

Homography와 영상 분할을 이용한 실내 영상으로부터 의 기하정보와 카메라 정보의 추출

박태준*, 권대현, 오광만
한국전자통신연구원 가상현실연구센터

Geometry and Camera Recovery for Indoor Images using Homographies and Image Segmentation

Tae-Joon Park, Daniel D. Kwon, Kwang-Man Oh
Virtual Reality Research Center, ETRI

요 약

본 논문에서는 다수의 실내 영상으로부터 영상을 촬영한 카메라의 속성정보와 실내 환경에 대한 기하정보를 추출하는 방법을 제안한다. BSP-Tree를 이용하여 주어진 실영상을 각각의 부분 영역이 실제로도 평면 영역에 해당되도록 분할하였으며, 특징점 대응을 통해 각 분할된 영역의 영상간 대응을 찾고 이로부터 각 분할 영역의 homography를 계산하였다. 또한 간단한 가정을 통해 계산된 homography로부터 각 분할영역에 대응된 평면의 방정식과 각 영상을 촬영한 카메라의 속성을 찾아낼 수 있음을 보였다. 본 논문에서 제안한 방법은 현재 본 연구팀이 구현 중인 영상기반 모델링 시스템에서 핵심적인 기능을 수행하리라 기대된다.

1. 서론

다수의 영상으로부터 그 영상에 나타난 환경의 기하정보와 각각의 영상을 촬영한 카메라의 속성을 추출하는 문제는 structure from motion이라는 이름으로 널리 알려져 있으며 많은 연구결과가 발표되어 있다 [1,2,3]. 하지만 기하정보와 카메라 속성 추출을 위해 매우 복잡한 계산을 필요로 하기 때문에 실시간 동작이 어려우며, 그 계산 과정에서 발생하는 여러 가지 오류로 인해 신뢰성 있는 결과를 얻기가 매우 힘들다는 문제가 있다.

최근에는 보다 제한된 경우에 대해 보다 쉽게

structure from motion 문제를 해결하고자 하는 시도가 발표되고 있다. 특히 컴퓨터 그래픽스 및 가상 현실 분야에서 복잡한 물체에 대한 모델링하는 과정에 이러한 방법론이 매우 활발하게 도입되고 있다 [4]. 특히 이 경우에는 물체가 컴퓨터 상에 그려졌을 때 이를 보는 사람들이 마치 실제 물체라고 느끼게 되는 정도의 정확도만이 요구되기 때문에, 물체에 대한 공학적인 분석이나 기타 매우 정밀한 계산결과를 요구하는 분야와는 달리 아주 간단한 연산만으로 어느 정도의 품질을 얻을 수 있다. 이러한 결과 중 이미 상용화된 대표적인 것이 Canoma와 Façade로서 [4], 이들은 주어진 환경이 육면체나 피라미드와 같이 아

주 간단한 몇 개의 기본 물체들로만 이루어져 있다고 가정하고, 주어진 영상으로부터 이들 기본 물체의 기하정보를 결정하는 조건들-예를 들면, 육면체의 각 꼭지점이 영상에서 나타나는 위치-을 사용자가 직접 마우스 등을 이용하여 입력하게 한 후 기하정보를 추출한다. 하지만 이들 시스템은 사용이 편리하고 계산과정이 간단하다는 장점에 반해, 복잡한 형태의 환경에 대해서는 적용이 곤란하다는 단점이 있다. 이에 반해 Photo Modeler와 같은 시스템에서는 복잡한 형태의 물체에 대한 기하정보 추출이 가능하지만, 사용하기가 아주 복잡하다는 단점이 있다.

본 논문에서는 사용이 간편하고 계산이 간단하면서도 복잡한 환경에 대한 기하정보 추출이 가능한 해결방법을 제시하고자 한다. 특히, 제시하는 방법은 실내 영상과 같이 대부분의 물체가 평면의 조합으로 구성된 환경에 잘 적용된다. 우선 BSP-tree를 이용한 영상 분할을 이용하여 각각의 영상을 각 부분영상이 실제로도 환경 내의 평면 부분에 대응되도록 분할한다. 이어서 각 분할 영상에 대한 특징점 추출과 추적을 통해 대응관계를 찾고, 이로부터 분할 영상간의 homography를 계산한다. 또한 분할 영상의 homography는 두 영상을 촬영한 카메라의 상대적인 위치변화와 분할 영상과 대응된 환경내의 평면의 방정식으로부터 유도될 수 있으므로 [5,6], 간단한 가정을 통해 homography로부터 이들을 계산해내는 방법을 제시한다.

