

실내환경 복원을 위한 다시점 카메라로
획득된 부분적 3차원 점군의 정합 기법

Registration Technique of Partial 3D Point
Clouds Acquired from a Multi-view Camera for
Indoor Scene Reconstruction

김세환, 우운택

Sehwan Kim and Woontack Woo

논문 2005-42CI-3-5

실내환경 복원을 위한 다시점 카메라로 획득된 부분적 3차원 점군의 정합 기법

(Registration Technique of Partial 3D Point Clouds Acquired from a
Multi-view Camera for Indoor Scene Reconstruction)

김 세 환*, 우 운 택**

(Sehwan Kim and Woontack Woo)

요 약

본 논문에서는 실내환경의 3차원 복원을 위해 다시점 카메라로부터 획득된 부분적인 3차원 점군에 대한 정합 기법을 제안한다. 일반적으로, 기존의 정합 방법들은 많은 계산량을 요하며, 정합하는데 많은 시간이 소요된다. 또한, 상대적으로 정밀도가 낮은 3차원 점군에 대해서는 정합이 어렵다. 이러한 문제점을 해결하기 위해 투영 기반 정합 방법을 제안한다. 첫 번째, 시간적 특성을 기반으로 변화량이 큰 3차원 점들을 제거하고, 공간적 특성을 이용하여 현재 화소의 주변 3차원 점들 참조하여 빈 영역을 채움으로써 깊이 영상 정제 과정을 수행한다. 두 번째, 연속된 두 장면에서의 3차원 점군을 동일한 영상 평면으로 투영하고, 두 단계 정수 매핑을 적용한 후 수정된 KLT (Kanade-Lucas-Tomasi) 특징 추적기를 사용해 대응점을 찾는다. 그리고 적응적 탐색 영역에 기반하여 거리 오차를 최소화함으로써 정밀한 정합을 수행한다. 마지막으로, 대응되는 점들에 대한 색을 참조하여 최종적인 색을 계산하고, 위의 과정을 연속된 장면에 적용함으로써 실내환경을 복원한다. 제안된 방법은 대응점을 2차원 영상 평면에서 찾음으로써 계산의 복잡도를 줄이며, 3차원 데이터의 정밀도가 낮은 경우에도 정합이 효과적이다. 또한, 다시점 카메라를 이용함으로써 몇 장면에 대한 색과 깊이 영상만으로도 실내환경의 3차원 복원이 가능하다.

Abstract

In this paper, a registration method is presented to register partial 3D point clouds, acquired from a multi-view camera, for 3D reconstruction of an indoor environment. In general, conventional registration methods require a high computational complexity and much time for registration. Moreover, these methods are not robust for 3D point cloud which has comparatively low precision. To overcome these drawbacks, a projection-based registration method is proposed. First, depth images are refined based on temporal property by excluding 3D points with a large variation, and spatial property by filling up holes referring neighboring 3D points. Second, 3D point clouds acquired from two views are projected onto the same image plane, and two-step integer mapping is applied to enable modified KLT (Kanade-Lucas-Tomasi) to find correspondences. Then, fine registration is carried out through minimizing distance errors based on adaptive search range. Finally, we calculate a final color referring colors of corresponding points and reconstruct an indoor environment by applying the above procedure to consecutive scenes. The proposed method not only reduces computational complexity by searching for correspondences on a 2D image plane, but also enables effective registration even for 3D points which have low precision. Furthermore, only a few color and depth images are needed to reconstruct an indoor environment.

Keywords: 실내환경 복원, 다시점 카메라, 가상환경 생성, 투영 기반 정합

I. 서 론

실세계 환경에 대한 영상 기반 3차원 복원 (3D reconstruction) 은 시각적인 사실감을 느끼며 가상환경을 체험하는데 있어 중요한 역할을 한다. 전통적으로 3차원 모델링 툴을 사용하여 실세계 환경에 대해 인위적

* 학생회원, ** 종신회원.

광주과학기술원 정보통신공학과
(Dept. of Info. & Comm., GIST)

* 본 연구는 광주과학기술원 실감방송 연구센터를 통한 정보통신부 대학IT연구센터 (ITRC) 사업의 지원에 의한 것입니다.

접수일자: 2004년12월18일, 수정완료일: 2005년5월10일

으로 (Artificially) 생성된 3차원 모델은 모델 생성을 위해 많은 시간을 요하지만 사실적이지 못하다는 단점이 있다. 그리고 능동적 거리 (Active Range) 측정 기법에 기반한 방법은 사실감은 유지하지만 고가의 장비를 필요로 하며, 모델을 생성하는데 많은 시간을 요한다. 또한, 텍스처와 3차원 정보를 획득하는 센서가 서로 다르기 때문에 이를 일치시키기 위한 처리가 요구되는 문제점이 있다. 이에 반해, 영상에 기반한 3차원 모델 복원 방법은 실사를 사용하므로 사실감을 보존할 뿐만 아니라 모델링 과정도 간단하다. 특별히, 일반 상용의 다시점 카메라 (Multi-view Camera)는 실시간으로 색 영상 뿐만 아니라 깊이 (depth) 영상도 동시에 제공하므로 실사 기반 모델링을 보다 용이하게 한다. 따라서 이러한 카메라를 통해 실세계 환경에 대한 3차원 복원을 수행하기 위해서는 여러 방향에서 획득된 3차원 데이터를 정교하게 정합하는 과정이 요구된다.

그 동안 실세계 환경의 3차원 복원을 위한 다양한 정합 방법들이 제안되었다. Besl 등에 의해 제안된 ICP (Iterative Closest Point) 알고리즘은 현재까지도 널리 사용되고 있다^[1]. Johnson 등은 색 정보를 결합한 Color ICP를 제안하고 이를 통해 실내환경을 복원하였다^{[2][3]}. 그리고 Blais 등은 모든 시점에서의 대응되는 부분 사이의 거리에 기반한 비용 함수를 최소화하기 위해 시뮬레이티드 어닐링 (Simulated Annealing)을 사용하였다^[4]. 또한 Levoy 등은 다양한 종류의 상 (Statue)에 대해 3차원 스캐너를 통해 얻은 점군을 복셀 (Voxel)에 기반한 방법을 통해 정합 과정을 수행하였으며, Kanade 등은 영상에 기반한 움직이는 객체의 3차원 복원을 위해 비슷한 방법을 적용하였다^{[5][6]}. 반면, Nishino 등은 여러 장의 거리 영상을 위한 강건한 정합을 구현하기 위해 M-estimator를 이용한 최적화 방법을 제안하였다^[7]. 특별히, Pulli는 데이터 획득 장치를 개발하고 평면의 원근 변형 (Planar Perspective Warping)을 이용한 투영 정합 방법을 제안하여 소규모 물체들을 모델링하였다^[8]. 이에 반해, Sharp는 불가변 특징 (Invariant Feature)을 정의하고 이를 이용한 ICP를 제안하였으며, Fisher 등은 원근 ICP (Projective ICP)를 이용하여 증강 현실 (Augmented Reality)에 적용하기도 하였다^{[9][10]}. 그러나 대부분의 방법은 고가의 정밀한 장비에 의존하며 3차원 모델을 생성하는데 상당한 시간을 요한다는 한계를 지닌다^{[4][5][9]}. 그리고 카메라를 사용하는 경우라 하더라도 근거리에 존재하는 물체로 그 대상을 제한하고 있다^[8]. 또한 3차원 깊이 정보가 상대적으로 큰 오차를 포함하

는 경우에는 효과적인 정합이 어렵다는 단점이 있다^{[11][12]}.

본 논문에서는 이러한 문제점들을 해결하기 위해 다시점 카메라로부터 획득된 색과 깊이 영상을 이용하여, 투영 (Projection)에 기반한 3차원 데이터의 정합 방법을 제안한다. 우선, 다시점 카메라로부터 획득된 3차원 데이터의 시공간 특성 (Spatio-temporal Property) 및 적응적 불확정 영역 (Adaptive Uncertainty Region)을 사용함으로써 깊이 영상 정제 (Depth Image Refinement) 과정을 수행한다. 두 번째, 연속된 두 시점에서의 3차원 점군을 동일한 영상 평면에 투영하고 겹치는 부분에 대해 두 단계 정수 매핑 (Two-step Integer Mapping)을 적용한 후, 수정된 KLT (Kanade-Lucas-Tomasi) 특징 추적기 (Feature Tracker)를 통해 대응점을 찾는다. 그리고 Levenberg-Marquardt 알고리즘을 적용하여 거리 오차를 최소화함으로써 두 집합의 3차원 점군 (3D Point Cloud)에 대한 정합을 수행한다. 마지막으로, 겹치는 부분의 대응되는 점들에 대해 선형적으로 최적화된 3차원 좌표를 구하고 새로운 색을 할당한다. 그리고 위의 과정을 연속적인 장면에 확장함으로써 실세계 환경을 복원한다.

