

영상정보를 이용한 3차원 정보의 정합

하승태^U 한준희
포항공과대학교 컴퓨터공학과
{hast, joonhan}@postech.ac.kr

3D Range Data Registration Using Corresponding Image

Seung-Tae Ha^U JoonHee Han
Dept. of Computer Science & Engineering, POSTECH

요 약

3차원 정보의 올바른 정합을 위해서 3차원 정보 자신 뿐만이 아니라 3차원 정보와 연관된 영상 정보를 이용한다. 먼저 영상의 정합을 수행함에 있어 서로 다른 두 영상간에 상관 윈도우를 씌워 상관계수를 계산하여 최적 정합점을 탐색한다. 본 논문에서는 카메라의 서로 다른 관점으로 인한 상관 윈도우의 뒤틀림을 3차원 초기 변환 행렬을 이용하여 보정하는 방법을 제안하고, 이에 의해 3차원 변환된 상관 윈도우를 정합에 이용함으로써 상관 계수의 정확도를 급격히 향상시킨다. 그 결과로 개선된 특징점 정합 결과로부터 영상 전반에 걸친 3차원 특징점 정합을 통해 이와 대응하는 3차원 정보의 정확한 정합 결과를 얻는다.

1. 서론

3차원 정보의 통합에 있어서 올바른 정합은 그 기초적이며 가장 중요한 역할을 한다. Besl[1]과 Chin[2]에서는 서로 다른 3차원 정보들이 주어진 상태에서, 대응되는 3차원 좌표들을 직접 구하여, 이 좌표들의 변환 좌표로부터 변환 행렬을 구해내는 방법이 시도되었다. 이 방법은 3차원 좌표의 대응을 직접 얻어내어야 하는 까다로운 작업을 요구 하였고, 따라서 에러 발생의 소지가 높다고 할 수 있다. Besl[3]은 ICP 알고리즘을 제안하였는데, 이는 3차원 통합을 위해 각 2.5차원 좌표간에 가까운 점 탐색(nearest point search)를 반복 수행함으로써 3차원 좌표의 대응을 얻었다. 이에 반하여 본 논문에서는 스테레오 정합(stereo matching)[4]이나 셀프-캘리브레이션(self-calibration)[5]과 같은 3차원 정보들을 이용한다. 이는 기존의 거리센서를 이용한 정보의 정합과는 달리 2차원 영상을 3차원 정합의 특징으로 삼을 수 있는 강점을 갖는다. 본 논문이 제안하는 방법에서는 초기 3차원 대응점의 추출 방법으로서 2차원 영상의 대응점을 이용, 간편하고 정확한 대응점의 추출이 가능하다.

또한 1단계 정합에서는 초기 변환행렬을 이용하여 변환된 상관 윈도우로서 초기 대응점을 개선한다. 2단계 정합에서는 무작위로 추출한 영상정보의 특징점들을 대상으로 대응점들을 찾음으로서 정확한 변환행렬을 구한다.

본 논문의 구성은 다음과 같다. Chapter 2에서는 대응점을 통한 변환 행렬의 산출법을, Chapter 3에서는 초기 3차원 변환을 이용한 개선된 상관 윈도우를 제안한다. 또한, Chapter 4에서는 이를 이용한 전체적 알고리즘의 구조를 보이고, Chapter 5에서는 실제 실험 결과를 보인다. 마지막으로 Chapter 6에서는 결론 및 개선점을 토론한다.

2. 대응점을 통한 3차원 변환 행렬의 도출

2.1 3차원 변환 행렬

서로 다른 두 3차원 좌표계의 변환 관계는 3차원 변환 행렬로서 나타내어질 수 있다. 이 때, 3차원 변환 행렬은 수식 (1)에 주어지는 것과 같은 3*4의 크기를 가지는 일반적인 행렬이다.

$$M = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \quad (1)$$

12개의 모르는 변수로 이루어진 이 행렬을 구하기 위해서는 서로 다른 3차원 좌표계간의 대응점이 일정 수 이상 존재하여야 한다. 따라서 좌표계간의 정확한 점대점 연관 관계를 구하는 것이 바로 3차원 변환 행렬을 구하는 필요충분조건이라 할 수 있다.

