

카메라 움직임 모델을 이용한 파노라마 영상 합성 Panoramic image reconstruction based on a camera motion model

이승현*, 이재호**, 이상길**, 강문기*

* 연세대학교 전기컴퓨터공학과

** 한국방송공사 기술연구소

Seung-Hyeon Rhee*, Jae Ho Lee**, Sang-Gil Lee**, and Moon Gi Kang*
Dept. of Electrical and Computer Engineering, Yonsei University
Technical Research Institute, Korean Broadcasting System

요약

디지털 영상 기기의 발전과 함께 여러 가지 실용적인 영상 처리 기법들이 연구 개발되고 있다. 화질의 선명도와 해상도가 향상되면서 더 많은 영상 정보를 한 화면에 담고자 하는 시도로서 파노라마 영상 합성이 주목을 받고 있다. 이는 영상 처리 분야에서 매우 중요한 문제로 다루어지고 있는 움직임 정보의 추출과 영상 복원 및 화질 향상 등 여러 가지 기반 기술이 요구되는 복합적인 영상 처리 기법이다. 본 논문은 특히 파노라마 영상 합성의 대상으로 적합한 축구 경기 장면을 파노라마 영상으로 재구성하는 효과기의 구현을 목표로 하였다. 별도의 장비를 활용하거나 촬영 기술상의 제한 또는 요구 없이 주어진 영상 시퀀스로부터 영상을 분석하고 합성한다는 전제하에 이루어졌으며, 카메라 움직임의 모델링, 역변환식에 의한 합성 영상의 재구성, 후처리 문제 등 시스템 전반을 다루었다.

1. 서론

축구 경기 등의 중계방송에서 카메라에 의한 시야는 전체 운동장을 담을 수 없는 한계를 갖는다. 이 문제를 해결하기 위해 각각의 부분을 촬영한 영상을 합성하여 전체적인 경기장의 모습을 재구성함으로써 시청자에게 보다 넓은 시야를 제공하는 영상 기법이 파노라마이다. 카메라를 상하좌우로 움직이면서 촬영하는 경우 각 부분 영상들을 포함하는 여러 프레임들을 조합하여 전체 영상으로 재구성하는 파노라마 영상 재구성은 각 프레임들간의 확대 배율 차이, 중심과 가장자리 부분간의 기하학적 왜곡, 그리고 움직이는 카메라에 의한 영상 번짐 현상 등을 보정, 해결해 주는 연구와 함께 프레임들을 변환하여 재구성해 주기 위한 연구가 수행되어야 한다. 또한 디지털 영상을 변환하는 경우, 정수 격자 위치를 벗어난

화소의 값들이 쓰이는 경우가 빈번하므로 이를 효과적으로 보간할 수 있는 알고리즘이 요구된다. 이와 같이 파노라마 영상 합성은 영상 신호 처리 전반에 걸친 이론과 기술이 요구되는 분야로서 디지털 영상 장치의 해상도가 향상되면서 보다 많은 정보를 한 화면에 담기 위한 요구의 증가로 더욱 활발히 연구되고 있다. 이는 가상 환경의 구축이나 초고해상도 연구[1][2][3][4]와 관련하여 특히 활발히 연구되고 있으며, 최근에는 디지털 부호화된 비트스트림의 정보를 활용하거나[5] 정확한 움직임 정보를 얻는 데 초점을 둔 연구 등이 이루어지고 있다.

본 논문에서는 2장에서 파노라마 영상 합성을 효과적으로 수행하기 위한 전체 시스템의 구조를 설명하고 이에 필수적인 영상 분석 과정에 대하여 설명한다. 또한, 파노라마 영상 합성에서 고려해야 할 후처리 문제에 대해서도