실세계 환경에 대해 제안된 방법을 사용한 3차원 복원 방법은 다음과 같은 특징을 갖는다. 우선, 다시점 카메라를 위한 새로운 정합 방법을 제안함으로써 실세계 환경에 대한 3차원 복원을 보다 용이하게 수행할 수 있다. 지금까지 능동적 거리 측정 기법에 기반한 정밀한 광학 센서를 사용하여 가상환경을 생성하는 연구들이 많이 진행되어 왔다. 이는 비교적 정밀한 데이터를 획득할 수 있으므로 여러 방향에서 획득한 3차원 데이터를 정합하는 과정이 보다 수월하다. 그러나 장비가 상당히 클 뿐만 아니라 모델링하는데 오랜 시간이 소요된다는 단점이 있다. 비록 정밀도는 상대적으로 낮지만 이미 보정되어 있는 다수의 렌즈를 이용해 3차원 데이터를 제공하는 다시점 카메라의 상용화는 이러한 과정을 보다 단순화할 수 있다^{[11][12]}. 이를 위해 3차원 공간이 아닌 투영된 2차원 영상 평면에서 대응점을 효과적으로 찾고, 이로부터 카메라의 자세를 추정하는 방법을 제안한다. 이를 통해 정밀도가 상대적으로 낮은 3차원 데이터의 경우에도 효과적인 정합이 가능하도록 한다. 또한, 3차원 모델의 정합 과정을 2차원 영상 평면을 기반으로 수행함으로써 보다 빠른 정합 과정을 수행할 수 있다. 마지막으로, 이러한 방법은 다시점 카메라를 단지 실내 환경의 임의의 몇 곳에 위치시켜 장면에 대한 3차

원 복원을 가능하게 함으로써 비교적 용이한 가상환경 생성에 적용 가능하다.

본 논문의 구성은 다음과 같다. II장에서는 깊이 영상 정제 과정에 대해 기술하고, III장에서는 초기 정합을 기반으로 특징점을 이용한 정밀한 정합 방법에 대해 설명한다. 그리고 IV장에서는 최적의 색 선택 방법 및 연속된 장면으로의 확장에 대해 다룬다. V장에서 실험 및 분석에 대해 기술한 뒤, VI장에서 결론 및 추후 과제에 대해 언급한다.

II. 깊이 영상 정제

능동적 거리 측정 기법에 기반한 정밀한 광학 센서를 사용하는 방법에 비해, 물체로부터 반사되어 오는 빛에 의해 생성되는 영상을 이용하는 수동적 기법 (Passive Technique)은 스테레오 부정합 (Stereo Mismatching) 문제로 인해 정합 과정을 어렵게 한다. 즉, 물체의 경계 영역과 같은 깊이 불연속 영역이나 균질한 영역 (Homogeneous Area)은 피할 수 없는 스테레오 부정합 문제를 발생시키는 부분으로, 정합 과정을 수행하기 위해서는 미리 제거해야 하는 부분이다. 깊이 영상 정제 과정은 스테레오 부정합 문제로 인해 깊이 영상에 존재하는 오차가 있는 점들을 깊이 값들의 시공간적 특성을 이용하여 제거한다. 깊이 영상 정제 과정은 크게 두 단계로 나뉘는데, 첫 번째 단계에서는 깊이 값들이 시간에 따라 변화하는 시간적 특성 (Temporal Property)을 이용하여 오차를 갖는 점을 제거하며, 두 번째 단계에서는 공간적 특성 (Spatial Property)을 이용하여 빈 영역 (Hole)을 채우는 역할을 한다.

위에서 언급한 영역들에서의 깊이 값들은 주변의 값들과 비교하여 급격한 변화를 갖으며, 깊이 값들 또한 안정적이지 못하여 시간에 따라 변화한다. 이렇게 오차를 갖는 점들의 시공간적 특성을 통해 깊이 영상의 신뢰하기 어려운 영역을 제거할 수 있다. 그림 1은 실제 세계의 두 장면에 대한 3차원 복원을 위한 흐름도를 보인 것이다.

1. 스테레오 부정합으로 인해 오차를 갖는 3차원 점

실험에 사용되는 다시점 카메라로부터 획득되는 깊이 영상은 물체의 경계 영역 및 균질한 영역과 같은 신뢰하기 어려운 영역에서 많은 오차 성분을 나타낸다. 이러한 영역에서의 오차를 갖는 점들을 포함하는 3차원 점군을 기반으로 정합 과정을 수행하는 경우에는 수용

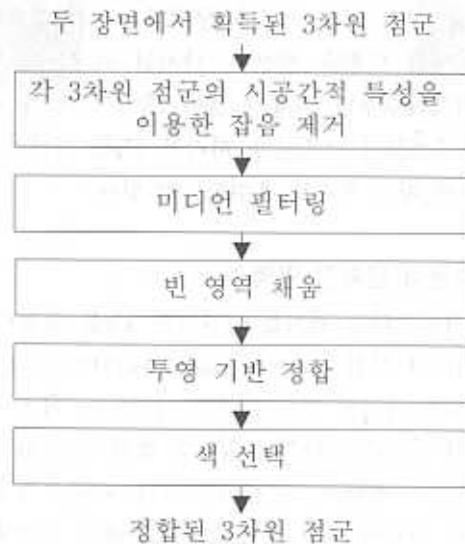


그림 1. 3차원 복원을 위한 흐름도
Fig. 1. Flow diagram for 3D reconstruction.

할 만한 정합 결과를 보장하기 어렵다. 따라서 보다 신뢰할 만한 정합 결과를 얻기 위해서는 동일한 장면에 대한 여러 장의 영상을 획득하고, 깊이 영상이 가지는 시공간적 특성에 기반하여 신뢰하기 어려운 영역을 제거해야 한다.

이를 수행하기 위해 동일한 장면에 대해 N_f 장의 깊이 영상을 획득한 후에 평균값을 계산한다. 그러나 움직임이 없는 정적인 장면에 대해서조차 동일한 화소에 대응되는 깊이 값이 시간에 따라 큰 변화량을 갖는 것들이 존재한다. 실제 세계 환경을 균등 확산면 (Lambertian Surface)이라 가정하지만, 이를 만족하지 못하는 물체들이 존재하기 때문에 시간, 카메라의 위치 및 조명 조건에 따라 화소값들이 달라진다. 이는 양안차 추정 (Disparity Estimation)시 물체의 경계 부분이나 균질한 영역에서 변화량이 큰 값을 생성한다.

깊이 영상에서의 각 화소에 대한 오차의 변화량이 가우시안 분포 (Gaussian Distribution)를 갖는다고 가정하고, 각 화소의 깊이 값을 조사하여 다음 조건에 해당되는 화소들은 제거한다. 즉, i 번째 화소에 대한 깊이 값의 변화량이 문턱값보다 크면 유효하지 않은 값으로 결정한다.

$$\sigma_i > \alpha Th_i \tag{1}$$

위 식에서, σ_i 는 i 번째 화소의 깊이 변화량에 대한 표준 편차를 나타낸다. 그리고 α 는 비례 인자 (Scale Factor), Th_i 는 i 번째 화소에 대한 문턱값 (Threshold

Value)을 나타낸다. 그러나 문턱값은 카메라로부터의 위치에 의해 영향을 받는다. 다시점 카메라의 특성에 따라 카메라로부터 물체가 멀어질수록 양안차 추정시 오차가 증가하기 때문이다. 따라서 Th 는 카메라로부터의 위치에 따른 함수로 표현되어야 한다.

2. 적응적 불확정 영역

카메라로부터의 위치를 고려하여 Th 를 결정하기 위해 카메라의 시선 (Line of Sight: LOS)에 따라 각 축 방향으로의 거리에 따른 오차를 계산한다. 이는 카메라로부터의 위치를 고려하여 일정한 범위 내에 있는 점들은 오차를 구별할 수 없기 때문이다. 이렇게 3차원 공간상에서 거리에 따라 적응적으로 변하는 유효한 공간은 3차원 공간에서 타원체 모양을 가지며 적응적 불확정 영역 (Adaptive Uncertainty Region)이라 명명한다.

