

실시간 파노라마 합성에서의 효과적인 outlier 제거 방법

*김범수 **조남익

서울대학교

*bskim@ispl.snu.ac.kr **nicho@snu.ac.kr

Efficient outlier removal algorithm for real-time panoramic stitching

*Kim, Beom Su **Cho, Nam Ik

School of Electrical Engineering and Computer Science, Seoul National University

요약

기존의 실시간 파노라마 합성 알고리즘에서는 매칭점과 입력 영상에서의 outlier를 구분하고 제거하기가 어렵기 때문에 노이즈가 많은 영상 또는 반복적인 패턴이 많은 영상에서 왜곡이 쉽게 발생하는 문제가 있다. 따라서 본 논문에서는 기존의 실시간 파노라마 합성 프레임워크에서 실시간 합성 조건을 만족시키면서 효과적으로 매칭점과 입력 영상에서의 outlier를 제거하는 방법을 제안한다. 이를 위해서 선형 모델에서 outlier를 제거하는 데 주로 사용되는 RANSAC 알고리즘을 실시간 파노라마 합성에서 사용되는 비선형 모델에 적용 가능하도록 수정하고 속도 향상을 위해서 사용되는 모델의 파라미터를 줄이는 방법을 제안한다. 이를 통하여 매칭점 중에 존재하는 outlier를 제거하고 전체 매칭점 중에서 inlier 비율을 이용하여 입력되는 영상 시퀀스에서 outlier 영상을 제거하는 방법을 제안한다. 실험 결과 기존의 방법에 비해서 합성 결과의 왜곡이 줄어드는 것을 확인하였다.

1. 서론

파노라마 합성은 컴퓨터 비전의 주요 연구 주제 중 하나로서 특징 점 추출과 매칭, 영상의 투영, 최적화 기법 등 다양한 기술을 요구한다. 디지털 카메라가 발명되기 전부터 파노라마 합성은 다양한 형태로 존재하였으며 현재에는 디지털 카메라와 PC의 보급으로 일반 사용자들도 범용적으로 이용하고 있다.[1] 초기의 파노라마 합성은 제한적인 모션 모델만을 이용하다가 점차 projective 모션 모델로 확장 적용되었으며, 픽셀의 밝기 차를 가장 작게 만드는 모션의 파라미터를 추정하는 방식에서 영상 간의 밝기 차이에 강인한 특징점 추출 및 매칭을 이용하는 방식으로 발전하였다. [1, 6, 7] 이 외에도 수백 또는 수천 프레임의 동영상상을 가지고 파노라마 합성을 수행하는 연구도 진행되었다. [8] 최근에는 스마트폰의 보급으로 모바일 기기에서 PC에 버금가는 처리가 가능하게 되었고 함께 장착된 카메라를 이용하여 실시간으로 파노라마 합성을 가능하게 하는 연구가 진행되었다. [4]

최근에 제안된 실시간 파노라마 합성은 모바일 기기에서 프리뷰 영상의 형태로 입력되는 비디오를 처리하여 실시간으로 파노라마 합성을 수행하며, 기존의 일반적인 파노라마 합성 방법을 간소화시켜 실시간 처리를 가능하게 하였다. [4] 특히 시간이 많이 걸리는 특징점 추출과 매칭 대신에, 트래킹에 기반하여 매칭점을 찾고 이를 통하여 합성에 걸리는 시간을 대폭 줄이게 된다. 하지만 일반적인 파노라마 합성과 달리 프레임 간의 매칭점을 구하는 대신, 원기둥 투영면에 이미 투영된 합성 결과와 현재 입력 영상 간의 매칭점을 구하는 방식을 이용하기 때문에 일반적으로 파노라마 합성에서 사용되는 outlier 제거 방법을 적용하기 힘든 단점이 있다. 이 때문에 노이즈가 많은 영상과 반복적인

패턴이 많은 건물이나 숲 영상의 경우 outlier로 인하여 결과가 쉽게 왜곡되게 된다. 뿐만 아니라 모션 블러와 같은 노이즈에 오염되거나 급격한 카메라 이동으로 트래킹이 실패하게 되는 프레임이 입력될 때, 이를 프레임 outlier로 판단하고 다음 영상을 기다려야 하지만 이 같은 추정이 쉽지 않은 문제점도 존재한다.

