

그래프에 기반한 전역적 정합 방법

김 현 우, 홍 기 상

포항공과대학교 전자전기공학과

Graph-Based Framework for Global Registration

Hyunwoo Kim, Ki-Sang Hong

Electrical and Computer Engineering Divison, POSTECH

{pccc,hongks}@postech.ac.kr

요 약

In this paper, we present a robust global registration algorithm for multi-frame image mosaics. When we perform a pair-wise registration recovering a projective transformation between two consecutive frames, severe mis-registration among multiple frames, which are not consecutive, can be detected. It is because the concatenation of those pair-wise transformations leads to global alignment errors. To overcome those mis-registrations, we propose a new algorithm using multiple frames for constructing image mosaics. We use a graph to represent the temporal and spatial connectivity and show that global registration can be obtained through the search for an optimal path in the constructed graph. The definition of an adequate objective function characterizing the global registration provides a direct manipulation of the graph. In the presence of moving objects, especially large ones compared with low texture backgrounds, by using the likelihood ratio as the objective function, we can deal with some of the most challenging videos like basketball or soccer. Moreover, the algorithm can be parallelized so it can be more efficiently implemented. Finally, we give some experimental results from real videos.

1. 서 론

영상 모자이크(image mosaic)을 자동으로 구현하는 기법은 비디오 해석과 장면 표현(visual scene representation)에 있어서 중요한 요소들 중의 하나이다. 이러한 분야의 대부분의 일들은 두 영상간의 프로젝티브 변환(projective transformation)을 예측하는데 주력하여왔다. 이러한 기법들은 인접 영상 정합(pair-wise local registration)이라고 부른다. 여러 장의 영상열로부터 하나의 영상 모자이크를 만드는 경우, 인접 영상간의 정합이 정확하게 예측되더라도 여러 장의 영상을 정합할 경우에 나타나는 전역적 정합 오차(global registration error)를 피할 수 없다. 인접 영상 변환들을 여러 번 거쳐서 정합을 시켜나갈 경우, 인접 영상간에는 작았던 오차들이 누적되어 전역적 정합 오차

가 나타나게 되므로 이에 대한 해결을 위해 전체 영상을 한꺼번에 정합하여 주는 전역적 정합 방법(global registration)이 필요로 되어진다. 두 가지 방법의 차이가 그림 1.에 나타나 있다. 이러한 현상은 영상 프레임이 시간상으로는 멀리 떨어져 있지만 공간상으로는 인접해 있어 중첩(overlap)이 생기는 경우에 두드러진다.



인접 영상 정합 방법

전역적 정합 방법

그림 1. 전역적 정합의 필요성

이러한 전역적 정합 방법은 90년대 후반기에 연구되기 시작했는데 인접 영상간의 정합에서 얻어진 프로젝티브 변환들로 구성된 선형 시스템의 해를 구하는 방법 [1]과 프레임을 모자이크에 정렬하는 방법 [2]이 제안 되어졌었다. 그러나, 이러한 접근 방법들의 주요한 단점은 인접 영상간의 정합이 하나라도 정확하지 않으면 나머지의 영상간의 정합에서 영향을 미쳐서 왜곡된 영상 모자이크를 만들게 된다는 문제점이 있다. 최근에는 특징점을 이용한 기법들(feature-based techniques)이 소개되어졌다 [3][4]. 이 방법들은 제곱오차를 최소화하는 개념에서 대응되는 특징점들의 삼차원에서의 광선 방향들(ray directions)을 최소화함으로써 전체 영상들에 대하여 정합을 할 수 있다. 비록 만족할 만한 결과를 보이지만 이 기법들은 일반적인 영상 모자이크보다는 파노라믹 영상 모자이크(panoramic image mosaic)을 구성하는데 중점을 두고 있다.

결론적으로 (1) 영상 모자이크이 새롭게 얻어지는 프레임들에 의해서 갱신되어야 하며 (2) 기준이 되는 프레임도 카메라의 움직임이 모자이크에서 나타나는 폐적의 이차원 위상기하(topology)에 적응하기 위해서 변할 수 있는 전역적 영상 정합 방법이 필요로 되어진다. 이러한 위상기하의 고려는 카메라의 체적(camera swath)이 S자와 같이 지그재그로 중첩이 될 경우에는 특히 중요하다. 이 경우, 영상 프레임이 시간상으로는 멀리 떨어져 있지만 공간상으로는 인접해 있어 중첩(overlap)이 생기는 현상이