실험에 사용되는 다시점 카메라는 내부적으로 카메라 보정 오차 (Calibration Error) p 와 영상간의 양안차 추정시 발생하는 상관도 오차 (Correlation Error) m 을 갖는다. 이때, 3차원 좌표는 다음과 같이 결정된다.

$$\frac{u}{f} = \frac{x}{z}; \frac{v}{f} = \frac{y}{z}; \frac{z}{f} = \frac{B}{d} \tag{2}$$

위 식에서, d 는 양안차, (u, v) 는 영상 중심에 대한 상대적인 현재 화소의 위치, B 는 기본선 (Baseline) 그리고 f 는 초점 거리 (Focal Length)를 나타낸다. 그리고 (x, y) 에서의 허용 오차 (Tolerance)는 보정 오차에 의해 다음과 같이 결정된다.

$$\Delta x = \frac{pz}{f}; \Delta y = \frac{pz}{f} \tag{3}$$

반면, z 에 대한 정확도 (Accuracy)는 다음과 같이 유도된다.

$$\Delta z = \frac{fB}{d-m} - \frac{fB}{d} \tag{4}$$

식 (3)과 식 (4)를 이용하여 카메라로부터 물체까지의 거리에 따른 허용 오차를 나타내면 그림 2에 보인 것과 같다. 여기서, 그림 2(a)는 거리에 따라 변화하는 Δx 또는 Δy 를 나타내고, 그림 2(b)는 Δz 를 나타낸 것이다. 그림에서 보는 것과 같이 Δx 또는 Δy 는 z 의 값에 따라 선형적으로 증가하지만 Δz 는 곡선의 형태를 보이고 있음을 알 수 있다. 이러한 값들은 각 좌표 축으로 가우시안 분포를 가지며 제곱 평균 제곱근

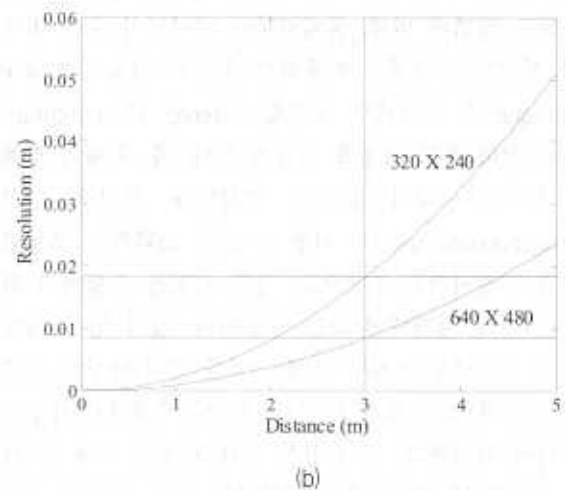
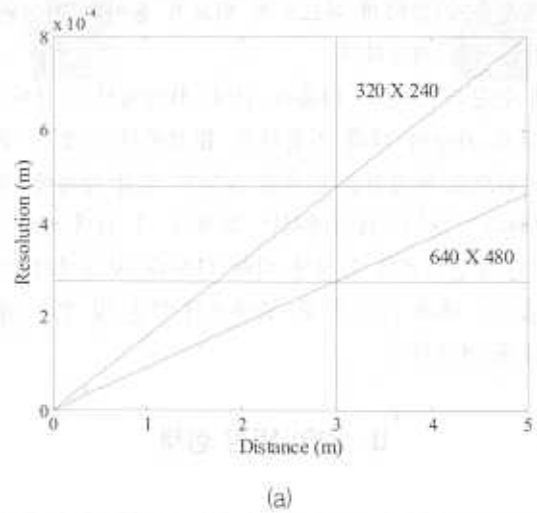


그림 2. 카메라로부터의 거리에 따른 허용 오차 (a) Δx 또는 Δy (b) Δz
 Fig. 2. Distance from camera vs. resolution. (a) Δx or Δy (b) Δz .

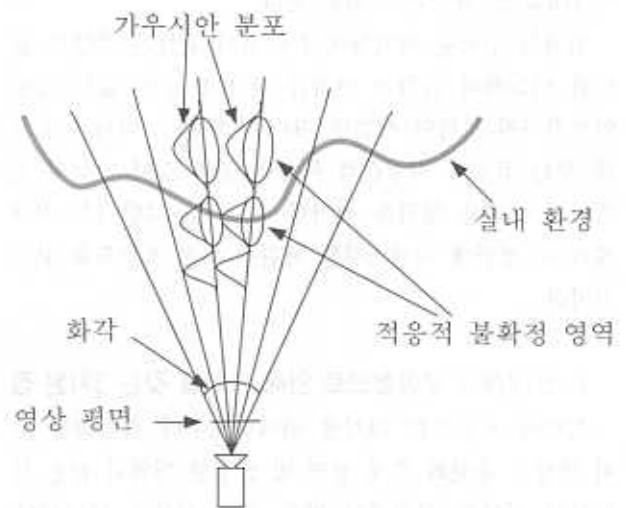


그림 3. 적응적 불확정 영역
 Fig. 3. Adaptive uncertainty region.

(Root Mean Square: RMS) 오차, 즉, 표준 편차 (Standard Deviation)에 대응되는 값들이다.

이와 같은 사실을 기반으로 일정한 화각 (Field Of view: FOV)을 갖는 다시점 카메라에 적용하면 그림 3에 보인 것과 같다. 즉, 거리가 카메라로부터 멀어짐에 따라 x, y 방향으로 선형적으로 증가하고, z 축 방향으로 그림 2(b)의 곡선에 따라 비례적으로 증가하는 가우시안 분포를 얻는다.

다시점 카메라로 획득된 영상의 각 화소에 대한 3차원 점이 주어지며 이는 3차원 공간에 거리에 따라 변하는 타원체 모양의 불확정 영역을 생성한다. 따라서 거리에 따라 적용적으로 변하는 불확정 영역이 고려된 Th 가 결정되며 이는 다음과 같이 표현된다.

$$\frac{(x-x_c)^2}{(\Delta x)^2} + \frac{(y-y_c)^2}{(\Delta y)^2} + \frac{(z-z_c)^2}{(\Delta z)^2} = 1 \quad (5)$$

위 식에서, $(x_c, y_c, z_c)^T$ 는 영상 평면의 각 화소에 대응되는 3차원 상의 좌표로서 타원체 모양의 중심이 카메라의 광학 중심 (Optical Center)에서 이동한 이동 벡터 (Translation Vector)를 나타낸다. 그리고 Δx , Δy 그리고 Δz 는 각 축으로의 불확정 거리 (Uncertainty Distance)를 나타낸다. 그러나 위의 모양은 카메라의 광학 중심을 기준으로 화각의 방향에 대해 일정한 각도만큼 회전되어야 한다. 따라서 식 (6)의 타원체 모양을 카메라의 광학 중심을 기준으로 식 (7)에 기반하여 회전한 후 $(x_c, y_c, z_c)^T$ 로 이동한다.

$$\frac{x^2}{(\Delta x)^2} + \frac{y^2}{(\Delta y)^2} + \frac{z^2}{(\Delta z)^2} = 1 \quad (6)$$

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = R_1 R_2 \begin{pmatrix} x \\ y \\ z \end{pmatrix} + \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$$

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & z_c/d & y_c/d \\ 0 & -y_c/d & z_c/d \end{pmatrix} \quad (7)$$

$$R_2 = \begin{pmatrix} d & 0 & x_c \\ 0 & 1 & 0 \\ -x_c & 0 & d \end{pmatrix} \quad d = \sqrt{y_c^2 + z_c^2}$$

위 식에서, $(x' y' z')^T$ 는 $(x_c y_c z_c)^T$ 를 고

려하여 카메라의 광학 중심을 기준으로 분포하게 될 카메라로부터의 위치에 따른 오차 분포에 대한 최종적인 적용적 불확정 영역을 나타낸다. 따라서 식 (1)은 카메라로부터 3차원 점이 위치하는 $(x_c, y_c, z_c)^T$ 를 고려한 함수로 재정의 된다.

$$\sigma_i > \alpha Th_i(x_c, y_c, z_c) \quad (8)$$

3. 빈 영역 채움

위의 과정을 통해 정적인 장면에 대해 매우 큰 변화량을 갖는 값들은 제거된다. 즉, 균등 확산면이라고 가정한 배경에 대해 이러한 특성을 만족하지 못하는 부분들은 제거된다. 그리고 물체의 경계 영역 및 균질한 영역에서와 같이 양안차 추정의 오차가 심하게 변화하는 부분들도 자동으로 제거된다. 본 과정 후에 존재하는 스폿 잡음 (Spot Noise)과 같은 점은 미디언 필터 (Median Filter)를 통해 제거된다.

