

파노라마 비전센서를 이용한 영상 기반 2차원 로봇 자세 측정

심성대^o, 김종성, 홍기상

포항공과대학교 전자전기공학과

Vision-based Robot 2D Pose Estimation using a Panorama Vision Sensor

Sung-Dae Sim^o, Jong-Sung Kim, Ki-Sang Hong

Department of Electrical Engineering, Pohang University of Science and Technology

purur@postech.ac.kr, kimjs@postech.ac.kr, hongks@postech.ac.kr

요 약

최근 지능형 로봇에 대한 관심이 증가하고 있다. 이에 따라 지능형 로봇을 제어하기 위하여 로봇의 위치 및 방향을 측정하는 기술이 요구되고 있다. 본 논문에서는 파노라마 비전센서를 이용하여 실내환경에서 로봇의 2차원 자세를 측정하는 문제를 다루었다. 본 논문에서 제안하는 방법은 에피폴라 기하를 이용하여 적은 수의 기준 영상을 가지고 더 정확한 자세측정을 할 수 있다. 그리고 실내 환경에서 적절한 기준 영상을 선택할 수 있다. 또한 물체가 가려지는 경우에도 로봇의 자세를 잘 측정할 수 있음을 실험을 통하여 보였다.

1. 서 론

지능형 로봇은 우리 주위에서 많이 응용될 수 있기 때문에, 최근 지능형 로봇에 대한 관심이 증가하고 있다. 지능형 로봇이 실제 환경에서 활용되기 위해서는 로봇의 위치와 방향 등의 자세를 측정하는 기술이 필수적인 요소가 된다. 이와 관련하여 음향센서, 레이저 센서, 주행기록계 등을 이용한 자세 측정 기술에 대한 연구가 다양하게 이루어지고 있다.

최근 영상 기술의 발달로 비전 센서를 기반으로 한 주변의 환경 지도를 작성 및 로봇의 네비게이션에 대해 많은 연구가 진행되고 있다. 비전 센서를 이용한 방법은 비용이 다른 센서에 비해 비교적 적고 구조적으로 간단하게 구성되는 등의 다양한 이점을 가지게 된다. 비전센서를 사용한 로봇 시스템에서는 평면 영상을 얻을 수 있는 비전센서를 이용하는 방법이 있고, 특수한 파노라마(panorama) 비전센서나 영상 처리 기법에 의해 생성된 파노라마 영상을 활용하는 방법이 있다. 이러한 비전센서를 기반으로 한 로봇 자세 측정 연구는 국내·외에서 다양하게 연구가 진행되고 있다.

파노라마 비전센서를 사용한 로봇 자세 측정 기술은 다양한 방법으로 연구가 진행되어 왔다. 주변 환경의 360도의 시야를 동시에 바라볼 수 있는 특징은 로봇 자세 측정에서 파노라마 비전센서를 많이 사용하는 이유이다. 파노라마 비전센서를 사용하여 로봇의 위치를 측정하는 방법에는 크게 영상 간의 비교를 통하여 위치를 측정하는 방식[1,2]과

특징점들을 이용하여 측정하는 방식[3,4]으로 나눌 수 있다. 영상 간의 비교를 이용하는 방식은 상당히 많은 수의 기준 영상을 필요로 하게 된다. 그리고 영상 간의 비교를 이용하여 측정하게 됨으로써 자세 측정 오차가 크게 나타난다. 특징점을 이용하는 방식 중 이미 알고 있는 특징점을 이용[3]하는 방식은 별도의 특징점이 주어야하는 불편함이 있다.

본 논문에서는 기준 영상의 수를 상대적으로 많이 줄이고, 주어진 특징점이 아닌 실내 환경의 자연적 특징점을 이용하는 방법을 제안한다. 이를 위해서 영상간의 에피폴라 기하를 이용하고 SIFT[5] 특징점을 특징점 추출 및 매칭에 활용한다. 또한 실내 환경에서 적절한 기준 영상만을 사용하여 기준 영상의 수를 줄이는 방법을 제안한다. 위와 같이 본 논문에서 제시한 방법으로 앞선 연구보다 나은 결과를 볼 수 있다.