이에 본 논문에서는 기존의 실시간 파노라마 합성 방법에서 실시간 조건을 만족시키면서 효과적으로 outlier를 제거하는 방법을 제안한다. Outlier를 제거하는데 주로 사용되는 RANSAC (RANDOM SAmple Consensus) 알고리즘은 사용되는 모델이 선형적인 경우 효율적이지만, 실시간 파노라마와 같이 투영된 결과 영상과 현재 카메라 영상 사이의 모델이 비선형적인 관계인 경우 수행시간이 크게 늘어나는 단점이 있다. 이를 해결하기 위해 모델의 비선형성을 제거하여 효율적으로 RANSAC을 적용 가능하도록 수정하고 RANSAC의 수행시간을 줄여 실시간으로 동작 가능하도록 사용되는 모델의 파라미터를 줄이는 방법을 제안한다. 또한 전체 매칭점 중에 inlier의 비율을 이용하여 현재 입력 영상의 inlier/outlier 여부를 판단하는 방법을 제안한다. 실험 결과, 제안하는 방법은 실시간 조건에 위배되지 않으면서 기존의 방법에 비해 효과적으로 outlier를 제거하여 합성 결과의 왜곡이 줄어드는 것을 확인하였다.

2. 기존의 연구

가. 프리뷰 형태의 실시간 파노라마 합성

파노라마 영상을 합성할 때, 일반적으로 한 장소에서 찍은 여러 장의 영상을 이용한다. 이때, 영상을 촬영하는 순간에는 파노라마 합성

결과를 예측하기 힘들기 때문에 사용자가 원하는 사각형 형태의 파노라마 결과를 얻지 못하는 경우가 많다. 이를 해결하기 위하여 파노라마 합성 결과를 프리뷰 형태로 실시간 처리하여 보여주는 방식이 연구되었다.[2] 이 연구는 정확한 합성 결과보다는, 파노라마 합성 결과를 유저가 미리 대략적으로 예측할 수 있도록 가이드라인의 역할에 초점이 맞춰져 있다. 카메라에 계속 입력되는 영상 시퀀스를 빠른 속도로 처리하여 파노라마 프리뷰 영상을 만들기 위해서 적은 수의 키프레임을 설정하고 키프레임과 현재 들어오는 프레임 사이에 특징점 매칭을 통하여 카메라의 위치를 계산하는 방법을 이용한다.[3] 이때 특징점 매칭 과정에서 RANSAC을 이용하여 효과적으로 outlier를 제거하고 비교적 정확하게 카메라의 위치를 계산할 수 있다.[5] 키프레임을 이용하여 전체 영상이 아닌 적은 수의 영상과 현재 입력 영상의 특징점을 비교하기 때문에 속도의 향상이 가능하고 실제로 실시간 처리를 할 수 있게 되었다. 하지만 모든 프레임에서 특징점을 추출하고 이를 다시 매칭하는 과정에서 오버헤드가 크기 때문에, PC에서 초당 1프레임 정도의 속도로 동작하며 완전한 실시간 파노라마 합성과는 거리가 있다.

