

# 3차원 객체 복원을 위한 SIFT 특징점 가중치 기반 반복적 점군 정합 방법

신동원 호요성

광주과학기술원 영상통신연구실

{dongwonshin, hoyo}@gist.ac.kr

## SIFT Weighting Based Iterative Closest Points Method

in 3D Object Reconstruction

Dong-Won Shin and Yo-Sung Ho

Gwangju Institute of Science and Technology (GIST)

### 요약

최근 실세계에 존재하는 물체의 3차원 형상과 색상을 디지털화하는 3차원 객체 복원에 대한 관심이 날로 증가하고 있다. 3차원 객체 복원은 영상 획득, 영상 보정, 점군 획득, 반복적 점군 정합, 무리 조정, 3차원 모델 표현과 같은 단계를 거쳐 통합된 3차원 모델을 생성한다. 그 중 반복적 점군 정합 방법은 카메라 궤적의 초기 값을 획득하는 방법으로서 무리 조정 단계에서 전역 최적 값으로의 수렴을 보장하기 위해 중요한 단계이다. 기존의 반복적 점군 정합 (iterative closest points) 방법에서는 시간이 지남에 따라 누적된 궤적 오차 때문에 발생하는 객체 표류 문제가 발생한다. 본 논문에서는 이 문제를 해결하기 위해 색상 영상에서 SIFT 특징점을 획득하고 3차원 점군을 얻은 뒤 가중치를 부여함으로써 점군 간의 더 정확한 정합을 수행한다. 실험결과에서 기존의 방법과 비교하여 제안하는 방법이 절대 궤적 오차 (absolute trajectory error)가 감소하는 것을 확인했고 복원된 3차원 모델에서 객체 표류 현상이 줄어드는 것을 확인했다.

### 1. 서론

3차원 복원 기술은 실제 세계에 존재하는 물체의 3차원 형상과 색상을 디지털화하는 기술로서 가상 현실, 증강 현실, 인터넷 쇼핑, 의료 영상, 공장 자동화 시뮬레이터 등 다양한 응용에서 널리 사용될 수 있다. 3차원 복원 방법은 크게 수동형과 능동형의 방식으로 분류할 수 있다. 수동형 방식은 다시점에서 촬영된 색상 영상을 이용하여 촬영한 영상의 명암이나 시차 등을 이용하는 방식으로 비용이 적고 객체에 대한 텍스처를 바로 획득할 수 있다. 한편, 능동형 방식은 특정한 적외선 패턴이나 초음파 등을 물체에 투사하여 그 변화량을 측정함으로써 3차원 형상을 복원하는 방법으로 수동형 방식에 비해 더 정확한 값을 계산해 낸다 [1].

본 논문에서는 RGB-D 카메라를 이용하여 수동형과 능동형 방식을 혼합하여 사용하는 3차원 복원 방법에 대해 중점적으로 설명한다. 그림 1은 일반적인 3차원 복원 방법의 흐름도를 나타낸다.



그림 1 일반적인 3차원 복원 방법의 흐름도

먼저 RGB-D 카메라로부터 색상 및 깊이 영상을 획득한 뒤 카메라 내부 파라미터를 이용하여 3차원 점군을 획득한다. 획득한 점군을 반복적 점군 정합 (iterative closest points) 방법을 이용하여 시점 간 카메라 외부 파라미터의 초기 값을 획득하고 무리 조정 (bundle adjustment) 단계에서 앞서 획득한 카메라 외부 파라미터를 전역적으로 보정한다. 마지막으로 3차원 모델의 점군 데이터를 TSDF (truncated signed distance function)와 같은 효율적인 방법으로 표현하는 단계를 거치면 완성된 3차원 모델을 획득할 수 있다 [2]. 이 과정에서 복원된 3차원 객체의 정확성을 높이기 위해서는 반복적 점군 정합 단계에서 카메라 외부 파라미터의 초기 값을 정확하게 계산하는 것이 중요하다.

기존의 반복적 점군 정합 방법으로는 점군 간의 Euclidean 거리를 이용하는 point-to-point 방법과 법선 벡터를 이용하는 point-to-plane 방법이 있다 [4][5]. 하지만 기존의 방법은 모두 시간이 지남에 따라 카메라 궤적에 대한 오차가 누적되어 객체가 표류하는 문제를 일으킨다.

따라서 본 논문에서는 누적 카메라 오차에 의한 객체 표류 문제를 해결하기 위해 색상 영상에서 SIFT 특징점을 얻고 반복적 점군 정합 단계에서 특징점에 가중치를 부여하여 객체 정합 능력을 높였다.