본 논문의 구성은 다음과 같다. 제 2 장에서는 제안된 로봇 자세 측정 기술에 대하여 설명한다. 제 3 장에서는 실내 환경에서 얻은 많은 수의 영상 중에서 적절한 기준영상을 선택하는 방법에 대하여 설명한다. 제 4 장에서는 앞서 제시된 방법을 이용한 자세 측정의 실험 결과를 살펴보고, 마지막으로 제 5 장에서는 결론과 앞으로의 연구 진행 방향에 대해서 살펴본다.

2. 로봇 자세 측정

본 장에서는 그림 1 과 같이 주어진 실내환경에서 몇 장의 기준영상들과 현재 위치에서 찍은 영상

사이의 관계를 파악하여 현재의 위치를 측정하는 방법에 대하여 설명한다. 본 논문에서 제안하는 로봇 자세 측정 시스템은 그림 2와 같다.

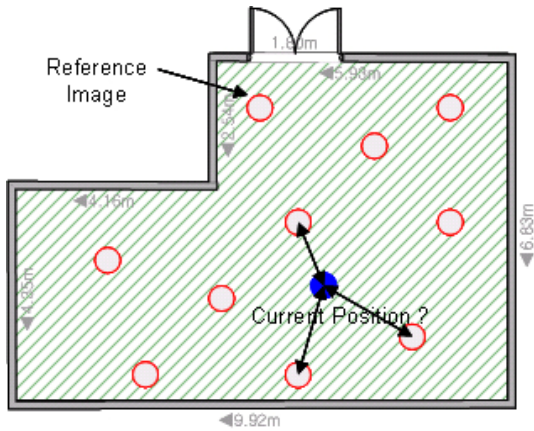


그림 1. 로봇 자세 측정

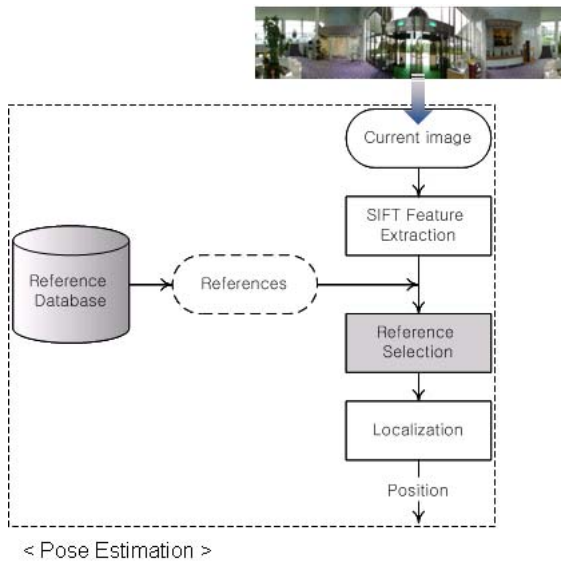


그림 2. 로봇 자세 측정 시스템

2.1 특징점 추출 및 매칭

파노라마 영상 사이에서 특징점의 매칭은 특징점들의 심한 크기 변화와 affine 변형 등에 의해서 일반적인 특징점 추출 및 매칭 방법으로는 어려운 일이 된다. 이를 해결하기 위해서 Scale-Invariant Feature Transform(SIFT)[5]를 사용한다. SIFT 특징점은 밝기 변화, 크기 변화, 회전 변화, affine 변형 등에 강인한 특성을 가진다.

SIFT 특징점은 다음과 같은 방법으로 추출된다. 먼저 영상을 scale-space 로 만들고 Difference-of-Gaussian(DoG) 함수를 이용하여 영상의 Scale-space 와 DoG 영상의 Scale-space 를 생성한다. 이 DoG 영상의 Scale-space 에서 특징점이 될 수 있는 위치를 추출한다. 특징점이 추출되면 특징점으로 추출된 위

치에서 주변의 픽셀의 gradient 의 크기와 방향을 가우시안(Gaussian) 가중치를 주어 히스토그램(histogram)의 형태로 만든다. 이를 1 차원의 벡터 형태로 표현한다.