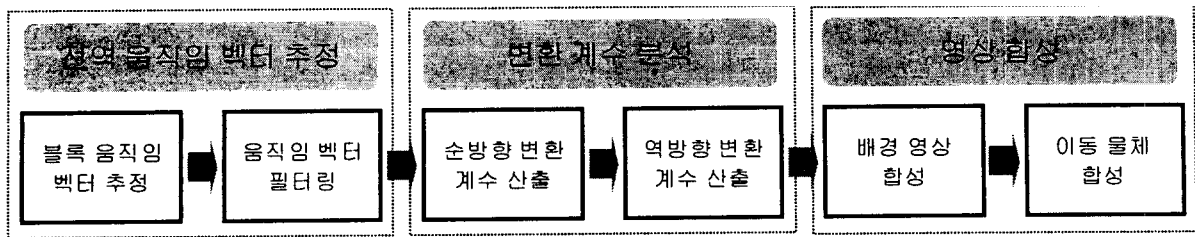


그림 1. 파노라마 영상 합성 장치의 구조

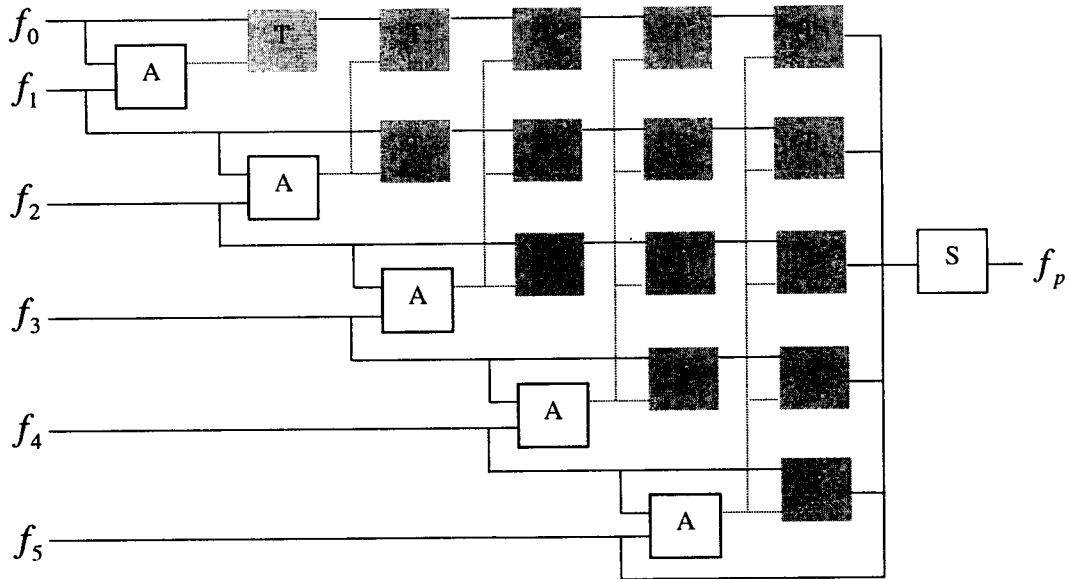


그림 2. 파노라마 영상 합성을 위한 영상 분석 및 변환부

다룬다. 마지막으로 3장에서 실험 결과와 함께 향후 연구과제를 제시함으로써 결론을 맺는다.

II. 파노라마 영상 합성 시스템

1. 영상 합성 시스템

그림 2는 파노라마 영상 합성 시스템의 블록 다이어그램을 보여주고 있다. 모든 영상을 특정한 기준 영상(f_5)의 영상 평면에 대응시키는 작업이므로, 기준 영상에 대한 각 영상의 변환 계수를 구해야 하고, 따라서 각 영상과 기준 영상 사이의 특징점 대응 관계를 분석해야 한다. 그러나 시간적, 공간적으로 거리가 있는 두 영상에서 공통된 특징점을 찾는 것은

매우 어려운 일이므로 인접한 프레임마다 특징점 대응 관계를 분석하여 변환 계수를 구한다. 다음 각 영상을 누적 변환시킴으로써 기준 영상에 대응하는 방법을 취한다. 따라서 모든 프레임에 대하여 연속적으로 변환 계수 및 영상 변환 작업을 수행하여 최종적인 파노라마 합성 영상을 출력한다. 그림에서 “A”로 표시된 블록은 연속되는 두 입력 영상으로부터 특징점 분석을 통해 변환 계수를 구하는 부분을 나타내고, “T”로 표시된 블록은 이렇게 얻어진 변환 계수와 변환 영상을 입력으로 받아 변환 결과를 돌려주는 부분을 나타낸 것이다. 전자는 카메라 움직임 모델에 근거한 3x3 계수 행렬의 산출 부분고, 후자는 이 3x3 행렬을 이용하여 입력 영상인 720x486 크기의 행렬의