나. 트래킹에 기반한 실시간 파노라마 합성

매 영상마다 특징점을 추출하고 이 특징점 간의 매칭을 계산하는 부분은 시간 소모가 크기 때문에 실시간 파노라마 합성의 걸림돌이 되었다. 이 부분의 속도를 향상시키기 위하여 매 영상마다 추출한 특징점 간의 매칭이 아닌 트래킹에 기반한 파노라마 합성 방법이 제안되었다.[4] 이 방법은 입력된 영상을 원기둥 형태의 투영면에 투영한 후 이 투영 결과에서만 특징점 추출을 수행하여 현재 입력 영상의 위치를 파악하는 맵으로 활용한다. 시간이 많이 걸리는 특징점 추출은 맵이 처음 생성되는 순간에 한번만 수행된다. 이렇게 추출된 특징점을 가지고 템플릿 매칭을 사용하여 현재 프레임의 어떤 점과 매칭이 되는지 대응점을 찾게 된다. 이 방법은 비디오 코딩에서 사용되는 motion estimation 기법과 유사하며, motion estimation에서는 블록의 모션을 찾는 데 반해 여기서는 미리 추출해놓은 맵의 특징점의 모션을 찾는다는 것이 차이점이다. 초당 수 프레임 이상의 영상이 입력된다고 할 때, 인접 영상 간에 시간적인 유사도 덕분에 모션이 범위가 크지 않게 되고 따라서 일반적인 파노라마 합성에서 사용되는 특징점 매칭 대신에 상대적으로 간략한 motion estimation 기법을 활용하여 수행속도를 높게 된다. 이 방법을 사용하여 모바일 기기에서 실제로 초당 30프레임의 속도로 파노라마 합성을 처리할 수 있다.[4] 하지만 트래킹에 기반하기 때문에 outlier를 제거하는데 한계가 있다. 특히 입력 영상과 맵 사이의 관계는 원기둥 투영이라는 비선형 모델로 정의되기 때문에 RANSAC을 적용하는데 시간 소모가 커지게 된다. 그 이유는 모델의 파라미터를 구하기 위하여 반복 알고리즘(iterative non-linear optimization)에 기반한 복잡한 최적화 기법을 RANSAC의 매 단계마다 반복하여 사용해야 하기 때문이다.

3. 제안하는 방법

가. 매칭점 사이에 존재하는 비선형 관계 제거

제안하는 방법은 트래킹을 사용하여 찾아진 매칭점들 중에서 outlier를 제거하는데 기본적으로 RANSAC을 이용한다. 하지만 위에

서 언급한 문제를 해결하기 위하여 RANSAC에서 사용할 모델을 간소화할 필요가 있다. 기존의 모델은 원기둥 투영 함수이며, 모델에서 쓰이는 파라미터는 4가지로 3차원 회전에 대한 파라미터 θ_i , ($i = 1, 2, 3$)와 카메라의 내부 파라미터인 초점 거리 f 이다. [6, 7] 여기서 f 는 미리 고정시켜 놓고 찾아야할 모델 파라미터는 θ_i 이다. 원기둥 투영 함수를 수식으로 표현하면 다음과 같다.

$$u = R\theta_c = R \tan^{-1}(X/Z), v = Y/\sqrt{X^2+Z^2}$$

여기서 R 은 원기둥의 반지름이고, θ_c 는 원기둥에서 호의 각도이다. X, Y, Z 는 현재 영상의 점을 월드 좌표계로 표현한 것이며 이를 계산하기 위해서는 θ_i 와 f 를 이용하여 구한 카메라 행렬의 역행렬을 이용하여 구한다.[5] 위 수식과 같이 현재 영상과 원기둥 투영을 이용하여 생성한 맵의 관계는 비선형이므로 RANSAC을 이용하여 파라미터를 계산하는 데 시간이 많이 걸리는 문제가 있다. 이를 해결하기 위해 입력 영상과 맵 사이의 비선형성을 제거한 후 RANSAC을 적용한다. 제안하는 방법은 맵 상의 특징점을 다시 카메라 평면으로 역투영함으로써 두 매칭점이 같은 평면에 위치하도록 하여 비선형성을 제거한다. 하지만 맵을 카메라 평면에 역투영하기 위해서는 어떤 회전 파라미터를 사용할지 결정해야 하는 문제가 있다. 현재의 입력 영상에서 사용된 회전 파라미터를 사용할 때 두 매칭점이 정확히 같은 평면상에 위치하게 되어 효과적으로 RANSAC을 수행할 수 있지만 그 값을 모르기 때문에, 이미 계산된 이전 프레임의 파라미터에 constant velocity 가정을 적용하여 예측한 파라미터를 사용한다. 역투영된 특징점의 좌표와 현재 입력 카메라 영상 위의 대응점은 같은 카메라 평면상에 위치하기 때문에 3x3 호모그래피 행렬로 정의되는 선형 관계가 성립한다.[5] 따라서 선형 모델에 기반한 RANSAC을 사용할 수 있고 결과적으로 빠르게 파라미터를 추정할 수 있게 된다.

