점을 보완하기 위해서 시간에 따라서 가중치를 부여해 정렬된 입력영상을 통합해 파노라마 영상을 획득하는 방법을 제안한다.

이 방법은 최근 영상의 정보를 더 많이 반영하기 위한 방법으로 영상의 시점이 최근일수록 가중치를 더 많이 부여하는 방법이다.

N개의 입력영상 $I_1, I_2, I_3, \dots, I_N$ 에 대해 식 (2)와 같이 영상의

시점이 최근일수록 시간에 따른 가중치를 더 많이 부여해서 파노라마의 밝기값을 결정한다.

$$I_P = \sum_{i=1}^N (W_i \times I_i), \quad W_i = \frac{t}{\sum_{k=1}^N k} \quad (2)$$

식(2)에서 I_P 는 파노라마 영상의 밝기값을 나타내고 W_i 는 i 시점의 가중치, I_i 는 i 시점 영상의 밝기값을 나타낸다.

그림 5는 시간적으로 증가하는 가중치를 고려한 영상 통합 방법을 보여주는 그림이다. 정렬된 각 시점의 영상에서 기준 좌표계에 대해서 같은 위치에 해당하는 화소의 밝기값에, 시간에 따른 가중치를 곱한 합을 파노라마 영상의 밝기값으로 결정한다.

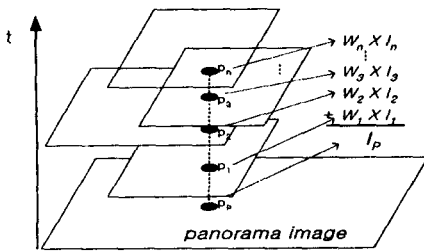


그림 5. 시간적 증가 가중치에 따른 영상 통합

4. 실험 및 결론

본 논문에서는 그림 6과 같이 카메라가 왼쪽으로 패닝(pan left)하는 동작이 포함된 실험 영상을 사용해서 파노라마 영상을 구성했다. 그리고 그림 7은 카메라의 동작 파라미터를 추출하기 위해 사용한 블록 정합 알고리즘을 통해 추출한 동작 벡터 및 필터링된 동작 벡터를 보여주고 있다. 본 논문에서는 블록 정합 알고리즘을 이용해서 추출한 동작 벡터에서 잡음을 제거하기 위해 동작 벡터를 필터링하여 사용한다.

그림 8은 입력 영상으로부터 본 논문에서 제안한 방법을 사용해서 구성한 파노라마 영상이다.

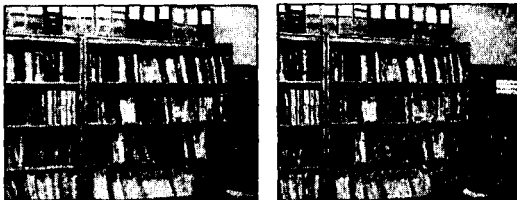


그림 6. 입력 영상 $t, t+\Delta t$

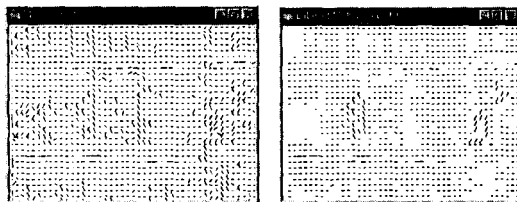


그림 7. 동작 벡터와 필터링된 동작 벡터

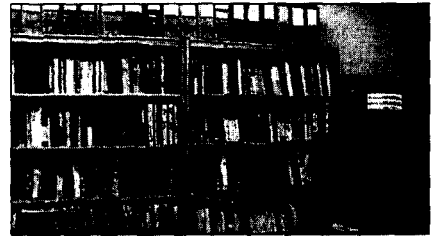


그림 8. 실험 영상으로부터 구성한 파노라마 영상

5. 결론

본 논문에서는 비디오 데이터를 표현하는 방법 중 원래의 입력 영상에 비해 보다 넓은 뷰를 제공해 영상검색에 유리하며, 입력 영상을 시공간적으로 통합하기 때문에 영상을 표현하는데 필요한 데이터의 양을 상당히 줄일 수 있는 파노라마 영상구성 방법에 대해서 설명했다.

기존의 파노라마 구성 방법은 카메라 동작을 보정하는 데 필요한 동작파라미터를 추출하는데 시간 복잡도가 너무 크고 지역적 최소화에 빠질 수 있다는 문제점이 있다. 이것은 파라미터를 추출하는데 사용되는 최적화 기법에 필요한 계산량이 많기 때문인데 본 논문에서는 이 문제점을 해결하기 위해서 블록정합알고리즘을 사용해 속도개선을 했다.

본 논문에서 주목할 점은 파노라마 영상을 구성하는데 있어서 중요한 과정중에 하나인 영상통합과정이다. 기존의 연구에서는 정렬된 입력영상을 간단히 미디언 필터링하는데 반해서 본 논문에서 제안한 방법은 입력영상의 최근정보를 반영할 수 있도록 시간에 따른 증가 가중치를 부여하는 방법이다. 이 방법은 입력영상의 시간에 따른 동적인 밝기값의 변화를 고려하는 방법으로 현재 시점과 가까운 영상일수록 더 많은 가중치를 부여해 최근정보를 잘 반영할 수 있도록 한다.

참고문헌

- [1] M. Irani, D. Anandan and S. Hsu, "Mosaic Based Representation of Video Sequences and Their Applications", In *Proc. of International Conference on Computer Vision*, pp.605-611, 1995
- [2] Frederic Dufaux, Fabrice Moscheni, "Background mosaicking for low bit rate video coding", *International Conference On Image Processing*, pp.673-676, 1996
- [3] Sunghoon Choi, Yongduck Seo, Hyunwoo Kim, "Soccer Game Analysis using Mosaicking and Color-based Tracking", *International Workshop on Computer Vision Technology*, pp.79-84, 1996.
- [4] Michal Irani, Steve Hsu, P. Anandan, "Video compression using mosaic representations", *Signal processing: Image Communication* 7, pp.529-552, 1995.
- [5] 김동근, 황치정, "활성 모델에 근거한 경계선 분할 및 모델 적합", *한국정보과학회 논문지 (B)*, Vol. 23, No. 2, February, 1996.