특징점의 매칭은 앞서 SIFT 알고리즘을 통해서 얻어진 특징점 벡터들을 서로 비교하여 feature space 에서 유클리디안 거리가 가장 가까운 것으로 매칭한다. 이 때 가장 가까운 거리가 두 번째 가까운 특징점과의 거리보다 일정 비율 이상 작아야 한다.

이러한 방법으로 특징점을 매칭하는 경우에 잘못 매칭되는 경우도 발생하게 된다. 이러한 점들은 대부분의 경우, 에피폴라 기하를 계산할 때 이상값(outlier)으로 RANSAC[9]을 통해서 제거 될 수 있다. 또한 본 논문에서는 잘못 매칭된 점들을 제거하기 위해서 2 차원 모션이라고 가정하고 있기 때문에 그림 3. (a)에서 보는 것과 같이 매칭된 특징점들 사이의 회전이 심한 것을 제거하고 그림 3. (b)에서 보는 것 처럼 카메라의 수직 중앙부분을 기준으로 위쪽은 위쪽, 아래쪽은 아래쪽으로 매칭이 되지 않은 매칭점들을 제거한다.

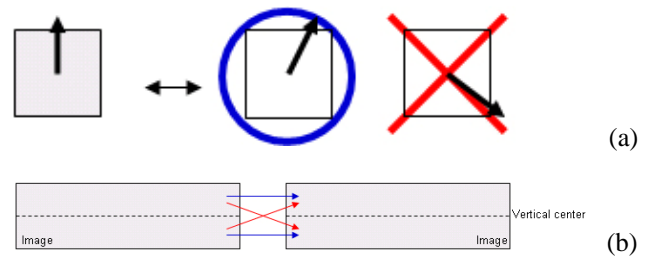


그림 3. 잘못된 매칭점 제거 방법

2.2 에피폴라 기하 계산

에피폴라 기하는 두 영상 사이의 관계를 파악하는 기본적인 기하이다. 로봇의 자세를 측정하기 위해서 기준 영상과 현재 영상 사이의 에피폴라 기하를 계산한다. 파노라마 영상 사이에서는 이센셜 행렬(essential matrix)을 계산한다. 이센셜 행렬 E 는 회전행렬 R 과 이동 벡터 t 에 의해서 다음과 같이 표현 될 수 있다.

$$E = [t]_x R \tag{1}$$

이센셜 행렬은 3x3 의 행렬로 표현 되는데 식(2) 과 같이 두 영상에서 매칭점의 실린더 영상 좌표 상의 3 차원 좌표를 활용하여 이센셜 행렬을 계산한다. 이 때, 특징점들 중에서 이상값(outlier)을 RANSAC 알고리즘을 이용하여 제거한다.

$$\hat{u}^T E \hat{u} = (\hat{u}', \hat{v}', \hat{w}') E \begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{w} \end{pmatrix} = 0 \tag{2}$$

여기서 \hat{u}, \hat{u}' 은 정규화된 실린더 영상 좌표 상의 3 차원 좌표이다. 좌표는 초점거리 f 와 실린더 좌표를 가지고 식(3)과 같이 표현할 수 있다.

$$\mathbf{u} = (f \cos \phi, f \sin \phi, z) \rightarrow \hat{\mathbf{u}} = \mathbf{u} / |\mathbf{u}| \quad (3)$$

이센셜 행렬의 특징은 두 개의 특이값(singular value)가 같고, 다른 하나의 특이값이 0 인 특징을 가지고 있다. 따라서 실제 이센셜 행렬에 잘 맞추어 주기 위해서 식(2)과 같이 8 점 알고리즘[6]으로 구한 이센셜 행렬 $\hat{\mathbf{E}}$ 을 식(4)와 같이 SVD 를 적용하여 특이값 분해를 한다.

$$\hat{\mathbf{E}} = \mathbf{U} \text{diag}(\sigma_1, \sigma_2, \sigma_3) \mathbf{V}^T \quad (4)$$

분해된 이센셜 행렬을 이센셜 행렬의 특이값 특징에 맞추어 식(5)와 같이 이센셜 행렬을 추정한다 [7].