## 2. 제안하는 반복적 점군 정합 방법

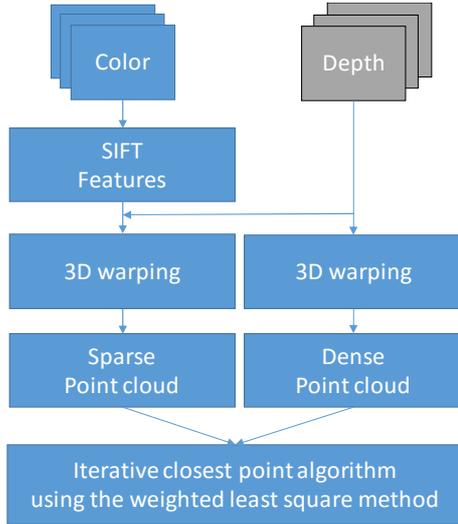


그림 2. 제안하는 방법의 흐름도

그림 2는 제안하는 방법의 흐름도를 보여준다. 먼저 RGB-D 카메라로부터 연속적으로 색상 영상과 깊이 영상을 획득한다.  $t$  시간의 색상 영상에 대해서 SIFT 알고리즘을 이용하여 영상의 특징 점을 검출한 뒤, 2차원의 특징 점들을 카메라 내부 파라미터와 깊이 정보를 이용하여 3차원의 점군  $p_s^t$ 를 얻는다 [3]. 깊이 영상에 대해서도 유사하게 영상 내의 깊이 값이 0이 아닌 모든 2차원 점에 대해 카메라 내부 파라미터와 깊이 정보를 이용하여 3차원의 점군  $p_d^t$ 를 얻는다. 그 후  $(t+1)$  시간의 색상 영상과 깊이 영상에 대해서 동일하게 점군  $p_s^{(t+1)}$ 과  $p_d^{(t+1)}$ 을 얻는다.  $t$  시간의 점군과  $(t+1)$  시간의 점군 사이의 회전 및 이동 변환 관계를 얻고자  $p_s^t$ 와  $p_s^{(t+1)}$  사이의 매칭 쌍에 대해서 추가적인 가중치를 부여하고 비용함수를 정의한 뒤 가중 최소 자승법을 이용하여 최적의 값을 얻는다.

### 3.1 특징점에 가중치가 적용된 point-to-point 방법

본 절에서는 기존의 point-to-point 방법에서 색상 영상으로부터 얻은 특징점에 가중치를 부여하여 점 군 간의 회전 및 이동 변환 관계를 얻는 방법을 유도한다. point-to-point 방법에서는 먼저 두 점군의 중점을 일치 시킨 후 식 (1)의 비용 함수  $E$ 를 최소화하는 회전 변환 파라미터  $R$ 를 얻고자 한다.

$$E = \sum_{i=1}^N w_i \| p_i^t - R p_i^{(t+1)} \|^2 \quad (1)$$

여기서  $p_i^t$ 는  $t$  시간의 색상 및 깊이 영상으로부터 얻은 점군에서  $i$ 번째 점의 좌표를 나타낸다. 식 (1)을 일부 전개하면 식 (2)를 얻는다.

$$E = \sum_{i=1}^N w_i (p_i^t p_i^t + p_i^{(t+1)} p_i^{(t+1)} - 2 p_i^t R p_i^{(t+1)}) \quad (2)$$

식 (2)를 최소화 하는 것은 식 (3)을 최대화 하는 것과 같다 [4].

$$\begin{aligned} F &= \sum_{i=1}^N w_i (2 p_i^t R p_i^{(t+1)}) \\ &= \text{Trace} \left( \sum_{i=1}^N w_i p_i^t R p_i^{(t+1)} \right) \\ &= \text{Trace} (RH) \end{aligned} \quad (3)$$

where  $H = \sum_{i=1}^N w_i p_i^t p_i^{(t+1)}$

식 (3)을 최대화 하는 행렬  $R$ 은 행렬  $H$ 를 특이 값 분해한 뒤 획득한 행렬들의 곱으로 계산할 수 있다.

$$\begin{aligned} H &= U D V^t \\ R^* &= V U^t \end{aligned} \quad (4)$$

이동 변환  $t$ 는 식 (5)를 이용해 간단하게 얻을 수 있다.

$$\begin{aligned} T^* &= p_c^t - R^* p_c^{(t+1)} \\ \text{where } p_c^t &= \frac{1}{N} \sum_{i=1}^N p_i^t, p_c^{(t+1)} = \frac{1}{N} \sum_{i=1}^N p_i^{(t+1)} \end{aligned} \quad (5)$$

여기서  $p_c^t$ 와  $p_c^{(t+1)}$ 는 각각  $t$ 와  $(t+1)$  시간에 얻은 점군의 중점을 의미한다.