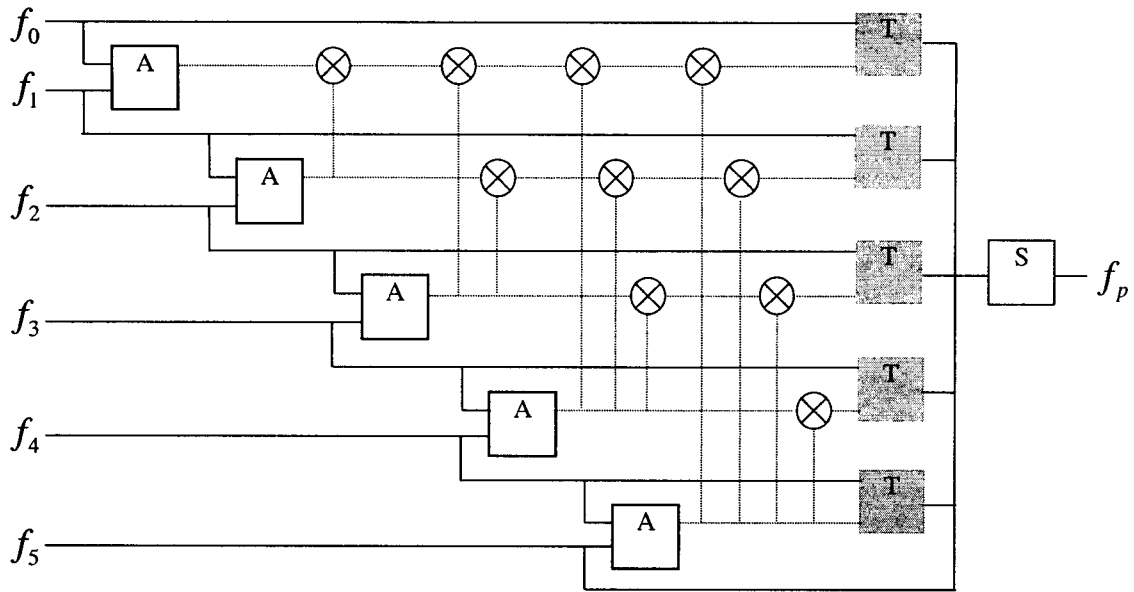


그림 3. 단순화된 영상 분석 및 변환부 구조

각 원소(좌표값)를 실제 변환하는 과정을 나타낸다.

영상 변환 작업에는 720x486 크기의 영상을 처리하기 때문에 이에 해당하는 크기의 메모리가 필요하다. 이는 입력 영상의 프레임 수가 증가할수록 많은 양의 메모리를 요구함을 의미하고, 메모리 요구량을 줄이기 위해서는 각 단계의 계산 결과를 저장하지 않고 매번 재수행하는 과정을 포함해야 하므로 시스템 구현 시 대용량의 메모리를 필요로 하거나 또는 실행 시간이 매우 길어지는 문제점을 안게 된다.

2. 단순화된 영상 합성 시스템

그림 2의 구조에서 보는 바와 같이 한 프레임을 기준 영상에 대응시키기 위해 기준 영상까지의 모든 변환식을 반복 적용하는 것보다 그 프레임과 기준 프레임 사이의 변환식을 구하여 이용하면 계산량을 크게 줄일 수 있다.

그림 3과 같이 단순화된 시스템에서는 중간과정에서의 영상 변환을 거치지 않고 변환 계수의 누적 연산을 통해 각 프레임의 기준 프레임에 대한 변환식을 구한 후, 최종적으로 파노

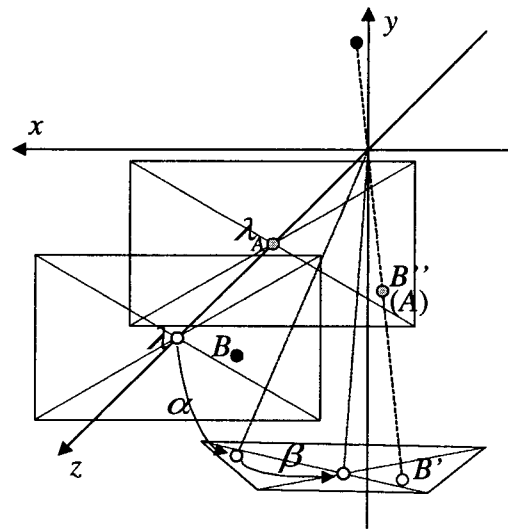


그림 4. 카메라 모델

라마 영상을 합성할 때 실제적인 영상 변환을 수행한다. 즉, P 가 720x486개의 화소로 이루어진 영상을 나타낸다고 할 때

$$A_{n-1}(A_{n-2} \cdots A_1(A_0(P)) \cdots)$$

의 형태를

$$A(P), \quad A = A_{n-1}A_{n-2} \cdots A_1A_0 \quad (1)$$

형태로 변형한다.