2. BSP-tree를 이용한 영상분할

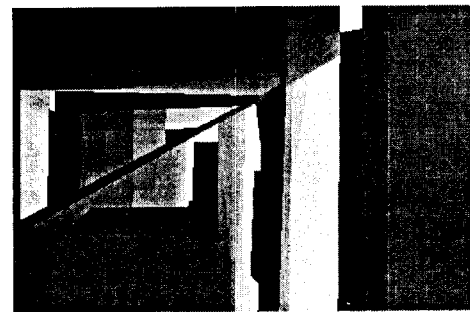
본 논문에서는 다음과 같은 방법으로 각각의 영상을 분할한다. 우선 주어진 영상에 Canny operation과 Hough 변환[7]을 적용하여 직선 성분을 추출한다 ([그림 1 (b)] 참조). 계속해서 구해진 직선성분을 이용하여 주어진 영상에 대한 BSP-tree 분할[8]을 얻는다 ([그림 1 (c)] 참조). 최종 영상 분할은 구해진 BSP-tree의 각 분할 영역 중, 서로 인접하면서 색상이 비슷한 영역을 병합하여 구한다 ([그림 1 (d)] 참조). 이러한 과정의 각 단계별 결과를 [그림 1]에 보였다.



(a) 초기 영상



(b) 추출된 직선성분



(c) BSP-tree를 이용한 영상 분할



(d) 최종 영상 분할

[그림 1] BSP-tree를 이용한 영상 분할

이렇게 구해진 각각의 분할 영역에 대해 특징점 추출과 추적을 통해 대응관계를 할당하게 되며, 할당된 대응관계로부터 homography가 계산되고, 계속해서 카메라 속성 추출과 평면 방정식 계산으로 이어지게 된다.

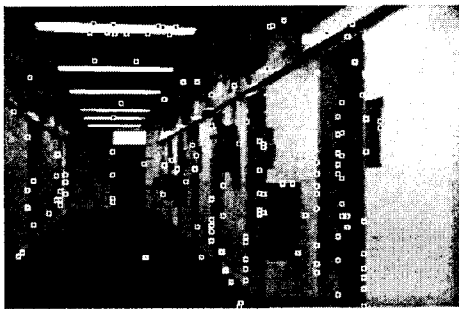
3. 특징점 추출/추적과 homography 계산

3.1 특징점 추출과 추적

본 논문에서는 KLT 특징점 추출 및 추적 알고리즘을 이용한다 [9]. 특징점 추출과 추적 결과는 [그림 2]에 보였다. 이러한 특징점 추적 결과는 대응관계의 형태로 저장이 되며, 다시 homography 계산과정에 활용된다.



(a) 첫번째 영상에서 추출된 특징점들



(b) 두 번째 영상에서 추적된 특징점들

[그림 2] 특징점 추출과 추적

3.2 Homography 계산

Homography는 서로 대응된 두 분할영역간의 2차원 projective mapping을 정의하는 3×3 행렬이다. 예

를 들어 어느 한 픽셀 (x_i, y_i) 가 다른 영상에서 (x_j, y_j) 에서 나타났고, 이 픽셀에 할당된 homography가 H 라고 한다면,

$$H \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = \begin{bmatrix} wx_j \\ wy_j \\ w \end{bmatrix} \quad (1)$$

의 관계가 성립된다.

Homography 행렬은 projective map이기 때문의 임의의 0이 아닌 λ 에 대해 H 와 λH 는 같은 map을 정의한다. 따라서 homography의 자유도는 8이며 4개의 대응관계로부터 구해질 수 있다. 편의를 위해 본 논문에서는 homography 행렬이라는 용어를 같은 projective map을 정의하는 3×3 행렬 중 마지막 원소가 1인 것을 뜻하는 용어로 사용하겠다.

각 분할 영역에 대해 다수의 대응관계가 존재하므로 본 논문에서는 이를 대응관계를 가장 잘 반영하는 homography를 찾기 위해 Levenberg-Marquardt 방법을 적용한다 [10]. 일단 homography가 계산되면 해당 분할 영역의 대응 관계 중 계산된 homography에 비추어 오류를 많이 발생시키는 대응관계를 제거하고 다시 homography를 계산하는 과정을 반복한다. 이 과정은 계산된 homography로 계산된 대응관계와 특징점 추적을 통해 얻어진 대응관계 사이의 차이가 충분히 적어질 때까지 계속된다. 특징점 추적단계에서 4개의 대응관계를 찾지 못한 분할 영역에 대해서는 주변 분할영역에 대해 계산된 homography 중 픽셀의 색상차이를 가장 적게 하는 homography를 할당한다.