### 3.2 특징점에 가중치가 적용된 point-to-plane 방법

본 절에서는 기존의 point-to-plane 방법에 가중 최소 자승법을 이용하여 점군 간의 회전 및 이동 변환 관계를 얻는 방법을 설명한다. point-to-plane 방법에서는 식 (6)의 비용 함수  $E$ 를 최소화하는 회전 및 이동 변환 파라미터의 변수가 포함된 벡터  $x$ 를 얻고자 한다 [5].

$$E = \sum_{i=1}^N w_i \| A_i x - b_i \|^2 \quad (6)$$

where

$$\begin{aligned} A_i &= \begin{bmatrix} n_{iz} p_{iy}^t - n_{iy} p_{iz}^t, \\ n_{ix} p_{iz}^t - n_{iz} p_{ix}^t, \\ n_{iy} p_{ix}^t - n_{ix} p_{iy}^t, \\ n_{ix}, \\ n_{iy}, \\ n_{iz} \end{bmatrix}, \\ b_i &= \begin{bmatrix} n_{ix} p_{ix}^{(t+1)} + n_{iy} p_{iy}^{(t+1)} + n_{iz} p_{iz}^{(t+1)}, \\ -n_{ix} p_{ix}^t - n_{iy} p_{iy}^t - n_{iz} p_{iz}^t \end{bmatrix}, \\ x &= [\alpha, \beta, \gamma, t_x, t_y, t_z] \end{aligned}$$

여기서  $n$ 은 법선 벡터를 나타내고  $\alpha, \beta, \gamma$ 는 각각  $x, y, z$ 축에 대한 회전 각도를 나타낸다.

식 (6)을 행렬 표현으로 나타내면 식 (7)과 같다.

$$E = \sum_{i=1}^N w_i r_i^2 = r^T W r = \| W^{1/2} r \|^2 \quad (7)$$

이를 미분하여 정리하면 식 (8)과 같다.

$$\frac{\partial E}{\partial x} = (b - Ax)' (W + W') (-A) = -2(b - Ax)' W A \quad (8)$$

식 (8)이 0이 되도록 하는 벡터  $x^*$ 의 값을 구하면 식 (9)와 같이 정리된다.

$$x^* = (A' W A)^{-1} A' W b \quad (9)$$

### 3. 실험결과

본 논문에서는 제안하는 방법의 정확성을 평가하기 위해 실제의 카메라 궤적과 제안하는 방법으로 예측된 카메라 궤적 사이의 거리를 비교하는 절대 궤적 오차 (absolute trajectory error, ATE)를 사용했다 [6]. 실험 시퀀스는 뮌헨공대에서 제공하는 RGB-D 영상 시퀀스 중 teddy 시퀀스를 사용했다 [7]. Teddy 시퀀스에서 0-1900 프레임을 실험에 사용했고 배경 제거 문턱치 값은 8000을 사용했다. 배경 제거 문턱치 값은 깊이 영상에서 객체에 대해서만 3차원 복원을 수행하기 위해 배경을 제거하는데 사용하는 문턱치 값이 낮을 의미한다. 이 실험 데이터에는 Microsoft Kinect v1에서 촬영된 색상 및 깊이 영상들과 각 장면에 대한 카메라의 회전 및 이동 변환 파라미터가 포함되어 있어 절대 궤적 오차를 비교할 수 있다.

본 실험에서는 이 영상 시퀀스에 대해 100프레임 단위로 반복적 점군 정합 방법들을 수행하여 얻은 절대 궤적 오차 값을 비교했다. 100프레임 이상의 영상을 사용하는 것은 반복적 점군 정합 방법 보다 무리 조정 방법을 사용하여 3차원 복원을 수행하는 것이 적절하기 때문이다.