그러나 본 과정을 통해 생성된 빈 영역이나 양안차 추정이 어려운 균질한 면에 대해서는 빈 영역 채움 (Hole Filling) 과정이 요구된다. 그림 4는 양안차 및 색 정보를 이용해 가상환경에 역투영된 3차원 데이터들을 보인 것으로, 오른쪽은 특정 영역을 확대한 것이다. 유효한 양안차 정보를 기반으로 양안차가 획득되지 않은 부분을 식 (9)를 통해 보완한다.

$$\begin{aligned} x_c &= ((1-u) \times x_r + u \times x_l + (1-v) \times x_l + v \times x_b) / 2 \\ y_c &= ((1-u) \times y_r + u \times y_l + (1-v) \times y_l + v \times y_b) / 2 \\ z_c &= ((1-u) \times z_r + u \times z_l + (1-v) \times z_l + v \times z_b) / 2 \end{aligned} \quad (9)$$

위 식에서, $(x_l, y_l, z_l)^T$, $(x_b, y_b, z_b)^T$, $(x_r, y_r, z_r)^T$ 및 $(x_c, y_c, z_c)^T$ 는 그림 4에 보인 것과 같이 현재 화소를 중심으로 2차원 영상의 각각 상하좌우에 해당되는 화소의 3차원 좌표를 나타낸다. 그리고 u 와 v 는 현재

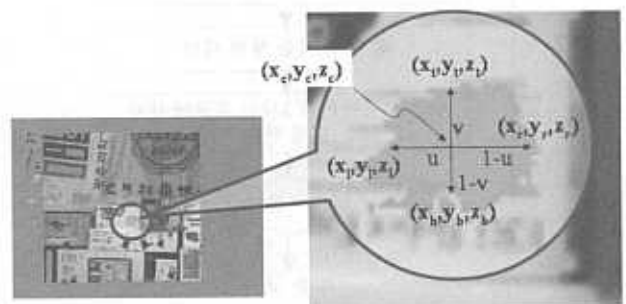


그림 4. 빈 영역 채움

Fig. 4. Hole filling.

화소를 중심으로 각각 가로 및 세로 방향으로의 비를 나타낸다. 이러한 과정을 통해 현재 화소에 대한 3차원 좌표인 $(x_c, y_c, z_c)^T$ 를 추정할 수 있다.

그러나 깊이 정보가 다른 영역에서 위와 같은 과정을 수행하는 경우에는 오히려 오차를 증가시킨다. 이러한 현상을 방지하기 위해, 물체의 경계 부분에서와 같이 불연속한 깊이 값이 문턱값 Th_{NM} 이상으로 발생하는 곳은 수행하지 않도록 한다. 또한, 상하좌우의 3차원 값을 조사하여 평면이라고 가정할 수 있을 만큼 크기가 작은 빈 영역에 대해서만 위의 과정을 적용한다.

III. 부분적 3차원 점군의 정합

2장에서 설명했던 것과 같이, 깊이 영상 정제 과정은 수동적 기법의 피할 수 없는 스테레오 정합 문제로 인해 발생하는 오차 부분들을 제거하고 3차원 점군의 오차 범위를 줄인다. 그러나 이러한 과정은 수동적 기법으로 인해 발생하는 오차가 발생하는 부분은 제거하지만, 획득되는 데이터는 여전히 능동적 거리 측정 기법에 기반한 정밀한 광학 센서를 사용하는 방법에 비해 정밀도가 낮기 때문에 정합 과정이 용이하지 않다. 즉, 다시점 카메라로부터 획득된 깊이 영상은 상대적으로 깊이 정보의 오차가 크기 때문에 ICP를 이용한 정합 과

정이 용이하지 않다. 이는 3차원 점군이 상대적으로 큰 오차를 포함하기 때문에 최단 거리를 기반으로 하는 ICP는 많은 오류를 야기시키기 때문이다. 따라서 원본 시점 (Source View)의 3차원 데이터의 대응점을 대상 시점 (Destination View)의 3차원 데이터에서 찾는 과정을 효과적으로 수행하기 위해 투영에 기반한 정합 방법을 제안한다. 그림 5는 그림 1에서의 투영 기반 정합 (Projection-based Registration) 과정에 대한 순서도를 보인 것이다.

1. 초기 정합

본 논문에서는 단일 다시점 카메라를 움직이는 로봇 (Mobile Robot)에 부착하거나 사람이 들고 다니며 움직이는 경우를 상정한다. 따라서 카메라 보정 패턴 (Calibration Pattern)과 같은 특정 물체에 의존하지 않고, 획득되는 영상에서의 특징점을 선택하고 이를 추적함으로써 단일 카메라의 자세를 연속적으로 추정할 수 있어야 한다. 이를 위해 우선 동일 평면 카메라 보정 패턴 (Coplanar Calibration Pattern)과 다시점 카메라의 구조적인 조건 (Structural Constraints)을 이용하여 기준 위치에서의 카메라의 기준 자세를 획득한다. 그리고 다시점 카메라를 움직이면서 추적되는 특징점에 대한 3차원 정보를 기반으로 강제 변환 (Rigid-body Transformation)을 수행한다. 마지막으로, 이를 최적화하기 위해 bundle adjustment를 수행한다¹³⁾¹⁴⁾. 이러한 과정을 통해 카메라간의 자세를 연속적으로 획득하고 이를 기반으로 각 위치에서 다시점 카메라로부터 획득된 3차원 점군을 역투영 (Back-projection) 시킴으로써

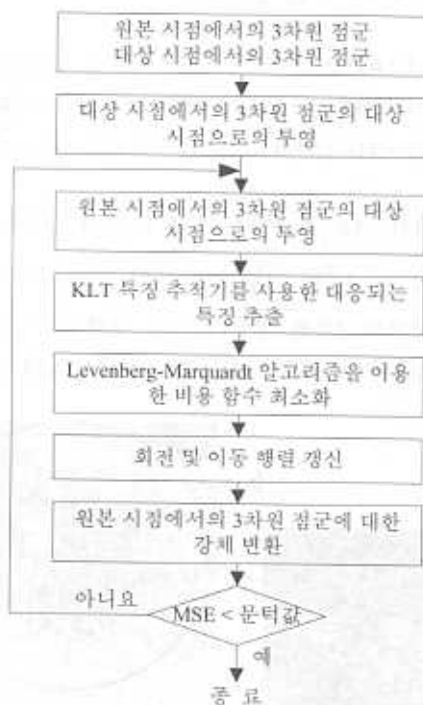


그림 5. 투영 기반 정합의 흐름도
Fig. 5. Flow diagram for projection-based registration.

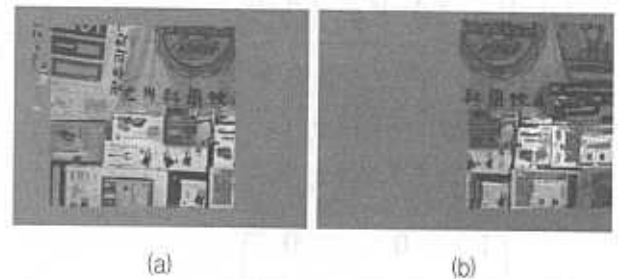


그림 6. 3차원 점군의 2차원 영상 평면으로의 투영 (a) 대상 시점의 3차원 점군의 대상 시점으로의 투영 (b) 원본 시점의 3차원 점군의 대상 시점으로의 투영

Fig. 6. Projection of 3D point cloud onto 2D image plane (a) projection of 3D point cloud of destination view onto its own view (b) projection of 3D point cloud of source view onto the destination view.

3차원 점군들 사이에 일치된 초기 정합 (Initial Registration) 과정을 수행한다.