4. 카메라 속성 및 평면 방정식 추출

본 논문의 결과는 첫 영상을 촬영한 카메라의 좌표계가 주어진 환경의 좌표계와 일치하며 카메라의 초점 거리는 1이라고 가정한다. 이 때 두 번째 영상을 촬영한 카메라의 transform 행렬 P_2 를

$$P_2 = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \quad (2)$$

라고 하고 첫번째 영상과 두 번째 영상에서 평면 방정식이 $ax + by + cz + 1 = 0$ 인 평면에 대응되는 영역을 각각 f_1 과 f_2 라고 하자. 이 때, 영역 f_1 에 대응되는 homography H 와 카메라 transform 행렬 P_2 , 그리고 평면의 방정식 $ax + by + cz + 1 = 0$ 사이에는

$$(p_{33} - cp_{34})H = \begin{bmatrix} p_{11} - ap_{14} & p_{12} - bp_{14} & p_{13} - cp_{14} \\ p_{21} - ap_{24} & p_{22} - bp_{24} & p_{23} - cp_{24} \\ p_{31} - ap_{34} & p_{32} - bp_{34} & p_{33} - cp_{34} \end{bmatrix} \quad (3)$$

와 같은 관계가 있다.

식에서 보는 바와 같이 homography 행렬 H 의 모든 원소와 카메라 transform 행렬 P_2 의 원소, 그리고 직선의 방정식을 이루는 세 변수간의 관계는 모두 선형시스템으로 이루어져 있다. 카메라 행렬의 자유도가 12이고, 평면의 방정식의 자유도가 3이므로, 적어도 3개 이상의 분할영역에 대한 homography가 주어진다면 선형 시스템에 대한 최적화 방법을 이용하여 쉽게 평면의 방정식과 카메라 속성을 구할 수 있다.

5. 결론

본 논문에서는 실내영상으로부터 이 영상을 촬영한 카메라의 속성정보와 촬영된 환경을 구성하는 평면의 방정식을 추출하는 방법을 제시하였다. BSP-tree를 이용한 영상분할과 특징점 추출 및 추적, Levenberg-Marquardt 방법을 이용한 homography 계산법이 적용되었으며, homography로부터 간단한 가정을 통해 카메라 속성과 평면 방정식을 추출하는 방법이 제안되었다.

향후 연구를 통해 3개 이상의 영상을 사용하는 경우에 환경에 대한 최적의 기하모형을 구축하는 방법과 특징점 추출과 추적 과정에서 발생하는 오류를 가능한 줄이는 방법을 포함시키고자 한다. 또한 단순히 카메라의 속성을 3×4 행렬의 형태로만 표현하는 것이 아니라 quaternion이나 Euler각에 대한 분석을 통해 이를 초점 거리, translation 성분, 그리고 rotation 성분으로 분석하여 보다 명확한 카메라 속성 분석이 가능하도록 하는 방법을 제시하고자 한다. 이 경우 카메라 transform 행렬의 자유도가 7이 되므로, 2개의

homography만으로도 평면 방정식과 카메라 속성분석이 가능해진다. 본 논문에서 제안한 방법은 본 연구팀에서 연구과제로 수행중인 영상 기반 모델링 시스템에 포함될 예정이다.

[참고문헌]

- [1] C. Baillard, C. Schmid, A. Zisserman, and A. Fitzgibbon, "Automatic Line Matching and 3D Reconstruction of Buildings from Multiple Views", In ISPRS Conference on Automatic Extraction of GIS Objects from Digital Imagery, IAPRS Vol.32, Part 3-2W5, pages 69-80, September, 1999
- [2] C. Baillard, and A. Zisserman, "Automatic Reconstruction of Piecewise Planar Models from Multiple Views", In Proc. IEEE Conference on Computer Vision and Pattern Recognition, pages 559-565, June, 1999
- [3] A. Fitzgibbon, G. Cross, and A. Zisserman, "Automatic 3D Model Construction for Turn-Table Sequences", In 3D Structure from Multiple Images of Large-Scale Environments, LNCS 1506, Springer-Verlag, 1998
- [4] P. E. Debevec and C. J. Taylor and J. Malik, "Modeling and Rendering Architecture from Photographs: A Hybrid Geometry - and Image-Based Approach", Computer Graphics (ACM SIGGRAPH '96), pp11-20, 1996
- [5] R. Szeliski, "Video Mosaics for Virtual Environments", IEEE Computer Graphics & Applications, Vol. 16, No. 2, pp22-30, 1996
- [6] R. Szeliski and H.-Y. Shum, "Creating Full View Panoramic Image Mosaics and Environment Maps", Computer Graphics (ACM SIGGRAPH '97), pp251-258, 1997
- [7] R. C. Gonzalez and R. E. Woods, "Digital Image Processing", Addison-Wesley, 1992
- [8] K. R. Subramanian, and B. F. Naylor, "Converting Discrete Images to Partitioning Trees", IEEE Transactions on Visualization and Computer Graphics, Vol. 3, No. 3, pp273-288, 1997
- [9] J. Shi and C. Tomasi, "Good Features to Track", IEEE Conference on Computer Vision and Pattern Recognition (CVPR94), June 1994
- [10] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, "Numerical Recipes in C: The Art of Scientific Computing, 2nd Ed.", Cambridge University Press, 1992