제안하는 방법의 점군 정합 방법에서는 색상 영상으로부터 얻은 특징 점에 부여하는 가중치의 크기가 정확성에 큰 영향을 미친다. 따라서 가중치 공간을 미리 정해놓고 가중치 공간 내의 가능한 가중치를 모두 적용하여 최소의 절대 궤적 오차를 산출한 뒤 기존의 방법에서 얻은 절대 궤적 오차와 비교했다. 본 실험에서 정한 가중치 공간 내의 가중치 값은 [0.01, 0.1, 10, 100]의 값을 가지며 SIFT 특징점의 수를 조정하는 GMT (good matches threshold)는 [2, 5, 10]의 값을 가진다.

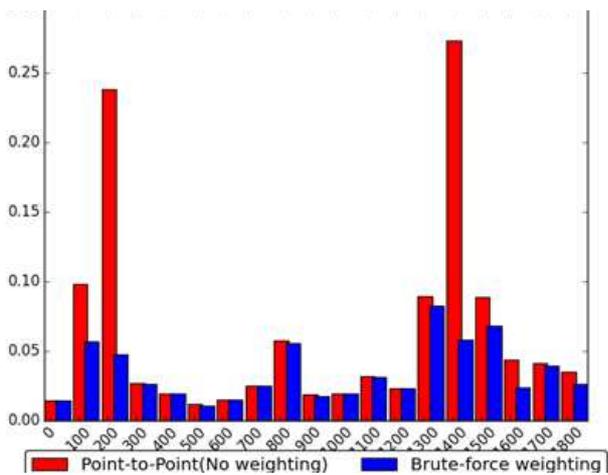


그림 3. point-to-point 방법에 대한 절대 궤적 오차

그림 3은 기존의 point-to-point 방법과 특징 점에 가중치를 적용한 point-to-point 방법의 절대 궤적 오차 결과를 구간별로 비교한 그래프이다. 대부분의 구간에서 제안하는 방법의 오차가 줄어들거나 비슷한 것을 확인할 수 있으며 특히 100-200, 200-300, 1400-1500 등의 구간에서 확연하게 차이를 보이는 것을 알 수 있다.

그림 4는 기존의 point-to-plane 방법과 특징 점에 가중치를 적용한 point-to-plane 방법의 절대 궤적 오차 결과를 구간별로 비교한 그래프이다. 이 결과 또한 대부분의 구간에서 제안하는 방법의 오차가 줄어들거나 비슷한 것을 확인할 수 있으며 특히 0-100, 200-300, 1700-1800, 1800-1900 등의 구간에서 확연하게 차이를 보이는 것을 알 수 있다.

표 1은 제안하는 방법에서 각 구간별 최소의 절대 궤적 오차를 보이는 가중치와 GMT의 값을 나타낸 것이다. 표를 분석해본 결과 기존의 방법과 비교하여 절대 궤적 오차가 큰 차이를 보이는 구간에서는 제안하는 방법이 높은 가중치를 가졌을 때 오차가 감소되는 것을 알 수 있었다. 또한 절대 궤적 오차가 기존의 방법과 비슷한 구간은 제안하는 방법이 낮은 가중치를 가진다는 것을 알 수 있었다. 이 분석이 의미하는 바는 구간 내 영상의 특성에 따라 가중치가 적절하게 선택되어야 하며 그에 따라 적절한 가중치를 선택하면 적어도 기존의 방법보다 절대 궤적 오차가 나쁜 결과를 보이지는 않는다는 점이다.

이 실험 결과를 바탕으로 무차별 가중치 대입을 통하여 가중치를 정하는 것이 아닌 구간 내 색상과 깊이 영상의 특성에 따라 적응적으로 최적의 가중치를 구하는 것을 추후에 연구하고자 한다.

