

Visual Servoing을 위한 3차원 물체의 인식 및 자세 추정

양재호** 정문호* 박민용**

한국과학기술연구원 지능로봇연구센터*, 연세대학교 전기전자공학과**

Recognition and Pose Estimation of 3-D Objects for Visual Servoing

Yang, JaeHo** Jeong, MoonHo* Park, Mignon**

Korea Institute of Science and Technology, Intelligent Robotics Research Center (KIST IRRC)*
Department of Electrical & Electronic Engineering, Yonsei University**

Abstract - 로봇이 어떤 물체를 인지하고 그 물체에 대해 어떤 작업을 하고자 할 때 특정 물체의 인식 문제, 3차원 정보를 획득하는 문제, 자세를 추정하는 문제 등 해결해야 될 문제들이 있다.

물체를 인식하는 과정에서는 주위 배경과 물체의 크기의 변화, 회전, 가려짐 등으로 인해 물체 인식을 어렵게 만드는 요소들이 있다. 2차원 이미지를 통해 3차원 정보를 추출하는 과정은 일반적으로 두 대의 카메라를 이용하여 스테레오 이미지를 통해 얻는다. 이 때 좌우 영상간의 매칭의 과정이 필요하다. 자세 추정의 문제는 카메라 좌표와 물체의 좌표간의 관계를 알아야 한다. Visual Servoing을 어렵게 만드는 많은 요인들이 있으며 본 논문에서는 물체의 크기, 회전, 이동에 불변인 디스크립터(descriptor)를 사용하는 SIFT(Scale Invariant Feature Transform)를 통해 3차원 물체의 인식과 자세를 추정하는 방법을 제시한다. 또한 자세 추정을 위해 2차원 Keypoint들의 매칭을 3차원 정보를 통해 검증하는 방법을 제시한다. (SIFT에 의해 추출된 point를 Keypoint라 명한다.)

2.1 SIFT를 이용한 2차원 DB 획득

Visual Servoing을 위한 물체 인식과 자세 추정은 주위 배경이나 가려짐, 크기 회전에 강인해야 하고 point들의 정보를 이용해야 하기 때문에 국소 정보를 이용한 접근법이 적합하다. 여기에서는 조명이나 회전, 크기에 강인한 SIFT(Scale Invariant Feature Transform)를 사용한다. [1]

물체 인식을 위해 SIFT를 수행하여 인식하고자 하는 물체에 대한 2차원 정보를 획득한다.

2.2 SIFT를 이용한 3차원 DB 획득

3차원 DB를 획득하기 위해서 먼저 회전축과 camera 간의 calibraion을 수행한다. [2]

3차원 정보 획득을 위해서 본 논문에서는 스테레오 카메라를 이용한다. 스테레오 카메라를 통해 스테레오 이미지를 획득하고 이미지 매칭을 통해 깊이 정보를 획득한다. 이미지 매칭은 SIFT를 이용하여 매칭을 수행한다.

1. 서 론

로봇이 인간과 공존하며 인간과 함께 어떤 작업을 하기 위해선 로봇의 Visual Servoing이 무엇보다 중요시 된다. 카메라를 통해 들어오는 영상을 이용하여 물체를 인식하고 Visual Servoing을 수행한다.

로봇의 Visual Servoing 을 위해서는 먼저 Target 물체를 정하고 그 물체의 위치와 방향을 인지해야 한다. 이미 알고 있는 DB를 통해 물체를 인식하고 스테레오 카메라를 통해 그 물체에 대한 3차원 좌표를 구한다. 또한 Visual Servoing을 위해서 그 물체의 자세를 추정해야만 한다. 물체의 자세 추정을 통해 Visual Servoing을 수행한다.

물체 인식을 위해 물체의 크기, 회전, 이동에 불변인 디스크립터(descriptor)를 사용하는 SIFT(Scale Invariant Feature Transform)를 사용한다. SIFT를 통해 물체를 인식한다.

3차원 정보 획득을 위해 스테레오 비전을 사용한다. 두 영상 간의 매칭은 SIFT를 이용하여 매칭을 수행한 후, 3차원 정보를 획득한다. 획득된 3차원 정보들을 가지고 원하는 물체의 자세 추정을 수행한다. 정확한 자세 추정을 위해 3차원 정보를 이용하여 위치 오차를 통한 매칭 검증을 수행한다.

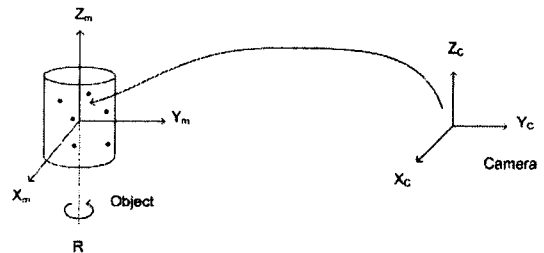
2. 본 론

그림 1 과 같이 Off line 과 On line 과정이 필요하다. Off line 과정에서 SIFT를 이용해서 물체 인식을 위한 2차원 DB와 자세추정을 위한 3차원 DB를 만든다. On line 과정에서는 구해진 DB를 가지고 물체인식과 자세추정을 수행한다.