2.2 변환행렬의 도출

2.2.1 강체 변환

두 좌표계의 2.5차원 정보들이 완전한 유클리디안(Euclidean) 정보일 때, 두 2.5차원 정보 사이의 변환은 강체변환(rigid-body transformation)의 성질이 있다. 이는 이동, 회전, 크기조정의 세가지 요소로 이루어진다. 대응하는 3차원 점의 집합이 주어졌을때, 이 점의 변환 관계가 강체변환이라고 한다면, Horn[6]에 의해 강체 변환 행렬을 구해 낼 수 있다.

2.2.2 일반적 변환

일반적 변환을 구하기 위해서는 [7]의 다중 선형 Regression을 이용한다. 서로 다른 3차원 좌표계의 \vec{x}_i 와 \vec{y}_i 가 대응하고 변환행렬이 M 이라고 할 때 $\vec{y}_i = M\vec{x}_i$ 의 관계가 있다. M 이 3*4 행렬이므로 이를 풀면 다음의 수식이 도출된다.

$$\begin{aligned} y_1 &= a_{11}x_{11} + a_{12}x_{21} + a_{13}x_{31} + a_{14} \\ y_2 &= a_{21}x_{11} + a_{22}x_{21} + a_{23}x_{31} + a_{24} \\ y_3 &= a_{31}x_{11} + a_{32}x_{21} + a_{33}x_{31} + a_{34} \end{aligned} \quad (2)$$

여러 대응점으로부터 M 을 구할때는 수식 (2)에 근접하는 M 을 구하여야 하는데, 이 때 가장 윗 식을 본다면 최소화되어야 할 에러는 수식(3)과 같다.

$$Err_1 = \sum_{i=1}^n (y_{i1} - a_{11}x_{i1} - a_{12}x_{i2} - a_{13}x_{i3} - a_{14})^2 \quad (3)$$

주어진 에러를 최소화하는 최적의 값은 수식 (4)와 같다.

$$\begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \end{bmatrix} = \begin{bmatrix} \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \sum x_{i1}x_{i3} & \sum x_{i1} \\ \sum x_{i1}x_{i2} & \sum x_{i2}^2 & \sum x_{i2}x_{i3} & \sum x_{i2} \\ \sum x_{i1}x_{i3} & \sum x_{i2}x_{i3} & \sum x_{i3}^2 & \sum x_{i3} \\ \sum x_{i1} & \sum x_{i2} & \sum x_{i3} & n \end{bmatrix}^{-1} \begin{bmatrix} \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \sum x_{i3}y_i \\ \sum y_i \end{bmatrix} \quad (4)$$

$a_{21} \sim a_{24}$, $a_{31} \sim a_{34}$ 도 위와같이 계산하면 변환행렬 M 의

12개 구성요소를 얻을 수 있다.

3. 개선된 상관계수의 계산

3.1 상관 윈도우의 뒤틀림

두 영상 사이의 대응관계를 설정하기 위해서는 영상간의 특징점 정합이 필요하다. 그러나 3차원 정합시의 특징점 정합시는 모델에 대한 영상을 얻은 두 카메라의 위치의 차이가 존재하므로, 대응하는 특징점의 정합시 시정 차로 인한 상관 윈도우의 뒤틀림이 생기게 된다.