3. 파노라마 합성에서 영상 분석

파노라마 합성이 성공적으로 이루어지기 위해서는 정확한 움직임 정보의 분석이 선행되어야 한다. 영상 취득 단계에 특수한 장비나 촬영기술 등의 추가적인 요구가 없어야 한다는 점을 고려할 때, 단일 카메라에 의한 영상의 분석에는 큰 한계가 있다. 따라서, 그림 4와 같은 단순화된 카메라 모델을 가정한다. 여기서는 카메라의 움직임을 패닝(β), 틸팅(α), 주밍(λ, λ_A)만으로 묘사하며 단일 카메라 환경에 따른 거리 정보 문제를 해결하기 위해 패닝과 틸팅이 고정된 초점을 중심으로 이루어지고, 초점 거리 이동은 초점의 이동 대신 영상 평면의 움직임에 의해 이루어진다고 가정한다. 3차원 좌표계에서의 회전 및 스케일링 변환 등을 표현하는 일반적인 방법 [6]을 이용하면 이러한 모델에서 변환 계수 행렬은 다음과 같다.

$$\begin{bmatrix} b'_{i,x} \\ b'_{i,y} \\ 1 \end{bmatrix} = k_i \begin{bmatrix} \cos \beta & -\sin \alpha \sin \beta & -\lambda \cos \alpha \sin \beta \\ 0 & \cos \alpha & -\lambda \sin \alpha \\ \frac{\sin \beta}{\lambda_A} & \frac{\sin \alpha \cos \beta}{\lambda_A} & \frac{\lambda}{\lambda_A} \cos \alpha \cos \beta \end{bmatrix} \begin{bmatrix} b_{i,x} \\ b_{i,y} \\ 1 \end{bmatrix}$$

이러한 모델은 패닝과 틸팅이 영상 평면상의 x 와 y 좌표 모두에 영향을 주는 여전히 복잡한 움직임 묘사이며, 실제로 근을 구하는 데 상당한 어려움이 있다. 이를 더욱 단순화하기 위해 패닝은 x 좌표값, 틸팅은 y 좌표값에만 영향을 주고 주밍은 공통으로 작용하도록 모델을 수정한다. 여기에 피사체와 카메라의 커리가 충분히 크다는 가정하에 근사화를 통하여 다음과 같은 변환 계수 행렬을 얻는다.

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{\lambda'}{\lambda} & 0 & \lambda \tan \beta \\ 0 & \frac{\lambda'}{\lambda} & \lambda \tan \alpha \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (2)$$

식 (2)는 특정 프레임 상의 (x, y) 위치에 화소가 패닝(β), 틸팅(α), λ 에서 λ' 로의 초점

거리 이동에 의해 기준 영상 평면의 (x', y') 위치로 대응되는 관계를 나타낸 것이다.

일반적으로 파노라마 영상 합성의 대상이 되는 축구 경기 장면에서 카메라 움직임은 다음과 같이 크게 3가지로 표현될 수 있다.

1. 패닝
2. 틸팅
3. 주밍

파노라마 영상 합성의 대상이 되는 장면의 경우 패닝, 틸팅, 주밍의 순으로 비중이 크며 카메라의 패닝과 틸팅은 영상 평면에서 수평/수직 이동으로 근사화해도 무리가 없다. 특히 이러한 가정하에서 움직임 정보 추출에 일반적으로 쓰이는 블록 정합 알고리즘을 사용할 수 있다.

블록 움직임 벡터들 중에는 실제 움직임 정보와 무관한 값들도 포함되어 있다. 이는 움직임 벡터가 움직이는 물체 각각의 분석을 통해 이루어지기 보다는 단순히 블록이 포함하는 화소의 차를 통해 얻어졌기 때문이다. 실제 움직임(true motion) 정보를 얻기 위해서는 일반적으로 매우 복잡하고 계산량이 많은 방법을 동원해야 한다. 본 연구에서는 블록 정합 알고리즘으로 얻어진 움직임 벡터로부터 카메라 움직임 정보를 얻기 위해서 블록 움직임 벡터를 적절히 필터링하는 방법을 통해 실제 움직임에 가까운 움직임 정보를 구하였다.