$$\mathbf{E} = \mathbf{U} \text{diag}((\sigma_1 + \sigma_2) / 2, (\sigma_1 + \sigma_2) / 2, 0) \mathbf{V} \quad (5)$$

2.3 이센셜 행렬 분해

이센셜 행렬은 이동벡터과 회전행렬에 의해서 정의 될 수 있다. 그리고 2 차원 모션이라고 가정하여 x,y 방향으로의 움직임과 z 축의 회전만 있다고 보면 식(6)과 같이 표현될 수 있다.

$$\mathbf{t} = [\cos \phi \quad \sin \phi \quad 0]^T, \mathbf{R} = \begin{bmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (6)$$

2 차원 모션을 가지는 이동벡터와 회전행렬을 조합하여 이센셜 행렬은 식(7)과 같이 표현된다.

$$\mathbf{E} = [\mathbf{t}]_x \mathbf{R} = \begin{bmatrix} 0 & 0 & \sin \phi \\ 0 & 0 & -\cos \phi \\ \sin(\theta - \phi) & \cos(\theta - \phi) & 0 \end{bmatrix} \quad (7)$$

따라서 이센셜 행렬에서 θ 와 ϕ 를 $\theta = \text{atan2}(e_3 / -e_4)$ 와 $\phi = \theta - \text{atan2}(e_7 / e_8)$ 로 구하여 이동벡터 \mathbf{t} 와 회전행렬 \mathbf{R} 을 결정한다.

2.4 자세 측정

앞서 구한 회전 행렬과 이동벡터를 가지고 로봇의 자세를 측정한다. 이를 위해서는 그림 4 와 같이 3 장의 기준 영상이 필요하다.

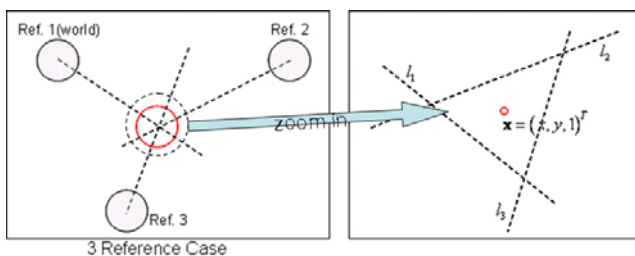


그림 4. 3 장의 기준 영상으로부터 자세 측정 방법

그림 4.(a)에서 보면 3 장의 기준 영상으로부터 현재 영상의 위치는 각각의 이동 벡터와 회전 행렬에 의해서 각 기준 영상의 좌표를 기준으로 $-\mathbf{R}^T \mathbf{t}$ 의 관계를 가진다. 이렇게 측정된 현재 영상의 위치와 기준 영상의 위치를 각각 한 점으로 보고 2 차원

평면 상의 직선벡터 l_1, l_2, l_3 를 구한다. 구해진 세 직선의 교점이 현재 영상의 위치라고 볼 수 있다. 현재 영상의 위치를 식 (8)에서 \mathbf{x} 의 해로 결정된다.

$$\begin{cases} l_1^T \mathbf{x} = 0 \\ l_2^T \mathbf{x} = 0 \\ l_3^T \mathbf{x} = 0 \end{cases} \Rightarrow \mathbf{L} \mathbf{x} = \mathbf{0} \quad \text{where } \mathbf{L} = \begin{bmatrix} l_1^T \\ l_2^T \\ l_3^T \end{bmatrix} \quad (8)$$

로봇의 위치는 위의 방법으로 측정하고 로봇의 방향은 각각의 회전행렬 $\mathbf{R}_1, \mathbf{R}_2, \mathbf{R}_3$ 로부터 각각의 각을 구하여 기준 좌표를 기준으로 한 각으로 변환한다. 이 변환된 세 각을 평균하여 기준 좌표계에 대한 로봇의 방향을 측정한다.

2.5 기준 영상 선택

실내 환경에서 주어진 많은 수의 기준 영상 중에서 현재 영상과 관계가 깊은 영상들을 세개 선택하여야 한다. 영상이 비슷하면 할수록, 즉, 영상간의 거리가 가까울수록 매칭이 이루어지는 SIFT 특징점 수는 늘어난다. 이를 그림 5 에서 확인할 수 있다.