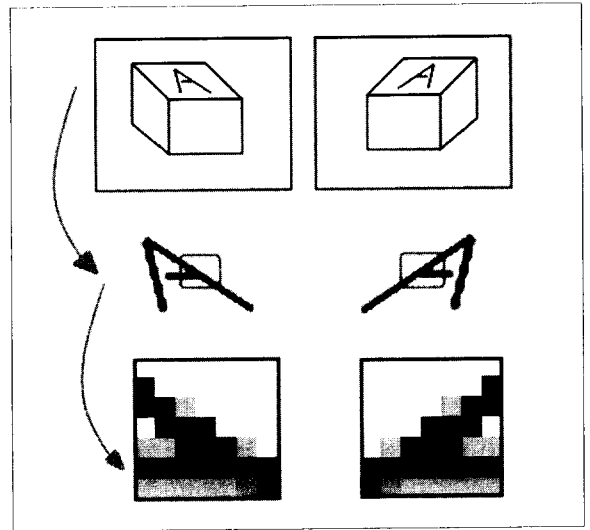


그림 1 : 시정에 따라 다른 물체의 영상과 상관 윈도우

그림 1에서 A자를 보면 카메라의 시정이 다르므로 두 영상에서 다르게 보이게 된다. 이러한 현상은 카메라의 위치차가 크면 클수록 다르게 보이는 정도가 커지게 된다. 따라서 초기 정합을 위해 사용되는 영상의 상관 계수를 계산할 때 변환된 영상을 이용하지 않으면, 정합의 정확성은 감소하고 잘못된 대응점을 찾을 확률이 높아진다.

3.2 3차원 변환을 이용한 개선된 상관 윈도우

영상에서 정합하고자 하는 두 특징점의 좌표를 m , m' 이라 하면, 영상에서 m , m' 에 직접 씌운 윈도우의 각 화소의 대응은 3.1절에서 언급한 이유로 정확한 대응이 되지 않을 수 있다. 즉, m 상관 윈도우의 (i, j) 번째의 화소는 m' 상관 윈도우의 (i, j) 번째 화소와 대응되는 화소가 아닐 수 있다. 이는 3차원 변환으로 인한 뒤틀림 때문인데, 반대로 이러한 3차원 변환의 대강 값을 알고 있다고 가정할 때 뒤틀림을 어느정도 보완할 수 있다는 뜻이 된다.

첫번째 영상에서 직접 씌운 상관윈도우 화소 (i, j) 의 밝기는 $I_{1(m)}(i, j)$. 두번째 영상에서 직접 씌운 상관 윈도우 (i, j) 의 밝기는 $I_{2(m')}(i, j)$ 라 정의한다. 또한, 첫번째 영상의 상관

원도우의 (i, j) 번째 화소에 대응하는 3차원 좌표를 (a_{ij}, b_{ij}, c_{ij}) 라 하고, 이를 초기 변환 행렬 M 에 의해 변환한 새로운 좌표를 $(a'_{ij}, b'_{ij}, c'_{ij})$ 이라 한다. 그렇다면 이는 첫번째 영상에서의 원도우를 두번째의 영상에서의 3차원 좌표로 변환한 것이 된다. 변환된 각 원도우의 좌표들을 다시 두번째 영상으로 대응 시키면 대응된 화소의 밝기가 변환된 원도우의 밝기값이 된다. 즉, 3차원 좌표 (a_{ij}, b_{ij}, c_{ij}) 가 두번째 영상으로 투사된 2차원 좌표를 $(\alpha_{ij}, \beta_{ij})$ 라 할때 보정된 새로운 원도우의 (i, j) 번째 밝기값은 수식 (5)와 같다.

$$I_2(m'_{\dots})(i, j) = I_2(\alpha_{ij}, \beta_{ij}) \quad (5)$$

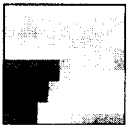


그림 2 : 첫번째 영상의 원도우

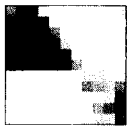


그림 3 : 두번째 영상의 원도우



그림 4 : 개선된 두번째 영상의 원도우

4. 알고리즘의 구조

Chapter2, Chapter3에서 제안한 방법을 바탕으로 서로다른 3차원 정보의 정합을 수행한다.

- 1) 영상에서 6개 이상의 초기 대응점을 얻는다.
- 2) 영상과 상응하는 3차원 정보간의 초기 대응 관계를 구한다.
- 3) 위에서 구한 초기 변환으로서 각 대응점에 대해 변환된 각각의 상관 원도우를 구한다.
- 4) 영상에서 상관 원도우를 이용하여 최적 대응점을 구한다.
- 5) 3~4)의 작업을 반복한다.
- 6) 영상에서 추출한 특징점 중 무작위로 충분한 수를 고른다.
- 7) 추출한 특징점들에 대해 3~4)의 과정을 반복한다.