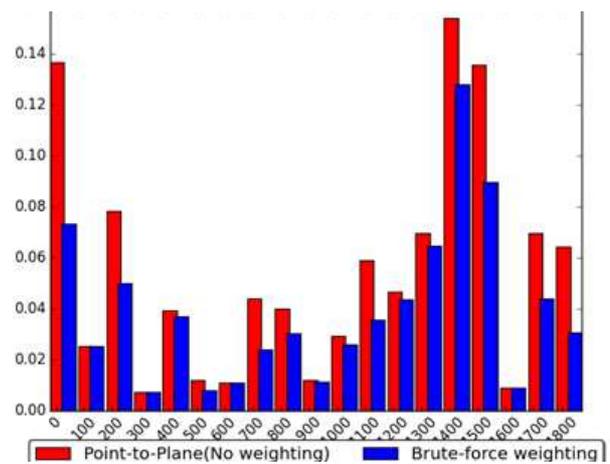


그림 4. point-to-plane 방법에 대한 절대 궤적 오차

표 1 제안하는 방법에서 각 구간별 최소의 절대 궤적 오차를 보이는 가중치와 GMT 값

	Section	0-	100-	200-	300-	400-	500-	600-	700-	800-	900-	1000-	1100-	1200-	1300-	1400-	1500-	1600-	1700-	1800-	1900-
		100	200	300	400	500	600	700	800	900	1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	
Weighted Point-to-Point	Weight	0.1	<b>100</b>	<b>100</b>	<b>10</b>	0.1	0.1	0.1	0.01	0.01	<b>10</b>	0.01	0.01	0.1	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>10</b>	<b>10</b>
	GMT	2	<b>10</b>	<b>5</b>	<b>2</b>	2	5	2	5	10	<b>5</b>	2	2	2	<b>10</b>	<b>2</b>	<b>5</b>	<b>5</b>	<b>10</b>	<b>10</b>	
Weighted Point-to-Plane	Weight	<b>10</b>	0.01	100	0.01	100	10	0.01	<b>100</b>	<b>100</b>	<b>10</b>	<b>10</b>	<b>10</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>10</b>	0.01	<b>100</b>	<b>100</b>	
	GMT	<b>10</b>	5	2	2	5	5	2	<b>5</b>	<b>2</b>	<b>2</b>	<b>5</b>	<b>5</b>	<b>2</b>	<b>5</b>	<b>5</b>	<b>2</b>	2	<b>5</b>	<b>5</b>	

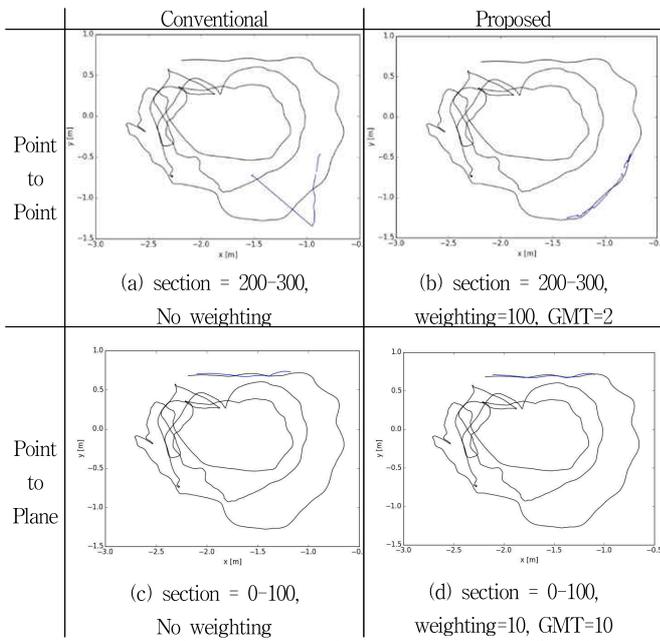


그림 5. 각각의 반복적 점군 정합 방법들을 이용해 예측된 카메라 궤적

그림 5는 각각의 반복적 점군 정합 방법들을 이용해 예측된 카메라 궤적을 보여준다. 그림 5(a)와 그림 5(b)는 teddy 시퀀스에서 200-300 구간의 색상 및 깊이 영상을 이용하여 예측된 카메라 궤적을 보여준다. 검은색의 선은 실제의 카메라 전체 궤적을 나타내고 파란색의 선은 현재 구간에서 예측된 카메라의 궤적을 나타낸다. 기존의 point-to-point 방법의 결과인 그림 5(a)에서는 실제 궤적에서 많이 벗어난 결과를 보여주는 반면 제안하는 point-to-point 방법인 그림 5(b)에서는 실제의 값에 가깝게 카메라의 궤적을 찾아주는 것을 확인할 수 있다. 그림 5(c)와 그림 5(d)에서도 역시 같은 구간에 대해 예측된 카메라 궤적이 제안하는 방법에서 더 정확하게 나오는 것을 알 수 있다.