빈번하게 발생하기 때문이다. 최근에 카메라 궤적의 위상 기하를 그래프를 사용하여 표현하고자 하는 노력이 되어 졌다 [5][6][7]. 특히, 논문 [7]에서는 프레임들간의 연결성(connectivity)과 대응점들(correspondences)을 한 번에 표현하는 통일된 틀(unified framework)을 제공하였다. 이는 (a) 명시적인 이차원 위상기하 표현(explicit 2D topology), (b) 직접적인 영상 평가를 통한 오차 측정(error measurement by direct image evaluation), 그리고 (c) 대응점에 대한 효율적인 탐색(the inexpensive search of correspondences)을 제공하고 있다. 하지만, 이론적인 개념 소개에 그치고 있으며 많은 부분들이 손으로 이루어져서 능구 경이나 축구 경이 같은 도전적인 영상열을 다룰 수가 없었으며 논문 [6]에 비해서 공헌도가 별로 없었다. 본 논문에서는 앞의 논문에서 제시되었던 통합적인 틀을 일반적인 어려운 영상열들을 해결하는 입장에서 개선하여, (a) 카메라의 평행이동(translation)에 의해 나타나는 평면 패러랙스(plane parallax)가 있는 경우 [8], (b) 정합시키려는 배경에 비해 움직이는 물체의 크기가 큰 경우, (c) 영상열에 줌 인과 줌 아웃이 있을 경우들에 적용되어지는 통합적인 강력한 전역적 정합 방법을 소개하고자 한다.

2. 접근 방법

영상 모자이크에서 프레임간의 시공간적인 연결성(temporal and spatial connectivity)을 표현하기에 적합한 그래프를 사용한다. 이 그래프들은 분할된 영상 모자이크의 격자점들에서 구성이 되며, 노드들(nodes)은 격자와 중첩되는 프레임의 마스크 영상이고, 아크들(edge costs)은 마스크 영상간의 매칭을 측정하기 위한 목적함수(object function)의 값들로 연결되어 있다. 이 프레임들로부터 구성된 그래프에서 그래프 알고리즘을 사용하여 최소거리(shortest path)에 해당하는 연결들을 찾아내어 대응점을 찾으므로 모든 프레임을 사용하여 전역적으로 정합을 하게 된다. 움직이는 물체가 있는 경우에는 위의 목적함수를 가능비(the likelihood ratio)로 정의한 후 그래프의 아크들의 연결성을 조정하여 같은 틀 안에서 구현되어질 수 있다.

3. 알고리즘

본 알고리즘은 세 부분으로 구성되어진다. 인접 영상간의 정합을 하고 영상 모자이크를 구성하는 인접 영상 정합을 위한 단계, 격자로 나누어진 영상 모자이크 위에서 격자별로 중첩되는 프레임에서 그래프를 구성한 후 대응점을 구해주는 전역적 영상 정합을 위한 단계, 마지막으로 새로 정의되어지는 모자이크를 위한 평면과 프레임간의 프로젝티브 변환을 조정하여 개선된 영상 모자이크를 구하는 강력한 영상 모자이크 구성을 위한 단계이다. 이에 대한 전체 흐름도가 그림 2에서 보여지고 있고 다음에서 단계별로 설명되어진다.