그림 6은 초기 정합 후에 대상 시점으로 투영된 영상을 보인 것이다. 그림 6(a) 및 그림 6(b)는 각각 대상 및 원본 시점에서 얻은 3차원 점군을 대상 시점으로 투영한 결과이다. 물체들과의 구별을 위해 바탕은 128로 설정되었다. 원본 시점에서의 3차원 점군을 대상 시점으로 투영할 때 발생하는 자기 폐색 (Self-occlusion) 현상은 카메라의 LOS를 이용하여 제거된다. 이상적인 경우에는 그림 6(b)가 그림 6(a)와 정확히 겹치지만, 실체는 양안차 추정 및 카메라 보정 등으로 인한 오차가 발생한다. 따라서 원본 시점에서의 영상인 그림 6(b)와 대상 시점에서의 영상인 그림 6(a)의 겹치는 부분에 대해 그 오차를 최소화 하는 원본 카메라의 투영 행렬 (Projection Matrix) P_S 을 찾음으로써 두 집합의 3차원 점군을 정합할 수 있다. 즉, 초기 정합 과정에서 획득된 카메라 사이의 관계는 양안차 추정 및 보정 등으로 인한 오차가 존재하므로 이를 보상하기 위한 정밀한 정합 (Fine Registration) 과정을 수행해야 한다.

2. 색과 질감 정보를 이용한 정밀한 정합

양안차 추정 및 보정 등의 오차로 인해 발생하는 초기 정합 결과에 대해 정밀한 정합을 수행하기 위해서는 정확한 대응점을 찾는 과정이 요구된다. 이를 위해 3차원 공간에서 대응점을 찾는 대신 상대적으로 오차가 큰 3차원 점에 대해서도 강건한 결과를 얻기 위해 2차원 영상 평면에서 색 (Color)과 질감 (Texture) 정보를 이용하여 정합하는 방법을 제안한다^[15]. 2차원영상 평면에서의 겹치는 부분에 대해 비용 함수 (Cost Function)를 정의하기 위해 색과 함께 질감 정보를 사용한다. 즉, 전체 영상을 일정한 크기의 블록 (Block)으로 나누고 원본 시점에서의 영상과 대상 시점에서의 영상에서 각 블록의 겹치는 부분 대해 색 기술자 (Color Descriptor)와 질감 기술자 (Texture Descriptor)를 적용하여 특징을 추출한다. 그림 7은 단일 블록에 대해 색과 질감에 대한 특징을 추출하는 과정을 보인 것이다.

비록 모델링하려는 객체가 균등 확산면이라 가정하지만, 동일한 배경을 획득하는 카메라에 대해서도 카메라의 위치에 따라 셰이딩 (Shading)의 차이가 발생한다. 일반적으로, 셰이딩은 물체의 휘도 (Luminance) 성분과는 달리 색도 (Chrominance) 성분에는 큰 영향이 없다. 따라서 원본 시점에서의 3차원 데이터의 대응점을 대상 시점에서의 3차원 데이터에서 찾기 위해 이러

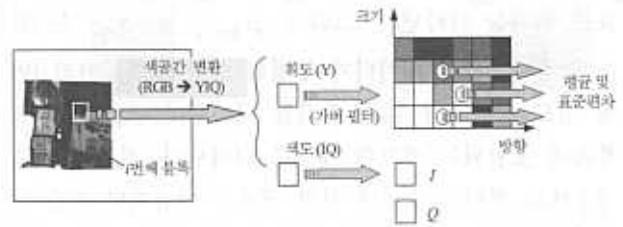


그림 7. 단일 블록에 대한 색과 질감 기술자
Fig. 7. Color and texture descriptors for one block.

한 특성을 반영한 색 공간의 선택은 매우 중요하다. 본 논문에서는 YIQ 색 공간을 사용하여 명도 성분을 색도 성분과 분리함으로써 셰이딩의 영향을 감소시킬 수 있다.

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.299 & 0.587 & 0.114 \\ 0.1687 & -0.3313 & 0.5 \\ 0.5 & -0.4187 & -0.0813 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} 0 \\ 128 \\ 128 \end{pmatrix} \quad (10)$$

그리고 색도 성분만을 고려한 색 기술자를 기반으로 다음과 같이 색도 성분에 대한 비용 함수를 정의한다.

$$ColorDiff_i = \sum_{j=0}^{N_{p,i}-1} \sqrt{(I_{Dst,j} - I_{Src,j})^2 + (Q_{Dst,j} - Q_{Src,j})^2} \quad (11)$$

위 식에서, $I_{Dst,j}$ 및 $Q_{Dst,j}$ 는 대상 시점의 3차원 데이터가 투영된 영상 평면에서의 j 번째 화소의 I 와 Q 값을 각각 나타낸다. 그리고 $I_{Src,j}$ 및 $Q_{Src,j}$ 는 원본 시점의 3차원 데이터가 투영된 영상 평면에서의 j 번째 화소의 I 와 Q 값을 각각 나타낸다. 또한, $N_{p,i}$ 는 i 번째 블록에 포함되는 화소 중에 투영된 원본과 대상 시점의 데이터가 모두 유효한 화소의 개수이다.

그리고 질감 기술자로는 가버 필터 (Gabor Filter)를 사용한다. 가버 필터의 크기 (Scale) (또는 주파수 (Frequency)) 및 방향 (Orientation) 특성은 효과적인 질감 분석을 가능하게 한다^[16]. 따라서 영상의 각 블록에 대해 M 개의 크기와 N 개의 방향을 갖는 가버 필터를 적용하면 실수부 (Real)와 허수부 (Imaginary)에 대해 각각 $M \times N$ 개의 영상이 획득된다. 각 영상에 대해 평균과 표준 편차를 특징으로 사용하여 다음과 같은 비용 함수를 정의한다.

$$TextureDiff_i = \sum_{j=0}^{N_{p,i}-1} \sqrt{(\mu_{Dst,j} - \mu_{Src,j})^2 + (\sigma_{Dst,j} - \sigma_{Src,j})^2} \quad (12)$$

위 식에서, $\mu_{Dst,j}$ 및 $\sigma_{Dst,j}$ 는 대상 시점의 3차원 데이터가 투영된 영상 평면에서의 j 번째 밴드의 평균과

표준 편차를 나타낸다. 그리고 $\mu_{Src,i}$ 및 $\sigma_{Src,i}$ 는 원본 시점의 3차원 데이터가 투영된 영상 평면에서의 i 번째 밴드의 평균과 표준 편차를 나타낸다. $N_{b,i}$ 는 i 번째 블록에 포함되는 밴드의 개수를 나타낸다. 질감이 다른 경우에도 색이 비슷하면 영역 구별이 어렵지만 질감 기술자를 통해 보다 정확한 구별이 가능하다. 이를 위한 최종적인 비용 함수는 다음과 같다.

$$E = \sum_{i=0}^{N_b-1} (ColorDiff_i + \alpha \cdot TextureDiff_i) \quad (13)$$

여기서, α 는 색과 질감 정보간의 가중치를 결정하는 인자로 사용되며, 실험적으로 결정된다. 그리고 N_b 는 영상에서의 블록의 개수를 의미한다.

따라서 각 시점에서 대응되는 블록에 대해 비용 함수의 오차를 최소화하는 원본 시점의 자세 (R_{Src}, T_{Src})를 대상 시점의 자세 (R_{Dst}, T_{Dst})를 기준으로 표현하면 다음과 같다.

원본과 대상 시점의 투영된 데이터에 대해 대응되는 블록들이 주어질 때,
비용 함수 E 가 최소가 되도록
(R_{Dst}, T_{Dst})를 기준으로 (R_{Src}, T_{Src})를 찾는다. (14)

위 식에서, E 는 각 블록에 대한 특징으로 표현된 비용 함수를 나타낸다. 이를 최소화하기 위해 비선형 최적화 (Non-linear Optimization) 방법인 Levenberg-Marquardt 알고리즘을 사용한다.

3. 대응되는 특징을 이용한 정밀한 정합

위에 제안된 각 블록에 대한 색과 질감 정보를 이용한 방법은 수렴 속도가 느리다는 단점이 있다. 이를 해결하기 위해 겹치는 부분에서의 대응되는 특징점 (Corresponding Features)을 이용한 방법을 제안한다. 이는 특징점을 이용하기 때문에 보다 빠른 수렴 속도를 보장하며 질감 정보를 계산하기 위해 요구되는 계산의 복잡도를 줄일 수 있다는 장점도 있다.