나. RANSAC의 간소화

호모그래피는 DOF가 8인 선형모델이기 때문에 파라미터를 추정하기 위한 최소의 특징점의 개수는 4이다. 하지만 초당 10프레임 이상을 처리하는 실시간 파노라마에서는 일반적으로 연속된 프레임 사이에 움직임의 크기가 작기 때문에 호모그래피보다 제약(cosnstraint)이 더 큰 모델을 사용하여도 충분한 결과를 얻을 수 있다. 정해진 search range 안에서 매칭점을 찾는 트래킹 알고리즘 덕분에 노이즈 크기가 크지 않고 따라서 호모그래피를 모션 모델로 사용하는 것은 오히려 오버피팅(overfitting)의 우려가 존재한다. 이를 해결하기 위해서 similarity 모델을 사용하여 RANSAC을 수행한다. 맵을 역투영할 때 현재 입력 영상에서 사용된 파라미터의 예측 값을 사용하기 때문에, 입력 영상과 역투영된 영상은 거의 같은 평면 상에 위치한다고 볼 수 있다. 따라서 일반적인 파노라마 합성과 다르게 similarity 모델로 충분히 매칭점 사이의 관계를 모델링 할 수 있다. Similarity 모델은 DOF가 4이기 때문에 2개의 매칭점으로 파라미터를 추정할 수 있으며 파라미터를 추정하는 최소 집합의 크기가 작기 때문에 호모그래피 모델보다 훨씬 노이즈에 강인하고 수렴 속도가 더 빠른 장점이 있다.

다. Outlier 영상의 판단

영상에 모션블러와 같은 노이즈가 심한 경우 또는 급격한 카메라 이동으로 기존 영상과 오버랩의 크기가 현저히 작은 경우 파라미터의 추정이 잘못될 수 있기 때문에 영상을 outlier로 판단하고 다음 프레임을 기다리게 된다. 기존의 알고리즘에서는 매칭점의 총 개수로 이러한 outlier 영상을 판단하게 된다. 하지만 outlier 영상을 판단하는 매칭점의 임계값(threshold)은 영상과 카메라에 따라 다른 값을 가져야 한다. 복잡하고 특징점이 많은 장면의 경우 outlier를 판단하는 임계값은 단조롭고 특징점이 많은 경우보다 커질 수밖에 없다. 또한 카메라 센서에 존재하는 노이즈의 양에 따라라도 임계값이 달라져야 한다. 이와 같이 영상에 따라 임계값이 달라질 수밖에 없기 때문에 정확한 임계값을 정하는 것은 어려운 문제다. 또한 매칭점은 search range안의 점 중에서 가장 유사한 점을 찾기 때문에 유사한 패턴이 반복되는 건물 영상 같은 경우 outlier 영상이지만 매칭점의 개수가 많은 경우도 존재한다. 이를 해결하기 위해 outlier 영상을 판단하는 기준을 매칭점의 개수가 아니라 총 매칭점의 개수에서 inlier의 비율로 정한다. 이 전략은 여러 장의 영상에서 불일치(inconsistent)하는 영상을 찾는 데 이미 사용된 적이 있으며 그 효과가 입증되었다. [6, 7] 정확한 확률 모델링을 위해서 i 번째 매칭점이 inlier인 확률을 independent binomial로 정의한다. Inlier 영상에서 하나의 매칭점이 inlier인 확률 p_1 을 0.9로 놓고, outlier 영상에서 하나의 매칭점이 inlier로 결정될 확률 p_0 을 0.3, 현재 입력 영상이 inlier일 확률을 0.5로 놓으면 다음과 같은 조건식을 만족할 때 입력영상을 inlier라고 판단할 수 있다.

$$n_i > 2 + 0.6n_f$$

여기서 n_f 는 모든 매칭점의 개수이고, n_i 는 이중 inlier의 개수를 의미한다. RANSAC으로 판단한 inlier의 개수가 위 식을 만족시키지 못하면 현재 입력 영상을 outlier로 판단하고 다음 영상을 기다리게 된다.