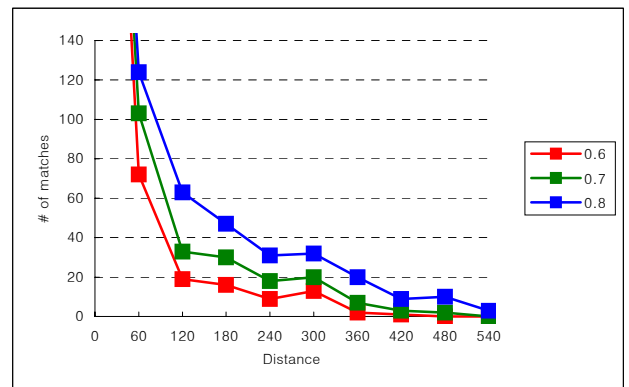


그림 5. 거리에 따른 SIFT 특징점 매칭 수

그림 5 에서 볼 수 있는 특징을 이용하여 매칭점들이 많이 뿔히는 기준 영상을 선택한다. 이러한 방법은 그림 6 의 순서도에서 확인할 수 있다.

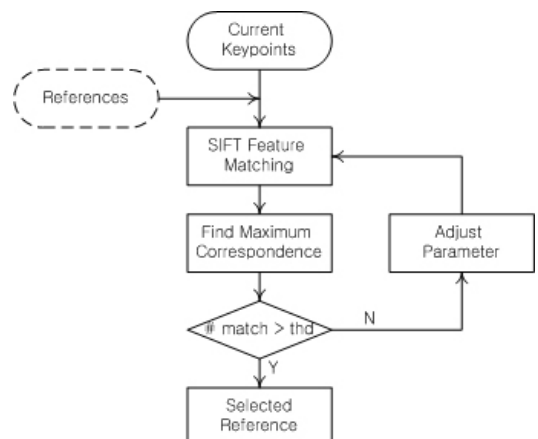


그림 6. 기준 영상 선택 알고리즘

3. 기준 영상 데이터 베이스 작성

실내환경에서 기준영상으로 활용하기 위하여 여러 위치에서 파노라마 영상을 얻는다. 하지만 이러한 파노라마 영상들을 모두 사용하기는 데이터 양이 많아지게 된다. 따라서 많은 수의 파노라마 영상 중에서 기준 영상으로 적합한 영상만을 선택하여 기준 영상 데이터 베이스를 만드는 과정이 필요하다. 적절한 기준 영상을 선택하기 위해 몇가지의 조건을 둔다.

3.1 기준 영상 채택을 위한 검색 영역 설정

먼저 기준 영상을 선택하는 방법은 전체 환경에서 동시에 진행하는 것이 아니라 하나의 주어진 시작점으로부터 연속적으로 진행해 나가면서 기준 영상을 선택한다.

다음의 기준 영상을 찾기 위한 검색 영역은 현재 영상에서 일정 거리의 범위 내로 제한한다. 또한 이전 기준 영상에서 현재 기준 영상으로의 방향에서 다음 번 기준 영상을 찾기 위해서 지그재그형태를 가지기 위해서 좌우 일정 각도의 범위 내로 제한한다. 이렇게 하면 기준 영상이 일직선상에 위치하는 경우를 막을 수 있다. 기준 영상을 일직선 상에 두지 않는 이유는 세장의 기준 영상을 쓰더라도 세장의 기준 영상이 일직선 상에 있을 때, 현재 영상이 그 일직선 상에 놓이게 되면 거리를 측정할 수가 없기 때문이다.

적절한 기준 영상을 검색하기 위한 영역은 그림 7과 같은 형태를 가지게 된다.