이 외에 파노라마 합성 후 발생 가능한 빈 화소를 효과적으로 채우는 문제도 고려해야 한다. 이를 해결하기 위하여 위하여 실제 파노라마 영상 합성을 위한 각 프레임의 변환 과정을 역변환을 이용하여 수행한다. 즉, 새로운 좌표를 구하기 위해서

$$X' = AX, \quad X = (x, y, 1), \quad X' = (x', y', 1)$$

대신

$$X = A^{-1}X' \quad (3)$$

을 이용한다. 식 (3)을 이용하면 재구성하고자 하는 영상의 화소에 대응되는 원영상 화소의

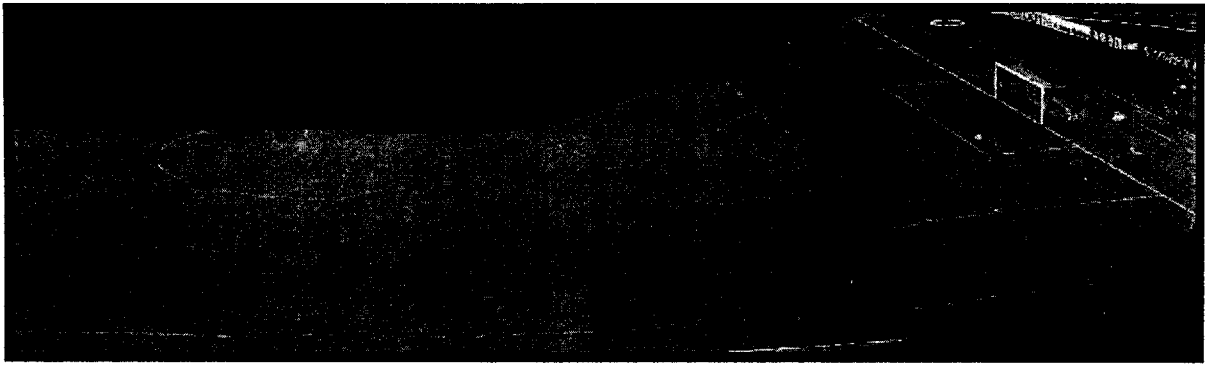


그림 5. 파노라마 영상 합성 결과

근사적인 위치를 구할 수 있으므로 필요한 화소의 값을 모두 얻을 수 있다. 동차(homogeneous) 좌표계 [6]를 이용함으로써 식 (2)와 (3)에서 보인 관계가 행렬로 간단히 표현된다.

4. 파노라마 합성에서 후처리 문제

파노라마 합성에 있어서 영상 정보의 조합뿐 아니라 시점이 다른 다수 영상을 한 화면에 나타냈을 때 각 부분의 움직이는 물체의 표현 방식도 중요한 문제이다. 각 영상에서의 선수와 공 등 움직이는 물체는 그 프레임의 시점에서만 의미가 있을 뿐이므로, 다른 영상(기준 영상)에 대응 변환되었을 때는 제거하는 것이 가장 타당한 표현 방식이다. 그러나, 움직이는 물체만 분리하기 위해서는 각 물체의 특성을 분석하고 판단하는 고도의 작업이 요구된다. 따라서 이 문제에 있어서도 파노라마 영상이 갖는 특성, 축구 경기를 대상으로 한다는 특수성 등을 효과적으로 활용하여 비용대 효과를 최대화할 필요가 있다.

정지된 배경과 움직이는 물체를 구분하는 판단 기준이 중요한 반면, 시스템의 복잡성을 최소화하고 계산량을 줄이기 위해서는 가능한 단순한 방법이 유리하다. 본 논문에서는 다음과 같이 통계적인 접근 방법을 채택하였다.