4. 실험 결과 및 분석

그림 1과 그림 2는 기존의 트래킹에 기반한 실시간 파노라마 합성 방법과 여기에 제안한 방법을 추가 적용한 결과를 비교한 것이다. 그림 1에서 보듯이 기존의 방법은 현재 입력 영상에서 매칭점의 개수가 충분히 많으면 이를 그대로 합성하기 때문에 숲이나 나무가 있는 부분에서 outlier 영상을 걸러내지 못하고 왜곡된 모습을 보인다. 반면에 새롭게 제안한 방법에서는 inlier의 비율로 outlier영상을 효과적으로 걸러내기 때문에 나무와 같이 특징점이 별로 없는 곳에서도 큰 왜곡 없이 합성된 모습을 볼 수 있다. 그림 2에서도 유사한 결과를 확인할 수 있다. 기존의 방법에서는 반복적인 패턴이 반복되는 건물 벽에서 outlier로 인하여 선이 똑바로 이어지지 못하고 끊어지는 모습을 볼 수 있다. 반면에 새로 제안한 방법은 반복적인 패턴에서도 강인하게 합성이 이루어짐을 확인할 수 있다. RANSAC에서 소요되는 시간은 2.67Ghz 인텔 i5 CPU를 사용한 PC에서 평균적으로 0.5ms 이하로 측정되었다. 따라서 제안하는 방법은 실시간 파노라마 합성에서 실시간 조건을 만족시키면서 합성 결과를 효과적으로 개선시킨다고 볼 수 있다.

5. 결론 및 향후 과제



그림 1. (위) 기존의 방법, (아래) outlier 제거 후

본 논문에서는 기존의 실시간 파노라마 합성에서 outlier를 제대로 처리하지 못하여 반복적인 패턴이 많은 건물이나 숲 등에서 결과가 왜곡되는 현상을 해결하고자 outlier를 효과적으로 제거할 수 있는 방법을 제안하였다. 실시간 파노라마 합성에서 사용된 트래킹에 기반하여 매칭점을 찾는 방식에 RANSAC을 적용하기 위해서 매칭점 사이의 비선형성을 제거하였고, 노이즈에 강인하고 속도를 향상시키기 위하여 RANSAC에 사용되는 모션 모델을 간소화 하였다. 또한 RANSAC에서 구한 inlier를 이용하여 매칭점에서 outlier 뿐만 아니라 영상 사이에서의 outlier를 효과적으로 제거하는 방법을 제안하였다. 제안된 방법의 결과를 기존의 방법과 비교하였을 때, 기존의 방법에서 왜곡이 심한 영상의 경우에도 효과적으로 왜곡을 억제하고 더 나은 합성 결과를 보여준다. 현재 실시간 파노라마 합성에서는 카메라의 초점거리(focal length)를 고정된 상태로 사용한다. 이는 트래킹에 기반한 실시간 파노라마의 한계로, 고정된 초점거리가 아닌 줌이 가능한 카메라를 이용하는 경우 추정해야 할 파라미터가 늘어나서 속도가 느려지기 때문이다. 또한 스케일 변화 때문에 트래킹 방식으로 매칭점을 찾기가 힘들어 outlier가 급격히 증가하는 단점이 있다. 향후, 제안한 방법을 확장 적용하여 스케일 변화가 있는 영상에서도 outlier를 효과적으로 제거하고