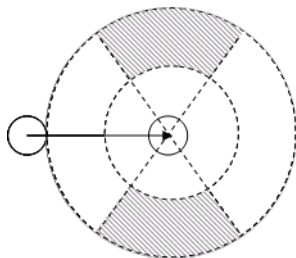


그림 7. 기준 영상 검색 영역

3.2 기준 영상 채택

많은 영상 중에서 기준 영상을 채택하기 위해서 앞 절의 검색 영역을 먼저 정하게 된다. 이러한 검색 영역 내에서 많은 후보 영상이 있을 수 있다. 이 중 하나를 선택하기 위해서 SIFT 특징점의 매칭 수가 일정 범위 내에 있는 것 중에서 기존의 영상과의 거리가 먼 것을 선택한다. 거리가 먼 것을 선택하면 전체 환경에서 기준 영상의 수를 상대적으로 줄일 수 있게 된다.

그리고 그림 7에서 보듯이 이전 방향에서 좌측과 우측 두 개의 검색 영역이 존재하게 되는데 둘 중에서 하나만 선택하면 기준 영상이 한쪽으로 치

우치는 현상이 나타난다. 따라서 양쪽 모두 가지기 위해서 이진 트리(binary tree)형태로 기준 영상을 저장해 나간다.

일단 기준 영상이 선택되면 그 주변의 일정영역 안의 영상들은 기준 영상의 후보에서 제거한다.

4. 실험 결과

본 논문에는 ㈜일리시스에서 개발된 파노라마 카메라인 PanoVix IPC-1101을 사용하였다. 실험 환경은 본 연구실의 환경에서 실험을 하였다.

4.1 특징점 매칭 결과

SIFT 특징점을 이용하여 특징점을 매칭한 결과는 그림 8과 같다. 그림 8에서 나타난 숫자로 같은 특징점임을 확인할 수 있다. 그림 8에서 보면 특징점들의 크기 변화와 보는 각도가 약간 달라졌음에도 불구하고 매칭이 잘 이루어지는 것을 확인할 수 있다.



그림 8. 특징점 매칭 결과

4.2 실내 환경에서 로봇 자세 측정

약 3m X 3m 크기의 사각 형태의 방에서 세장의 기준 영상을 가지고 실험을 하였을 때 41개의 측정 위치에서 위치와 방향의 평균 오차는 5.9cm 와 1.6°였다. 그림 9는 실험 환경과 실험 결과를 나타낸 그림이다.

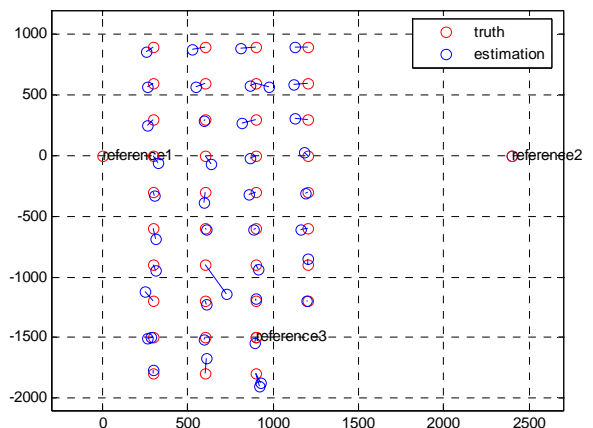


그림 9. 세장의 기준 영상을 이용한 로봇 자세 측정

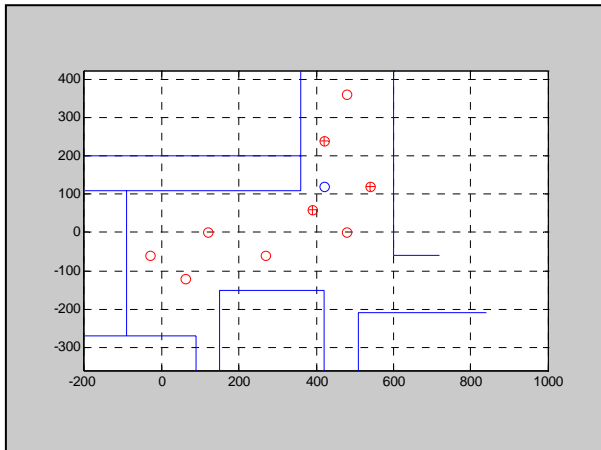
앞의 그림 9와 같은 환경에서 로봇 자세를 측정 한 결과는 표 1에서 볼 수 있다.