영상에서의 모든 화소에 대해 대응점을 구하는 것이 일반적이겠지만, 시간이 많이 소요되며 무늬(texture)가 없을 경우 상관계수의 의존성이 떨어지므로, 모서리 화소를 비롯한 특징점을 무작위로 선택하여 상관 계수를 구한다. 충분한 숫자의 특징점을 선택할 경우 3차원 정보 전반에 걸친 정확한 정합을 수행할 수 있다.

5. 실험 결과

실험은 그림 5,6과 같은 모자이크 상자를 이용하였다. 서로 다른 각도에서 얻은 두개의 3차원 스테레오 정보를 그 대상으로 하였으며, 10개의 초기 대응점을 영상에서

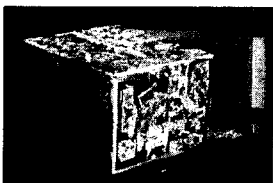


그림 5: 좌측에서 본 모자이크 상자

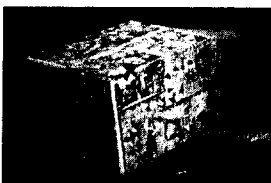


그림 6: 우측에서 본 모자이크 상자

수작업으로 구하였다.

표 1은 정합시에 개선된 상관 계수를 보여주고 있다. 3차원 변환을 적용하지 않은 초기 대응점의 상관계수, 3차원 변환을 적용한 초기 대응점의 상관계수, 추출한 특징점에 대한 상관 계수가 각각 나타나 있다. 그림 7은 영상을 기반으로 한 정합의 결과이다.

| window size | initial (per point) | refined (per point) | for sample (per point) |
|-------------|---------------------|---------------------|------------------------|
| 5 | 0.263684 | 0.769743 | 0.670006 |
| 7 | 0.204289 | 0.742564 | 0.760541 |
| 11 | 0.231317 | 0.676981 | 0.649757 |

<표 1> : 개선된 상관계수

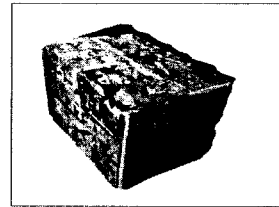


그림 7 : 정합 결과

7. 결론

2차원 영상에 의해 개선된 3차원 정합 방법에 대해 알아보았다. 영상을 이용한 정합 방법은 기존의 정합방법과는 달리 직관적이며 국부 최소값에 빠질 우려가 없다는 강점이 존재한다. 단점으로는 무늬가 거의 없을 경우 정합에 어려움이 있다. 향후에는 기존 ICP 알고리즘등과 혼합한 형태의 새로운 정합 방법이 가능하다고 할 수 있겠다.

6. 참고문헌

- [1] P. Besl and R. Jain, "Three dimensional object recognition", *ACM Comput. Surveys*, vol 17(1), pp 75-145, 1985
- [2] R. Chin and C. Dyer, "Model-based recognition in robot vision", *ACM Comput. Surveys*, vol 18(1), pp 67-108, 1986
- [3] P. Besl and Mckay, "A method for registration of 3d shape", *IEEE Trans. PAMI*, vol 14(2), pp. 239-255, 1992
- [4] S. T. Barnard and M. A. Fishler, "Computational Stereo", *Computing Surveys*, vol 14(4), pp. 553-572, 1982
- [5] O. Faugeras, !.-T. Luong, and S. Maybank, "Cameras self-calibration: theory and experiments", *ECCV*, pp 321-334, 1992
- [6] B. Horn, "Closed-form solution of absolute orientation using unit quaternions", *J. Opt. Soc Amer. : A.*, vol 4(4), pp. 629-642, 1987
- [7] S. Shapra and R. Canale. *Numericak Method of Engineers (2nd Ed.)*, McGraw-Hill Company, 1990