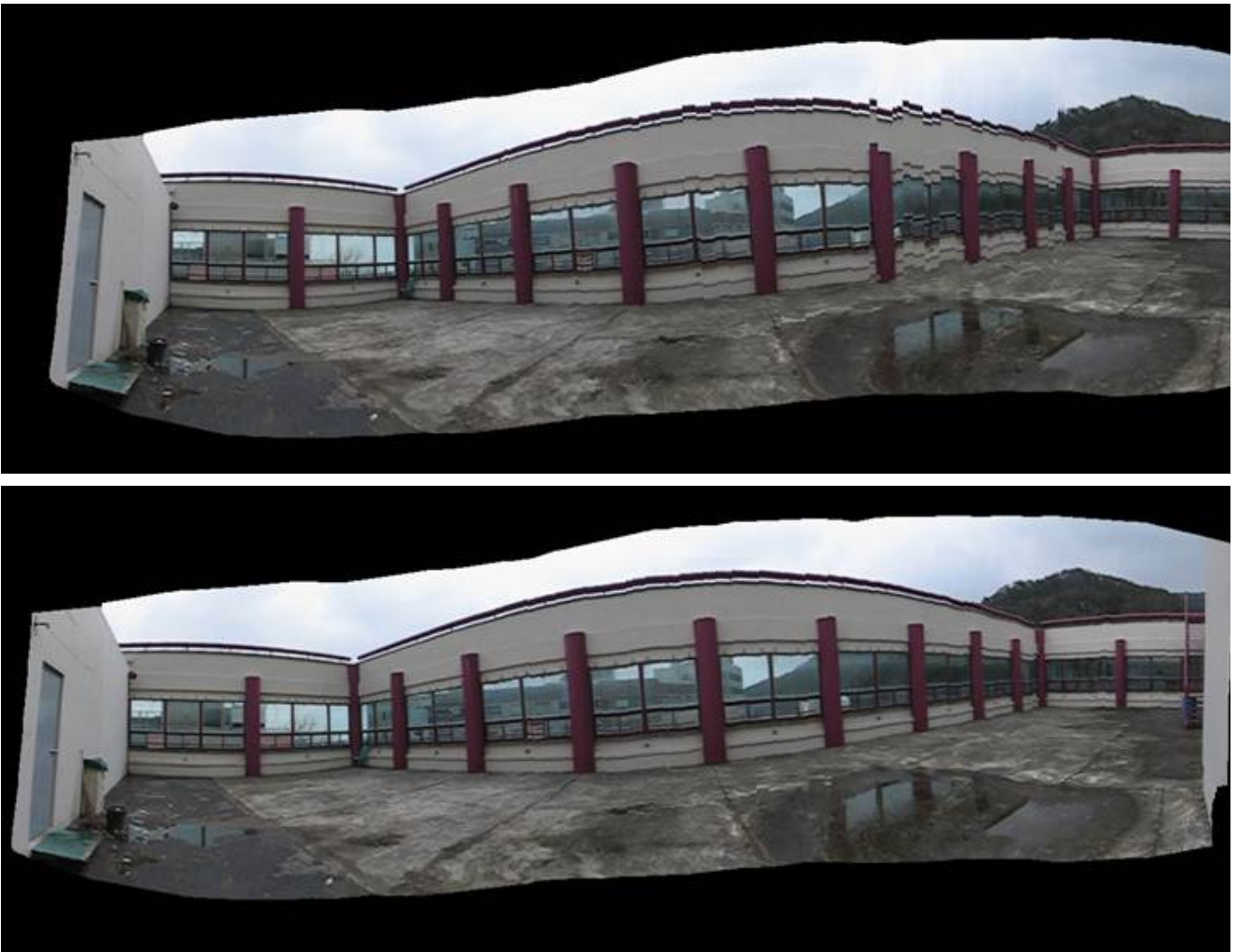


그림 2. (위) 기존의 방법 법, (아래) outlier 제거 후

이를 실시간으로 합성하는 방법을 연구하고자 한다.

감사의 글

이 논문은 2011년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(20110001271)

참 고 문 헌

[1] R. Szeliski, "Image alignment and stitching: a tutorial," *Found. Trends. Comput. Graph. Vis.*, vol. 2, no. 1, pp. 1-104, 2006. [Online]. Available: <http://dx.doi.org/10.1561/0600000009>

[2] P. Baudisch, D. Tan, D. Steedly, E. Rudolph, M. Uyttendaele, C. Pal, and R. Szeliski. "Panoramic viewfinder: providing a real-time preview to help users avoid flaws in panoramic pictures," *In Proceedings of the 17th Australia conference on Computer-Human Interaction (OZCHI '05)*, Australia, 2005.

[3] D. Steedly, C. Pal, and R. Szeliski, "Efficiently Registering Video into Panoramic Mosaics," *In Proceedings of the 10th International Conference on Computer Vision (ICCV '05)*, 2005.

[4] D. Wagner, A. Mulloni, T. Langlotz, and D. Schmalstieg, "Real-time panoramic mapping and tracking on mobile phones," *Virtual Reality Conference (VR '05)*, 2010.

[5] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[6] M. Brown and D. Lowe, "Recognising panoramas," *in Proceedings of Ninth IEEE International Conference on Computer Vision (ICCV '03)*, 2003.

[7] M. Brown and D. G. Lowe, "Automatic panoramic image stitching using invariant features," *International Journal of Computer Vision (IJCV '07)*, vol. 74, no. 1, pp. 59-73, 2007.

[8] R. Szeliski, "Video mosaics for virtual environments," *IEEE Computer Graphics and Applications*, 16(2):22-30, 1996.