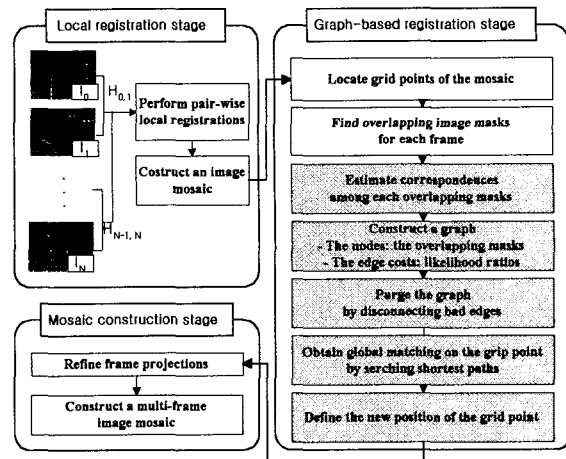


그림 2. 전체 흐름도

인접 영상 정합

영상열 $S = \{I_0, \dots, I_K\}$ 이 평행이동을 포함한 임의의 카메라에 의해 취득 되어졌다고 하자. 각각의 인접한 영상들 I_{i-1} 와 I_i 사이의 프로젝티브 변환 $H_{i-1,i}$ 을 구하게 된다. 이 과정에서 사용되어진 계산방법은 다음과 같다. 각각의 영상에 대하여 가우시안 피라미드를 구성하고 가장 해상도가 낮은 영상에 대하여 주파수 영역(frequency domain)에서의 위상차를 이용하여 영상간의 평행이동(image translation)을 구해낸다. 이를 초기치로 하여 다음으로 높은 해상도의 영상 피라미드들에 대해서는 다층적으로(hierarchically) 특징점들을 추출하고 특징점간의 상관관계(correlation)가 가장 높은 것들을 매칭시켜 대응점들을 구한다. 이 대응관계는 선형시스템으로 표현되는 데 반복적인 최소제곱방법을 사용하여 프로젝티브 변환 $H_{i-1,i}$ 을 구할 수 있다 [9]. 모자이크를 구성하기 위해서 하나의 기준 프레임에 대한 변환 관계를 계산하게 되는 데, 기준 프레임을 I_{ref} 라고 할 때 임의의 프레임 I_i 로부터의 변환 P_i 은 프레임 ref와 i사이의 인접 영상간의 프로젝티브 변환들을 연속 연결(concatenation)하여 얻어낼 수 있다.

전역적 영상 정합

앞에서 설명한대로 인접 영상간의 정합만으로 구성된 영상 모자이크는 정합오차가 커서 영상 해석이나 장면표현에 사용되어질 수 없다. 따라서, 전역적 영상 정합이 필요로 되어진다. 먼저 앞 단계에서 구해진 영상 모자이크 위에 균일한 격자점들(grid points) $L = \{L_1, \dots, L_N\}$ 을 분포시킨다. 이 격자점들을 특징점들로 사용하여 대응점을 구할 것이다. 각각의 격자점 L_i 에 대하여 해당하는 프레임들과 그 프레임에서 해당되는 격자점들의 위치를 구하고, 각각의 프레임에서 해당위치 주변의 일정크기의 원도우에 해당하는 영상 영역들 $M = \{M_1, \dots, M_N\}$ 을 잡아

이들을 중첩 영상 마스크들(overlapping image masks)이라고 부른다. 각각의 격자점 위에서 이 마스크들 사이의 영상측정을 통하여 대응관계를 구해내고자 한다.

우선 위의 중첩관계를 나타내는 마스크들을 노드들로 하여 그래프를 구성한다. 임의의 한 노드는 아크들을 사용하여 나머지 모든 노드들에 연결되어있다. 이 아크들에 정보들을 넣어두고 그래프 알고리즘들을 효율적으로 사용하여 전역적 정합을 할 수 있는 것이다. 각각의 아크는 임의의 두 노드들(마스크들)을 연결해 주는데, 두 마스크 간의 상관관계 (correlation)를 사용하여 계산된 매칭과 인접 영상 정합에 의해 구해지는 매칭과의 차이값과 움직이는 물체가 있는 지를 측정해 주는 목적함수인 가능비 (likelihood ratio) [10]가 저장되게 되어진다. (이 가능비 목적함수는 차후에 설명되어진다.)

다음으로 움직이는 물체가 들어있는 확률이 크거나 정합이 되지 않았을 프레임들간의 관계를 그래프에서 없애주기 위해서 아크들의 가능비 목적함수들의 중앙값 (median)을 기준으로 값이 큰 아크들을 제거하여준다. 이렇게 가지치기가 된 그래프에 Dijkstra 의 알고리즘을 적용하면, 하나의 기준 프레임(마스크)으로부터 나머지 프레임들로의 최단거리를 구하게 되어, 이 때의 대응관계가 전체의 중첩된 영상들을 모두 고려한 최적의 정합관계가 된다 [11]. 기준프레임을 잡는 것은 시간적으로 가장 빠른 영상을 잡아주던지 인접 영상을 구성할 때 잡았던 기준 프레임과 가장 시간적으로 근접 영상을 사용하였다. 마지막으로, 수정된 대응점들을 사용하여 현재의 격자점의 위치를 조정하여 새로운 격자점을 모자이크위에서 구하게 되는데 이는 새로운 대응점들을 모두 모자이크위에 재투영(reprojection)을 시킨후 이 투영된 점들의 무게중심을 구함으로 이루어진다.