표 1. 로봇 자세 측정 오차 및 평균 계산 시간

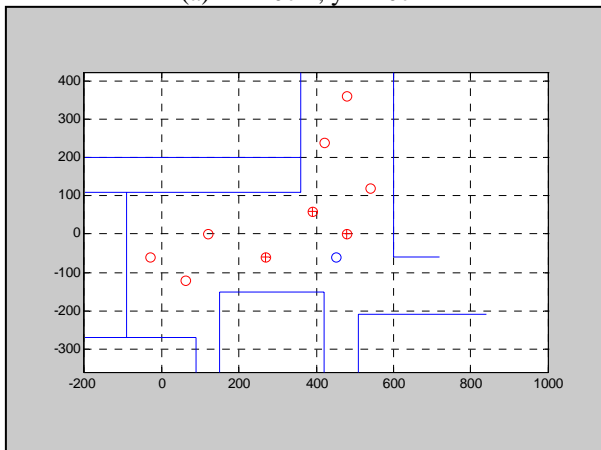
Position Error(mm)	58.9879
Orientation Error(degree)	1.6289
Total time(sec)	2.234

4.3 기준 영상 선택 결과

현재 영상의 위치를 알기 위해서 필요한 기준 영상을 선택해야 한다. 7m X 6m 크기의 실험실에서 9개의 기준 영상을 임의로 주었을 때 현재 영상에서 SIFT 특징점 매칭을 통해서 기준 영상을 선택한 결과는 그림 10에서 볼 수 있다. 그림 10에서 보듯이 기준 영상 선택이 잘 이루어짐을 알 수 있다.



(a) x=420cm, y=120cm



(b) x=450cm, y=-60cm

그림 10. 현재 영상 자세 측정을 위한 기준 영상 선택 결과 (O : reference image, O : current image, ⊕ : selected reference image)

선택된 세개의 기준 영상을 바탕으로 로봇의 자세를 측정한 결과 그림 10.(a)의 경우에는 1.5cm의 위치오차와 0.1°의 방향 오차를 나타내었다. 그림

10.(b)의 경우에는 4.2cm의 위치오차와 0.4°의 방향 오차를 나타내었다. 따라서 선택된 기준 영상으로부터 로봇의 자세 측정이 잘 이루어짐을 볼 수 있다.

4.4 기준 영상 데이터 베이스 작성 결과

앞서 3장에서 소개한 기준 영상 작성 방법을 이용하여 주어진 실내 환경에서 얻어진 많은 영상 중에서 적절한 기준 영상을 선택하였다. 실험은 7m X 6m 크기의 실험실에서 54개의 주어진 영상을 이용하였다. 기준 영상 데이터 베이스를 작성한 결과는 그림 11에 나타나 있다.

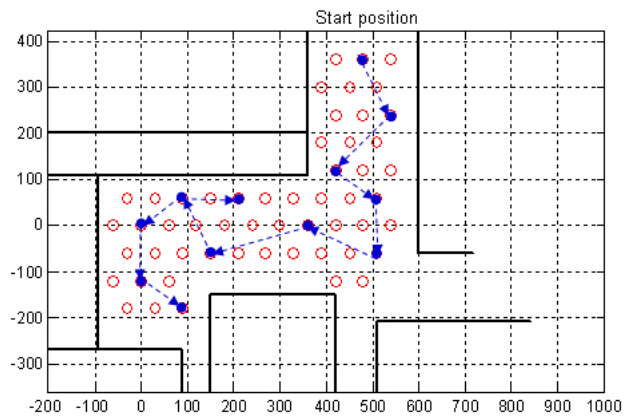


그림 11. 적절한 기준 영상 채택 결과

그림 11에서 처음에 주어진 시작점에서 출발하여 연속적으로 기준 영상이 선택되어 가는 과정을 볼 수 있다. 기준 영상을 작성할 때 이진 트리의 형태를 가지게 되므로 좌,우의 방향으로 동시에 진행해 나가게 된다.