1. 시간상의 통계적 특성을 활용하는 방법
2. 공간상의 통계적 특성을 활용하는 방법

우선, 각 영상을 적절한 크기의 블록으로 나누어서 처리한다. 배경 영상을 만드는 과정에서 매 입력 프레임이 변환될 때마다, 직전 단계까지의 입력 영상으로부터 만들어진 배경 영상과 현재 프레임의 변환 결과를 블록 단위로 비교하여 복잡도가 높은 블록이 움직이는 물체에 해당된다고 판단하는 것이다. 이는 파노라마 변환 이후 각 변환된 영상의 동일 좌표상의 화소는 실제 동일한 물체에 대응된다는 점과, 경기장이 비교적 단순한 패턴으로 구성되어 있음에 착안한 것으로, 여기서 '복잡도'는 선수나 공을 포함하는 경우 단순한 패턴으로 구성된 같은 위치에 있는 다른 프레임의 블록에 비해 높은 값이 나오는 척도를 선정하면 된다. 본 개발에서는 분산을 활용하였다. 즉, 매 프레임의 변환시 모든 블록의 분산을 현재까지 만들어진 배경 화면의 같은 위치의 블록의 분산과 비교하여 현재의 것이 더 낮으면 배경일 가능성이 이전의 것보다 크다고 판단하는 것이다. 움직이는 물체가 나타나는 부분이 주로 운동장이고, 운동장의 패턴이 매우 단순하다는 점 때문에 비교적 우수한 결과를 얻을 수 있다. 오인 판단의 가능성을 최소화하기 위해서는 각 화소 위치의 정확한 변환이 필수적이다. 이는 배경과 선수 및 공을 구분하기 위한 목적도 있으나, 무엇보다도 다수 영상의 이음 부분을 자연스럽게 하는 데 있어서 근본적인 해결 방법이다.

또한, 식 (3)에서 보인 방법에 따라 각 화소의 값을 읽어오는 과정에서, 저역통과 필터링 역할을 하는 마스크 오퍼레이션을 적용함으로써 파노라마 영상 합성시 부표본화에 의해 나타나기 쉬운 에일리어싱을 효과적으로 억제할 수 있다.

III. 실험 결과 및 결론

실험은 실제 중계 방송에 이용된 영상을 대상으로 이루어졌다. 일반적으로 파노라마의 대상이 되는 영상은 좁은 공간 영역에서 복잡한 카메라 움직임을 통해 얻어진 것들보다는 비교적 넓은 영역에서 일관성있는 카메라 움직임을 통하여 얻어진 영상들이다. 그림 4에 보인 파노라마 영상 합성의 결과는 수평이동이 주가 되고 약간의 초점 거리 이동이 고려된 예이다.

카메라 움직임이 비교적 단순하다는 점, 즉 움직임의 방향과 크기에 변화가 크지 않다는 점은 움직임 정보를 추출하기가 비교적 용이하다는 장점으로 작용한다. 그러나, 카메라의 급격한 움직임에 의해 영상의 흐려짐(blur) 현상이 결과에 미치는 영향은 상당히 크다. 이는 축구와 같은 역동적인 대상의 촬영에서 피할 수 없는 문제이기 때문에 파노라마 영상 합성 시스템에서 이를 해결하여야 한다.

또한, 본 논문에서 제안한 방법은 초점 거리에 대한 가정 때문에 주밍에 취약하다는 문제가 있다. 이를 해결하기 위해 주밍의 비중이 큰 시퀀스의 경우, 변환 계수에 그 영향을 적절히 반영하는 처리가 필요하다.

마지막으로, 전체 시스템의 속도 향상을 위한 다양한 최적화 기법이 요구된다. 이러한 점들을 고려하여 시스템을 성능을 개선하고 안정화한다면 실제 방송에 매우 효과적으로 적용할 수 있을 것이다.

참고 문헌

- [1] S. E. Chen, "Quicktime VR – an image-based approach to virtual environment navigation," in *ACM Computer Graphics (SIGGRAPH'95)*, pp. 29-38, Aug. 1995
- [2] M. Irani and S. Peleg, "Improving resolution by image registration," in *CVGIP: Graphical Models and Image Processing*, vol. 53, no. 3, pp.231-239, May 1991
- [3] S. Mann and R. W. Picard, "Virtual bellows: Constructing high quality stills from video," in *Proc. IEEE Int. Conf. Image Processing*, Austin, TX, Nov. 1994, vol. I, pp. 363-367.
- [4] A. Zomet and S. Peleg, "Applying super-resolution to panoramic mosaics," in *Proceedings Fourth IEEE Workshop on Applications of Computer Vision (WACV'98)*, Los Alamitos, CA, pp.286-287.
- [5] Y. Altunbasak and A. J. Patti, "A Fast Method of Reconstructing High-Resolution Panoramic Stills from MPEG-Compressed Video," *IEEE Workshop on Multimedia signal processing*, Redondo Beach, CA, pp. 99-104, Dec. 1998
- [6] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, Reading, MA: Addison Wesley, 1993.