원본 시점에서의 3차원 데이터를 대상 시점으로 투영할 때 실수 (Floating-point Number) 값을 갖는다. 이러한 값들은 그림 9(a)와 같이 화소가 값을 갖지 못하는 부분이 발생한다. 이러한 부분은 수정된 KLT 특징 추적기를 통해 대응되는 특징점을 찾을 때 잘못된 결과 (False Alarm)를 발생시킨다. 따라서 이러한 부분을 제거하면서, 동시에 투영되는 영상을 그대로 유지하는 방법이 요구된다. 이러한 요구 사항을 만족시키기 위해 그림 8과 같은 두 단계 정수 매핑 방법을 사용한

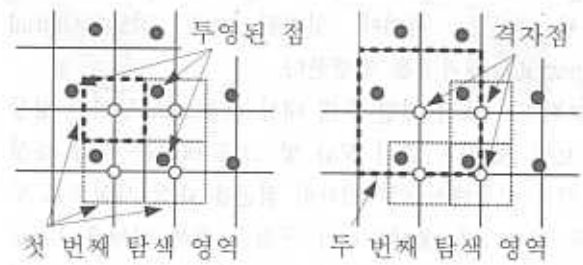


그림 8. 두 단계 정수 매핑
Fig. 8. Two-step integer mapping.



그림 9. 두 단계 정수 매핑 (a) 적용 전 (b) 적용 후
Fig. 9. Two-step integer mapping (a) before (b) after.

다. 실선은 영상에서 격자점이 존재하는 부분으로 원은 격자점, 검은 원은 원본 시점에서의 3차원 데이터가 대상 시점으로 투영된 화소를 나타낸다.

첫 번째 단계에서는 각 격자점을 중심으로 탐색 영역을 x 축, y 축 방향으로 $-0.5 \sim +0.5$ 로 설정한다. 그리고 탐색 영역 내에 존재하는 점들에 대해 거리에 따른 가중치를 고려하여 격자점에 대한 색을 결정한다. 투영된 대부분의 점들은 이 단계에 해당되며 특히, 대부분의 경우 한 개의 화소만이 포함된다. 그러나 첫 번째 과정에 포함되지 않는 투영된 점들이 존재하는데 이를 위해 두 번째 과정이 수행된다. 즉, 탐색 영역을 $-1.0 \sim +1.0$ 으로 설정하고 동일한 과정을 수행한다. 그림 9는 그 결과를 보인 것이다. 그림 9(a)는 그림 6(b)의 특정 부분을 확대한 것을 나타내며, 그림 9(b)는 두 단계 정수 매핑을 적용한 후의 결과를 보인 것이다. 이러한 방법을 통해 투영되는 값이 없는 격자점을 제거하고, 동시에 투영되는 원영상을 그대로 유지함으로써 수정된 KLT 특징 추적기 성능을 개선시킬 수 있다.

영상 평면의 겹치는 부분에 대해 비용 함수를 정의하기 위해 특징점을 찾고 대응되는 특징점들 사이의 유클리디언 거리 (Euclidean distance)를 사용한다. 이상적인 경우, 대응되는 특징점들은 완전히 겹치게 되므로 다음과 같은 비용 함수를 최소로 함으로써 카메라의 자세를 추정하게 된다. 즉, 각 시점의 N_{feat} 개의 대응되는 특징점들에 대해 그 오차를 최소화하는 원본 시점의 자세

(R_{Src}, T_{Src}) 는 대상 시점의 자세 (R_{Dst}, T_{Dst}) 를 기준으로 구해진다. 이는 다음과 같이 표현된다.

원본과 대상 시점의 투영된 데이터에 대해 대응되는 특징점들이 주어질 때, 비용 함수 E 가 최소가 되도록

$$E = \sum_{i=1}^{N_{cor}} \sqrt{(x_{Dst,i} - x_{Src,i})^2 + (y_{Dst,i} - y_{Src,i})^2} \quad (15)$$

(R_{Dst}, T_{Dst}) 를 기준으로 (R_{Src}, T_{Src}) 를 찾는다.

위 식에서, $(x_{Dst,i}, y_{Dst,i})$ 및 $(x_{Src,i}, y_{Src,i})$ 는 대상 시점에 투영된 대상 시점 및 원본 시점에서의 3차원 데이터의 2차원 영상에서의 (x, y) 좌표를 나타내며, E 는 그 사이의 유클리디언 거리를 나타낸다. 이를 최소화하기 위해서 Levenberg-Marquardt 알고리즘을 사용한다.

IV. 3차원 점군에 대한 표면 복원

1. 융합 (Integration)

2차원 영상 평면으로 투영된 데이터를 기반으로 정합을 수행한 후, 실세계 공간에서는 물리적으로 동일한 점임에도 불구하고 3차원 공간상의 데이터들은 양안차 추정 오차 때문에 동일한 점을 점유하지 못한다. 따라서 물리적으로 동일한 두 3차원 점들은 동일한 3차원 좌표를 갖도록 하는 과정이 요구된다. 즉, 대상 시점을 기준으로 원본 시점에 대한 카메라 파라미터를 추정 한 후에 겹치는 부분에서의 중복되는 3차원 데이터를 다듬는 (Trimming) 과정 및 색 선택 과정이 요구된다.

이를 위해서는 각 카메라로부터 획득되는 각 화소에 대한 3차원 점들이 일정한 오차 범위를 가지고 존재하는 적용적 불확정 영역 및 그 영역에서의 3차원 점의 존재 확률이 가우시안 오차 분포를 갖는다는 사실을 이용한다. 양안차 영상 정제 과정 후에 정밀한 정합을 수행한 3차원 점군은 물리적으로 동일한 점에 대해서는 동일한 영상 평면으로 투영된다. 즉, 두 장면에 대한 정밀한 정합이 수행된 후에는 그림 6(a) 및 그림 6(b)에서와 같이 대상 시점에서의 겹치는 부분의 각 격자점은 두 장면의 대응점이 된다. 이는 원본 시점에서의 3차원 데이터와의 대응점을 알 수 있다는 것을 의미한다. 따라서 이렇게 대응되는 점들에 대해 선형 삼각화 방법 (Linear Triangulation Method)을 사용하여 겹치는 부분에서 최종적인 3차원 좌표를 획득할 수 있다^[17].

이러한 과정을 통해 두 장면에 대한 최종적인 정합을 수행한 후에는 카메라의 위치에 따라 서로 다른 조명 조건 (Lighting Condition)의 변화를 고려하기 위해 색

조정 (Color Adjustment) 과정이 요구된다. 여기서는, 환경에 존재하는 모든 재질 (Material)이 균등 확산면을 만족한다고 가정한다.

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \left(v \times \begin{bmatrix} R_{Dst} \\ G_{Dst} \\ B_{Dst} \end{bmatrix} + u \times \begin{bmatrix} R_{Src} \\ G_{Src} \\ B_{Src} \end{bmatrix} \right) / (u + v) \quad (16)$$

여기서, u 와 v 는 겹치는 부분의 왼쪽 및 오른쪽 가장자리에서 현재 화소까지의 거리를 나타낸다. 그리고 R_{Dst} (또는 G_{Dst}, B_{Dst}) 및 R_{Src} (또는 G_{Src}, B_{Src})는 현재 화소의 빨간색 또는 녹색, 파란색을 각각 나타내며, R' (또는 G', B')은 겹치는 부분에서의 최종적인 빨간색 또는 녹색, 파란색을 각각 나타낸다.

2. 실내환경의 3차원 복원

위에서 제안한 정합 방법을 통해 연속된 두 위치에서 획득된 영상 및 3차원 점군에 대해 복원된 3차원 장면을 얻을 수 있다. 따라서 이러한 과정을 확장하여, 실세계 환경에 대해 다시점 카메라를 연속적으로 위치시켜 획득한 3차원 점군의 집합에 위와 같은 과정을 적용하여 최종적으로 3차원 복원을 수행할 수 있다. 그림 10은 이에 대한 개념도를 보인 것이다.