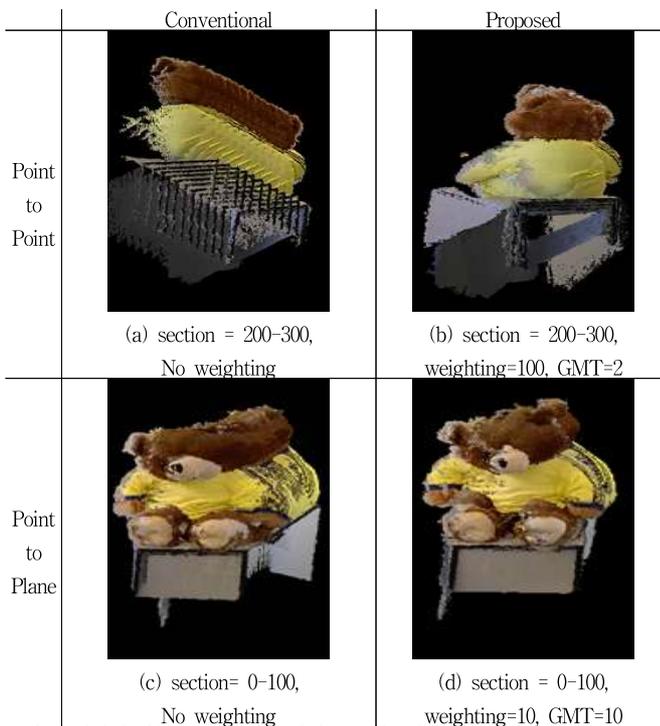


그림 6. 각각의 반복적 점군 정합 방법들을 이용해 복원된 3차원 객체

다음으로 그림 6에 각각의 반복적 점군 정합 방법들로 복원된 3차원 모델의 결과를 나타냈다. 그림 6(a)와 그림 6(b)는 200-300 프레임의 색상 및 깊이 영상을 이용하여 3차원 복원을 수행한 결과이다. 그림 6(a)는 기존의 point-to-point 방법을 이용한 결과인데 시간이 지남에 따라 누적된 카메라 궤적 오차가 복원된 3차원 모델의 정확성에 나쁜 영향을 미치고 있는 것을 확인할 수 있다. 반면, 그림 6(b)는 제안하는 point-to-point 방법을 이용한 결과로서 누적된 카메라 궤적 오차가 줄어들어 3차원 모델이 정확하게 복원된 것을 확인할 수 있다.

다음으로 그림 6(c)와 6(d)는 0-100 프레임의 색상 및 깊이 영상을 이용하여 3차원 복원을 수행한 결과이다. 그림 6(c)는 기존의 point-to-plane 방법을 이용한 결과이며 마찬가지로 시간의 흐름에 따른 누적 카메라 궤적 오차 때문에 3차원 객체 모델이 공간상에서 정확한 위치를 찾지 못하고 떠다니는 모습을 볼 수 있다. 그에 반해 그림 6(d)는 제안하는 point-to-plane 방법을 이용한 결과이며 비교적 정확한 3차원 모델을 얻은 것을 확인할 수 있다.

#### 4. 결론 및 향후 연구

본 논문에서는 3차원 객체 복원을 위해 색상 영상에서의 SIFT 특징점과 가중치를 이용한 반복적 점군 정합 방법을 제안했다. 제안하는 방법은 기존의 반복적 점군 정합 방법에서 발생하는 객체 표류 문제를 줄이고 실제의 카메라 궤적에 가깝게 카메라의 궤적을 예측한다. 하지만 현재는 각 구간에 대해 다양한 가중치를 모두 대입하여 절대 궤적 오차가 최소가 되는 가중치를 획득하여 기존의 방법과 비교하였으나 추후에는 구간 내 색상과 깊이 영상의 특성에 따라 적응적으로 최적의 가중치를 획득하는 방법에 대해 연구할 것이다.

#### 감사의 글

이 논문은 2015년도 정부(미래창조과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No.2011-0030079)

#### 참고문헌

- [1] D. W. Shin and Y. S. Ho, "Implementation of 3D object reconstruction using a pair of kinect cameras," APSIPA, pp. 1-4, 2014.
- [2] B. Curless and M. Levoy, "A Volumetric Method for Building Complex Models from Range Images.," SIGGRAPH, pp. 303-312, 1996.
- [3] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints.," IJCV, vol. 60, no. 2, pp. 91-110, 2004.
- [4] K. S. Arun, T. S. Huang, and S. D. Blostein, "Least-Squares Fitting of Two 3-D Point Sets.," PAMI, vol. 9, no. 5, pp. 698-700, 1987.
- [5] K. L. Low, "Linear least-squares optimization for point-to-plane icp surface registration," Chapel Hill, 2004.
- [6] F. Endres, J. Hess, N. Engelhard, and J. S., "An evaluation of the RGB-D SLAM system," ICRA, pp. 1691-1696, 2012.
- [7] <http://vision.in.tum.de/data/datasets/rgbd-dataset/download>