<그림 2> SIFT를 이용한 스테레오 매칭

3차원 DB를 획득하기 위해서 정해진 Rotation을 수행해 가며 SIFT를 통한 스테레오 매칭을 수행한다.



<그림 3> 3차원 정보 획득

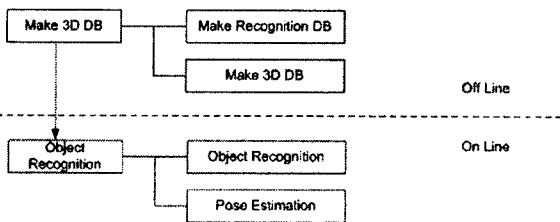
SIFT를 이용해서 구해진 3차원 정보를 이용해 자세추정을 위한 3차원 DB를 구한다.

$$\begin{bmatrix} x_{ic} \\ y_{ic} \\ z_{ic} \end{bmatrix} = {}^c T_m \times \begin{bmatrix} x_{im} \\ y_{im} \\ z_{im} \\ 1 \end{bmatrix} \quad \begin{array}{l} x_{ic} : i \text{ th Camera Coordinate} \\ x_{im} : i \text{ th Model Coordinate} \end{array} \quad (1)$$

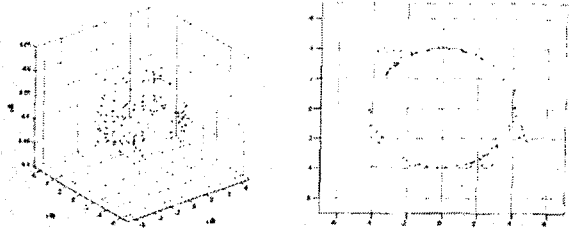
${}^c T_m$: Model to Camera Transform

$$E = \sum_{i=1}^n \| X_{ic} - {}^c T_m \cdot X_{im} \|^2 \quad (2)$$

식 (2)의 Error function을 최소화 하는 X_{im} 을 powell method를 사용하여 구한다.[3]



<그림 1> 전체 흐름도



〈그림 4〉 3차원 DB 획득

물체 인식과 자세 추정을 위한 DB를 만드는 Off line 과정이 끝났으면 실제 영상을 통해 물체 인식과 자세 추정을 통해 Visual Servoing을 수행한다.

2.3 물체 인식

2.1에서 SIFT를 통해 구해진 DB를 가지고 실제 들어오는 영상에 대한 물체 인식을 수행한다.



〈그림 5〉 SIFT를 통한 물체 인식

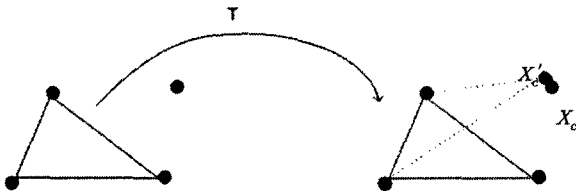
물체가 인식이 되면 SIFT를 통해 구해진 Keypoint 들을 이용하여 자세 추정 및 검증을 수행한다.

2.4 자세 추정

이미 구해진 3차원 DB를 통해 입력영상으로부터 물체를 인식하고 그 물체의 자세를 추정한다. [4] 물체 인식의 경우 많은 Keypoint들이 검출되며 또한 잘못된 Keypoint들이 매칭이 되어도 물체 인식에는 영향을 미치지 않는다. 하지만 Visual Servoing을 위한 자세추정은 Visual Servoing을 위해 거리를 두어야 하기 때문에 추출되는 Keypoint들의 수도 많지 않고 잘못된 Keypoint들이 매칭이 될 경우 자세 추정에 영향을 미친다. 또한 우리가 흔히 접할 수 있는 물체들을 보면 대부분 비슷한 무늬들로 디자인 되어있다. 이런 경우 앞면과 뒷면을 구분할 수 없다. 물체를 인식하는 경우에는 문제가 되지 않지만 3차원 물체의 자세를 추정하는 데는 문제가 된다.

3차원 정보를 이용하여 추출된 SIFT의 Keypoint 들을 다시 검증하여 3차원 자세 추정을 수행하는 방법을 제시한다. 3차원 정보들을 가지고 2차원 Keypoint들의 매칭을 검증한다.