그림 10에 보인 것과 같이 실내환경으로부터의 상대적인 거리를 이미 알고 있는 카메라 보정 패턴을 특정한 곳에 위치시킨다. 그리고 실세계 좌표계의 원점을 기준으로 카메라 보정을 통해 카메라의 내부 (Intrinsic) 및 외부 (Extrinsic) 파라미터를 추정한다. 즉, Tsai의 방법을 응용하여 다시점 카메라의 구조적인 정보를 이용한 카메라 내 보정 (Intra-calibration) 방법을 사용하여 $[R T]_0$ 을 구한다^[13]. 여기서, $[R T]_0$ 은 카메라 보정 패턴의 특정 위치로부터의 카메라 자세를 의미한다. 그리고 카메라를 임의의 거리만큼 옮긴 후에 환경에 대한

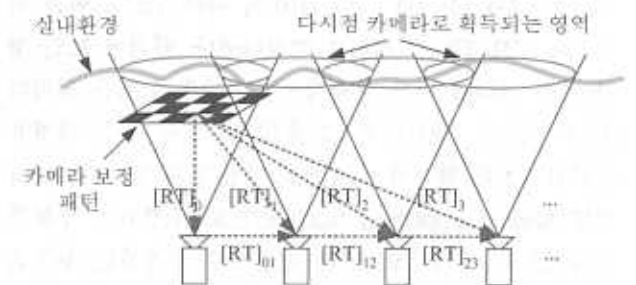


그림 10. 실내환경 복원에 대한 개념도
Fig. 10. Indoor scene reconstruction.

3차원 데이터를 획득한다. 이때, 카메라의 위치는 바로 전의 카메라 위치에서 획득된 데이터와 중복되는 영역이 1/3 정도 존재하도록 한다. 겹치는 영역이 너무 작으면 정밀한 정합 과정을 수행한 결과가 부정확해진다. 따라서 이러한 조건 하에 획득된 연속된 3차원 데이터에 위에 언급된 방법을 적용한다. 이렇게 함으로써 기준 카메라 위치에서 두 번째 카메라까지의 자세 $[R \ T]_0$ 를 획득할 수 있다. 이는 기준 카메라로부터 두 번째 카메라까지의 자세 $[R \ T]_1$ 을 계산할 수 있음을 의미한다.

이와 같은 방식으로 실내환경에 대해 각 위치에서의 기준 카메라로부터의 상대적인 자세 $[R \ T]_i$ 를 얻을 수 있고 이를 확장하여 실내환경에 대해 3차원 복원을 수행할 수 있다. 따라서 다중 기본선 스테레오 (Multi-baseline Stereo)를 통한 깊이 정보 추정 과정을 생략할 수 있다. 또한, 계산 시간을 줄일 수 있는 효과가 있다.

V. 실험 결과 및 분석

본 실험은 일반적인 실내환경의 정상적인 조명하에서 수행되었다. 영상 및 3차원 점군 획득을 위해 다시점 카메라인 Digiclops를 사용하였으며, Xeon 2.8 GHz CPU 컴퓨터에서 수행되었다. Digiclops는 IEEE1394 방식의 다시점 카메라로서 동일한 평면 위에 있는 3개의 렌즈를 이용하여 양안차 추정을 통해 실시간으로 3차원 정보를 획득한다. Digiclops는 ICX084AK CCD 센서를 사용하며, 카메라의 초점 거리 $f = 6 \text{ mm}$, 기본선 $B = 10 \text{ cm}$ 이다. 실험을 위해 7×5 개의 격자점을 가지고 격자간의 간격은 10.6 cm 인 평면 패턴을 사용하였다.

그림 11은 깊이 영상 정제 과정의 결과를 보인 것이다. 그림 11(a) 및 그림 11(b)는 원 영상과 대응되는 양안차 영상을 나타낸다. 이를 통해 얻어진 3차원 점군 및 거리에 따른 깊이 영상 정제와 빈 영역 채움 과정의 결과를 그림 11(c)와 그림 11(d)에 나타내었다. 이를 위해 $N_f = 20$, $T_{thd} = 0.15$ 로 설정하였다. 결과에 보인 것과 같이 거리에 따라 적응적으로 유효하지 않은 영역의 3차원 점들을 제거하고 그 후에 빈 영역 채움 과정을 수행함으로써 양안차 추정으로 인한 오차 및 비균등 확산면 (Non-Lambertian Surface)으로 추정되는 부분을 효과적으로 제거하고, 또한 평면으로 추정되는 부분을 채웠음을 알 수 있다.

그림 12는 대응되는 두 영상에서의 대응되는 특징점을 찾고 그 사이의 거리를 최소화한 결과를 보여준다.

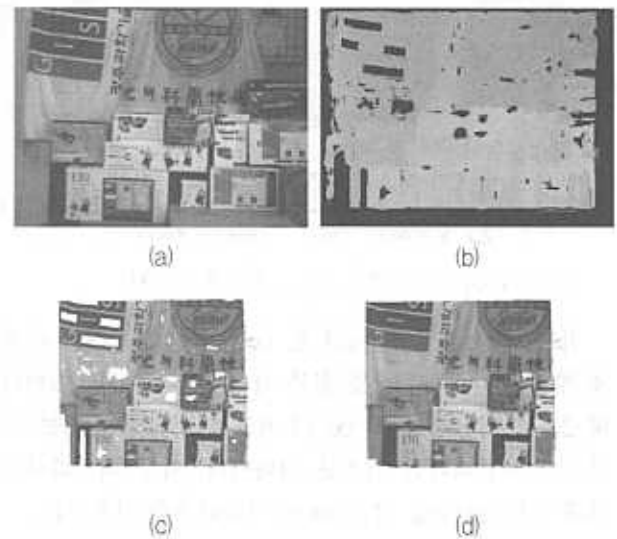


그림 11. 깊이 영상 정제 과정 (a) 원 영상 (b) 깊이 영상 (c) 깊이 영상 정제 과정 적용 전의 3차원 점군 (d) 깊이 영상 정제 과정 적용 후의 3차원 점군

Fig. 11. Depth image refinement (a) original image (b) depth image (c) 3D point cloud before depth image refinement (d) 3D point cloud after depth image refinement.

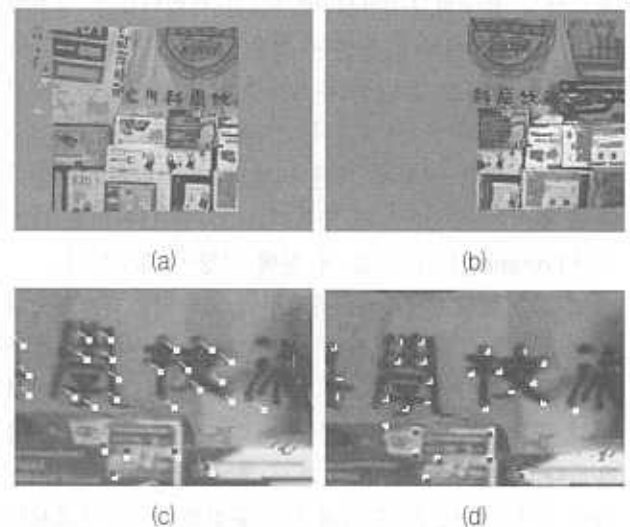


그림 12. 대응되는 특징 사이의 거리 오차 최소화 (a) 대상 시점의 3차원 점군의 대상 시점에서의 투영 (b) 원본 시점의 3차원 점군의 대상 시점에서의 투영 (c) 영상 (a)의 특정 영역 확대 (d) 거리 오차를 최소화 한 후에 영상 (b)의 특정 영역 확대

Fig. 12. Error minimization between corresponding features (a) projection of destination view data onto its own view (b) projection of source view data to the destination view (c) enlarged area of (a) (d) enlarged area of (b) after error minimization.