강건한 영상 모자이크 구성

위의 단계들을 다 거치면 새로운 모자이크 평면과 각각의 프레임에서 그에 대응하는 격자점들의 위치를 가지게 되고 이 관계를 사용하여 인접 영상 정합으로부터 구해진 프로젝티브 변환들 $P = \{P_1, \dots, P_K\}$ 을 조정하면 정밀한 영상 모자이크를 구성할 수 있게 된다. 이 때 잘못된 매칭들을 없애주고 강건한 모자이크 구성을 위해서 최소제곱중앙값 기법(least median squares method)을 사용하여 파라미터들을 갱신해 준다.

4. 움직임 검출을 위한 목적 함수

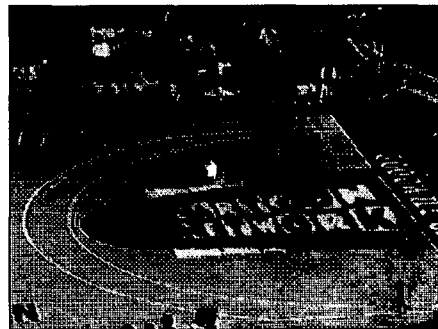
배경에 대해서 매칭되어지지 않고 움직이는 물체에 의해서 매칭되어지는 경우를 그래프에서 없애주기 위해서는 이에 맞는 목적함수를 사용하여야 한다. 이를 위해서 움직임 검출에 사용되는 측정치가 가능비(likelihood ratio)라는 것이다[10]. 두 프레임에서 마스크들 안의 계조값의 평균이 각각 μ_1, μ_2 이고 분산이 σ_1, σ_2 일때, 가능비는 다음과 같이 정의되어진다.

$$\lambda = \frac{\left[\frac{\sigma_1 + \sigma_2}{2} + \left(\frac{\mu_1 - \mu_2}{2} \right)^2 \right]^2}{\sigma_1 \times \sigma_2}$$

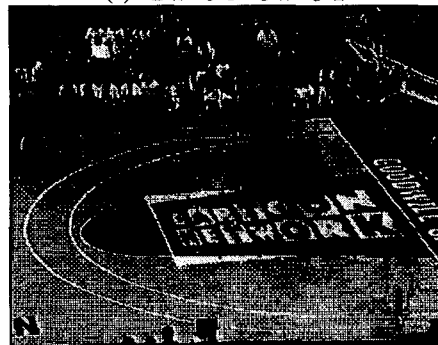
값이 1 이 될 때 계조값의 분포가 동일한 영상이 되는 것이고, 값이 커질수록 움직이는 물체가 들어갈 확률이 높아지는 것이다.

5. 실험 결과

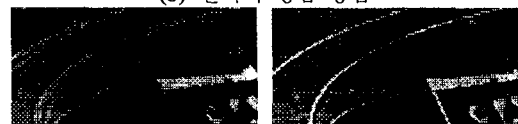
실험에 사용되어지는 영상은 농구, 축구, 테니스 경기에서 얻어졌으며 프레임의 수는 각각 25, 30, 25 장으로 구성 되어있다. 농구의 경우는 움직이는 선수들과 관중들이 영상에서 차지하는 부분이 크고 경기를 따라가기 위해서 카메라의 움직임이 크기 때문에 정합하는 것이 쉽지않은 것으로 알려져 있다. 축구의 경우는 경기장에 기하학적이 특징점들이 없고 선수들이 그 역할을 해서 부정합을 만들어 내며 테니스의 경우는 마지막 영상열에서 좁아터지는 부분이 일반적인 정합을 어렵게 한다. 그에 대한 결과들이 그림 3-5 에서 보여진다. 또한, 그림 1 에서는 plane parallax 가 있는 경우에 대한 정합을 보여주고 있다. 인접 영상간의 정합의 경우 정합 오차로 인한 고스트 현상(ghost-like blending)를 볼 수가 있으며 전역적 방법에서는 이러한 오차가 개선되었음을 확연하게 비교하여 볼 수 있다.