이 실험에서 54개의 영상 중에서 12개의 영상만이 기준 영상으로 선택이 되었다. 12개의 영상을 기준 영상 데이터 베이스로 두고 실내환경에서 현재 영상의 자세를 측정하였을 때, 위치와 방향의 각각 평균 오차는 6cm와 0.3°였다.

이 같은 자세 측정 결과를 관련된 다른 논문들과 비교하면 제시된 방법을 사용하여 좀 더 정확하게 자세를 측정할 수 있었다. 결과의 비교는 표 2에 나타나 있다.

표 2. 로봇 자세 측정 오차 비교

Method	Position Error(cm)	Orientation Error(°)	Test Environment
Proposed	6	0.3	7m X 6m
[1]	14	.	4.5m X 5m
[2]	13	.	4.5m X 5m
[3]-A	9	.	4.5m X 7.9m
[3]-B	15	.	4.5m X 7.9m
[4]	25	3.5	6m X 9m

4.5 교합(occlusion)이 있는 경우 자세 측정 실험

실내환경에서 지능형 로봇 혼자서 돌아다니는 경우에는 교합이 일어날 일이 없으나, 지능형 로봇과 함께 사람이나 다른 로봇이 같이 있게 되면 교합에 의하여 환경 정보를 가리게 된다. 이러한 경우에 대하여 영상에 교합영역을 설정하여 실험을 하였다.



(a) 교합 영역이 없는 영상



(b) 50%의 교합 영역이 있는 영상
그림 12. 교합 영역이 있는 영상

그림 12는 교합 영역이 없는 영상과 50%의 교합 영역이 있는 영상을 비교한 것이다. 이러한 영상에서의 실험은 표 3에서 교합 영역이 있더라도 위치 측정이 대부분 잘 이루어지는 것을 확인하였다.

표 3. 교합 영역이 있는 영상에서 자세 측정 오차

Occlusion	Position Error(cm)	Orientation Error(°)
0%	5.3	0.16
10%	8.7	1.18
20%	6.6	0.38
30%	15.3	0.59
40%	4.1	0.02
50%	1.3	0.10

5. 결론

본 논문에서는 실내 환경에서 파노라마 비전센서를 사용하여 로봇의 자세를 측정하는 방법을 제안하였다. 제안된 방법은 적절한 기준 영상을 선택하고 현재 영상과의 에피폴라 기하를 이용하여 더욱 정확한 자세 측정 결과를 갖도록 하였다. 또한, SIFT 특징점을 이용하여 에피폴라 기하를 적용하는데 어려운 문제점이 되었던 특징점 매칭 문제를 해결하였다. 그리고 실내환경에서 기준영상의 수를 적게 가지도록 적절한 기준 영상을 자동으로 선택하도록 하였다. 또한, 교합 영역이 존재하는 영상에 대해서도 로봇의 자세 측정이 잘 이루어졌다.

현재는 주어진 영상에 대해서 실험을 하였으나 앞으로 자동적으로 영상들을 얻고 적절한 기준 영

상을 추출하는 방법에 대한 부가적인 연구가 필요하다.

6. 참고문헌

- [1] M. Artac, M. Jogan and A. Leonardis, "Mobile Robot Localization Using an Incremental Eigenspace Model", ICRA 2002
- [2] M. Jogan and A. Leonardis, "Robust Localization Using Panoramic View-Based Recognition", ICPR 2000
- [3] Mark Fiala, "Linear Markers for Robot Navigation with Panoramic Vision", CRV 2004
- [4] David C.K. Yuen and Bruce A. MacDonald, "Natural Landmark based localization system using panoramic images", ICRA 2002
- [5] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", IJCV 2004
- [6] S.B. Kang and R. Szeliski, "3-D Scene Data Recovery using Omnidirectional Multibaseline Stereo", Cambridge Research Lab Technical Report 1995
- [7] Roland Bunschoten, "Mapping and Localization from a Panoramic Vision Sensor", Thesis 2003
- [8] R. Hartley and A. Zisserman, "Multiple View Geometry", Cambridge Univ. press, page238-241, 2000
- [9] M. A. Fischler, R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", Comm. of the ACM, Vol 24, pp 381-395, 1981