2.4.1 3차원 정보 이용한 2차원 Keypoint 매칭 검증



〈그림 6〉 Keypoint 검증

추출된 Keypoint들을 통해 Transform을 구한 후 그 정보를 이용하여 다른 점들의 위치를 추정하여 실제 매칭된 Keypoint들을 검증한다. Transform을 구하기 위해선 최소 3점이 필요하다.[4] 처음 Transform을 구하기 위한 Keypoint들은 SIFT 매칭에서 가장 높은 매칭률을 가지는 Keypoint들을 이용하여 오차를 최소로 줄인다. 구해진 Transform을 이용하여 한 개의 Keypoint의 검증이 끝나면 다시 Transform을 구하고 그 정보를 이용하여 또 다른 Keypoint들을 검증하여 정확도를 높인다. 2차원 Keypoint를 3차원 정보를 이용하여 검증한다.

$$\begin{bmatrix} X'_c \\ Y'_c \\ Z'_c \end{bmatrix} = {}^m T_c^{-1} \times \begin{bmatrix} X_m \\ Y_m \\ Z_m \end{bmatrix} \quad (3)$$

${}^m T_c$: Camera to Model Transform
 X'_c : Estimated Camera Coordinate
 X_m : Model Coordinate

추정된 카메라 좌표(X'_c)와 실제 매칭된 카메라 좌표(X_c)를 비교하여 Keypoint를 검증한다. 두 점의 Correlation coefficient를 구하여 매칭의 정

확도를 판별한다. Correlation coefficient는 식 (4)에 의해서 정의 된다. [5]

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}_i)(x'_i - \bar{x}'_i)}{\sqrt{\sum_{i=1}^n (x_i - \bar{x}_i)^2 \sum_{i=1}^n (x'_i - \bar{x}'_i)^2}} \quad (4)$$

r 은 Correlation coefficient이며, 0과 1사이의 값을 가진다. n 은 매칭된 Keypoint의 수이다. Correlation coefficient의 값이 높을수록 매칭의 정확도가 높다고 판단할 수 있다. 3차원 정보를 이용해서 구해진 추정값을 통해 매칭된 2차원 입력값을 검증한다. Correlation coefficient의 값에 따라 매칭을 검증한다. <표 1>은 실제 10개의 Keypoint가 추출된 모델을 가지고 실험을 하였다. 먼저 올바르게 매칭된 5개의 Keypoint들을 가지고 Transform을 구한 후 나머지 Keypoint들을 추정하여 실제 매칭된 값과의 Correlation coefficient를 구하여 검증하는 과정을 보여준다.

X_c (입력값)		X'_c (추정값)		Correlation Coefficient	
X	Y	X	Y	X	Y
-0.069828	0.000049	-0.0697	0.0012	0.999914324	0.999886235
-0.014338	0.067914	-0.0139	0.0671	0.999945474	0.999891968
-0.013511	0.079082	-0.0132	0.0788	0.999944162	0.999927506
-0.044010	0.062421	-0.0440	0.0629	0.999922514	0.999947947
-0.035235	0.066463	-0.0350	0.0667	0.999926252	0.999943331

〈표 1〉 Correlation coefficient

<표 1>에서 알 수 있듯이 추정된 값들을 이용한 검증으로 매칭된 Keypoint들은 정확히 매칭이 되었다는 것을 알 수 있다.

3. 결 론

Visual Servoing을 위해서는 물체 인식이 먼저 수행 되어야하며 무엇보다 정확한 자세 추정이 필요하다. 이를 위해 매칭을 수행하고 3차원 정보를 이용해서 검증을 하여 정확도를 높였다. SIFT는 2차원 정보를 이용한다. 그리고 회전과 크기에 불변한 특성이 있기 때문에 3차원 모델에 이용하기에 부적합하다. 본 논문에서는 SIFT의 2차원 정보를 이용하여 3차원 모델의 자세를 추정하고 검증하는 방법을 제시했다. 3차원의 데이터를 사용하여 2차원의 데이터를 검증하여 2차원 데이터를 실제 3차원 물체에 적용했을 때의 오차를 최소화 하였다. SIFT는 각각의 Keypoint들만을 가지고 매칭을 수행하고 검증한다. 하지만 본 논문에서는 유사한 Keypoint들의 매칭 문제와 매칭 오차를 줄이기 위해 주위 Keypoint들간의 3차원 관계를 이용하여 검증하였다. Correlation coefficient를 통해 검증하여 추정된 Keypoint들과 실제 매칭된 Keypoint들이 올바르게 매칭 되었는지 확인할 수 있다.

〈참 고 문 헌〉

- [1] Lowe, D.G. , "Object recognition from local scale-invariant features", In International Conference on Computer Vision, Corfu, Greece, pp. 1150-1157., 1999.
- [2] ROGER Y. TSAI, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3. No. 4, August 1987.
- [3] William H. Press, Saul A. Teukolsky, William T. Vetterling, Brian P. Flannery Numerical Recipes, second edition pp. 417-424
- [4] Berthold K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions", Journal of the Optical Society of America A, Vol. 4, page 629, April 1987.
- [5] Ramesh Jain and Rangachar Kasturi and Brian G.Schunck, Machine Vision, McGRAW-HILL, pp.295-297, 1995.