(a) 인접 영상 정합 방법

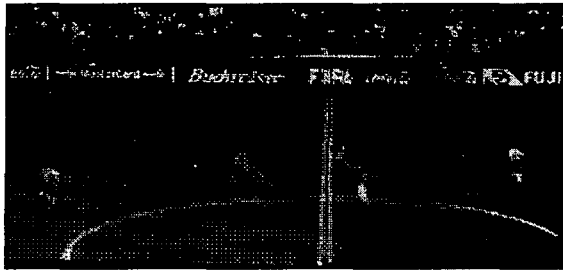


(b) 전역적 정합 방법

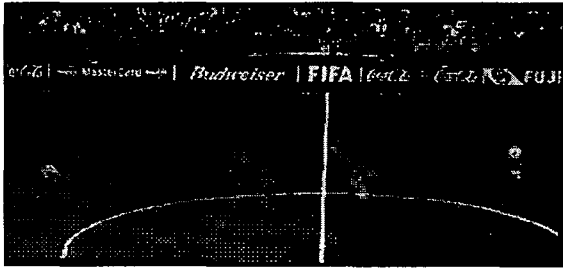


(a)와 (b) 모자이크의 줌인

그림 3. 농구 경기의 영상 모자이크

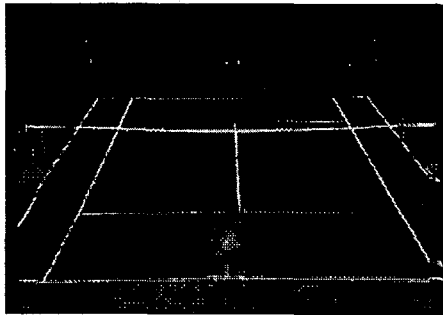


(a) 인접 영상 정합 방법

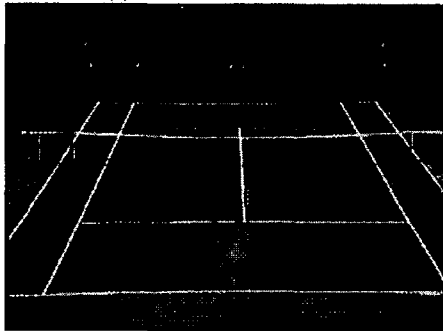


(b) 전역적 정합 방법

그림 4. 축구 경기의 영상 모자이크



(a) 인접 영상 정합 방법



(b) 전역적 정합 방법

그림 5. 테니스 경기의 영상 모자이크

6. 결론

본 논문에서는 그래프에 기반한 전역적 정합방법에 대해 제안하였고, 이는 인접 영상 정합 방법의 전역 정합 오차를 개선하여 주었다. 또한, 기존에 다루기 어려웠던 도전적인 농구 경기와 같은 영상열에 대해서도 정확한

정합 결과와 고해상도의 영상 모자이크를 얻는 것이 가능하게 하였고 이는 실제의 영상열에 대한 실험 결과를 통해 보여졌다.

이러한 통일된 틀 안에서 영상열의 모션에 의한 흐릿함을 개선할 수 있고 영상열에서 없는 정보를 추출하는 고해상도(super resolution)의 모자이크의 구현이 가능하다. 이를 위한 연구가 현재 진행중이며, 또한 움직이는 물체들의 추적알고리즘이 그래프의 정보를 사용하여 쉽게 추출되어질 수 있을 것으로 여겨진다.

7. 참고 문헌

- [1] Davis, J.E.[James Edward], Mosaics of Scenes with Moving Objects, CVPR98(354-360).
- [2] Sawhney, H.S., Kumar, R., Gendel, G., Bergen, J., Dixon, D., Paragano, V., VideoBrush: Experiences with Consumer Video Mosaicing, WACV98(Session 2A).
- [3] Capel, D.[David], Zisserman, A.[Andrew], Automated Mosaicing with Super-resolution Zoom, CVPR98(885-891).
- [4] Shum, H.Y.[Heung-Yeung], Szeliski, R.[Richard], Construction and Refinement of Panoramic Mosaics with Global and Local Alignment, ICCV98(953-958).
- [5] Hsu, S.C., Sawhney, H.S., Influence of Global Constraints and Lens Distortion on Pose and Appearance Recovery from a Purely Rotating Camera, WACV98(Session 4B).
- [6] Sawhney, H.S., Hsu, S., Kumar, R., Robust Video Mosaicing Through Topology Inference and Local to Global Alignment, ECCV98(xx-yy).
- [7] Kang E.Y., Cohen I., Medioni G., A Graph-based Global Registration for 2D Mosaics, ICPR2000.
- [8] Irani, M., Anandan, P., A Unified Approach to Moving Object Detection in 2D and 3D Scenes, PAMI(20), No. 6, June 1998, pp. 577-589.
- [9] Hartley, R.I.[Richard I.], and Sturm, P.[Peter], Triangulation, CVIU(68), No. 2, November 1997, pp. 146-157.
- [10] Jain, R.C., Kasturi, R., and Schunck, B.G., Machine Vision, New York: McGraw-Hill1995.
- [11] Cormen, T.H., Leiserson C.E., and Rivest, R.L., Introduction to Algorithms, MIT Press, 1992.