즉, 그림 12(a)와 그림 12(b)의 영상 간에 수정된 KLT 특징 추적기를 적용하여 그 사이의 거리를 최소화 하도

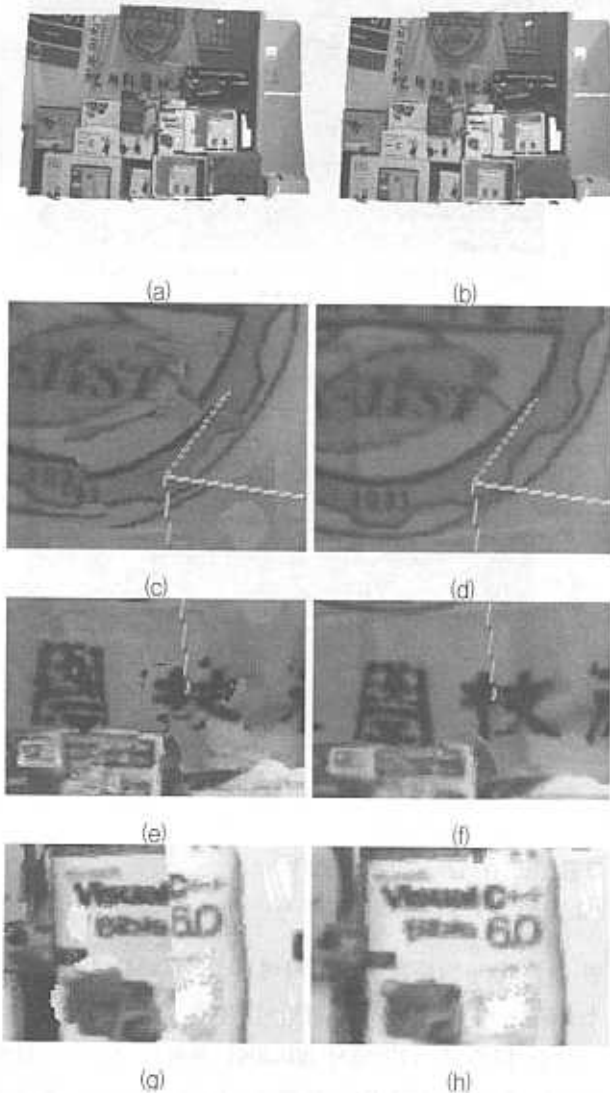


그림 13. 정합 결과 (a) 결합된 3차원 점군 (b) 정합된 3차원 점군 (c) 영상 (a)의 특정 부분 확대 1 (d) 영상 (b)의 특정 부분 확대 1 (e) 정합된 3차원 점군 (c) 영상 (f)의 특정 부분 확대 2 (d) 영상 (b)의 특정 부분 확대 2 (g) 정합된 3차원 점군 (c) 영상 (h)의 특정 부분 확대 3 (d) 영상 (b)의 특정 부분 확대 3

Fig. 13. Registration results (a) combined 3D point cloud (b) registered 3D point cloud (c) enlarged area 1 of (a) (d) enlarged area 1 of (b) (e) enlarged area 2 of (a) (f) enlarged area 2 of (b) (g) enlarged area 3 of (a) (h) enlarged area 3 of (b).

록 하였다. 그림 12(a) 및 그림 12(b)는 초기 위치 및 최종 위치를 나타낸 것이다. 그리고 그림 12(c) 및 그림 12(d)는 각 영상의 특정 부분을 확대한 것이다. 결과에 보인 것과 같이 대응점의 거리를 효과적으로 최소화하였음을 알 수 있다.

그림 13은 정합 과정을 나타낸 것으로 그림 13(a)는

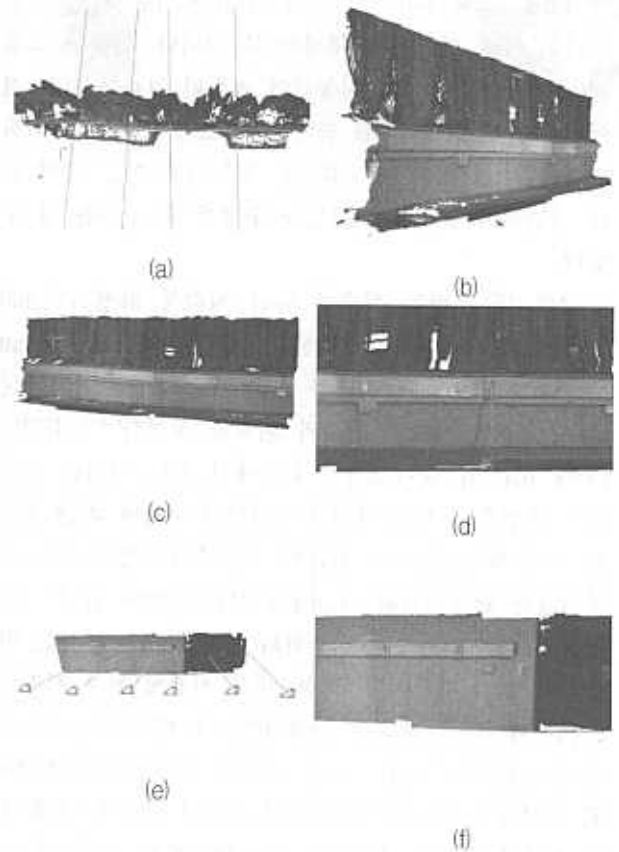


그림 14. 실내환경 복원 (a) 정합된 4 시점에 대한 위쪽 장면 (b) 왼쪽 장면 (c) 정면 장면 (d) 영상 (c)의 특정 영역 확대 (e) 정합된 6 시점에 대한 정면 장면 (f) 줌인 장면

Fig. 14. Indoor scene reconstruction (a) top view of 4 registered views (b) left view (c) front view (d) enlarged area of (c) (e) front view of 6 registered views (f) zoomed-in scene.

양쪽에서 획득된 3차원 점군을 결합한 것이다. 그리고 그림 13(b)는 제안된 방법을 사용하여 정합을 수행한 결과를 보인다. 반면, 그림 13(c) 및 그림 13(d), 그림 13(e) 및 그림 13(f) 그리고 그림 13(g) 및 그림 13(h)는 각 영상의 일부분을 확대한 것을 나타낸 것이다. 실험 결과에서 보는 것과 같이 정합 전후를 비교해 보면 원의 테두리 모양, 한자나 영어 등을 살펴보면 정합이 효과적으로 수행되었음을 확인할 수 있다.

그림 14는 제안된 방법을 실내의 커튼이 포함된 단열면에 대해 수행한 결과를 보인 것이다. 즉, 한 대의 다시점 카메라를 벽면을 따라 위치시켜가며 2차원 영상 및 3차원 깊이 정보를 획득한다. 획득된 데이터에 제안된 방법을 적용함으로써 다음과 같은 결과를 얻을 수 있다. 그림 14(a)는 카메라를 4회 이동하여 촬영한 후, 벽면에 대한 정합 결과를 위에서 본 장면이다. 반면, 그

림 14(b), 그림 14(c) 및 그림 14(d)는 이를 왼쪽, 정면 그리고 특정 영역을 확대해서 본 것이다. 그리고 그림 14(e)는 카메라를 6회 이동하여 정합된 결과를 보인 것이며, 그림 14(f)는 이를 줄인 하여 보인 것이다. 다시점 카메라를 통해 3차원 점군을 획득하고 이를 통해 실내 장면에 대한 3차원 복원을 수행할 수 있음을 알 수 있다.

그림 15는 반복 횟수에 따라 시각적 화질 (Visual Quality)을 나타내는 최대 신호 대 잡음비 (Peak Signal to Noise Ratio: PSNR)를 기반으로 기존의 ICP 및 Color ICP와 제안된 방법과 비교한 것이다^{[1][2]}. ICP나 Color ICP와 같이 3차원 공간에서 최단 거리를 통해 찾는 방법은 다시점 카메라의 양안차 추정에 의한 오차로 인해 획득된 3차원 점군에 적용하기 어렵다. 이러한 방법들은 단지 3차원 공간에서 점들 사이의 최단 거리만을 고려하고 이웃한 점들과의 관계에 대한 정보는 이용하지 않기 때문이다. 또한, 대상 시점에 원본 시점의 3차원 데이터를 단순히 투영시키므로 대응되는 화소를 찾기 어렵다는 단점이 있다. 그러나 제안된 방법은 2차원 영상에서 일정한 탐색 영역 내에서 블록에 대해 주변 정보를 이용해 상관도가 높은 대응되는 특징점을 찾는다. 그리고 그 거리를 최소화하기 때문에 보다 신뢰할 만한 결과를 보임을 실험 결과를 통해 알 수 있다. 성능 비교를 위해 겹치는 부분에 대해 다음과 같은 척도를 사용하였다.

$$PSNR = 20 \log_{10} \frac{255}{\sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (Y_{Src,i} - Y_{Dst,i})^2}} \quad (dB) \quad (17)$$

여기서, N 은 양쪽 영상에서 동시에 유효한 화소의 개수를 나타내며, $Y_{Src,i}$ 및 $Y_{Dst,i}$ 는 원본 및 대상 시점에서의 데이터를 투영한 영상 평면에서의 i 번째 대응되는 특징점의 휘도 값을 나타낸다.

그림 15에서 관찰할 수 있는 것과 같이 기존의 ICP 및 Color ICP 방법에 비해 제안된 방법에서의 시각적 화질이 개선되었음을 알 수 있다. 또한, 색 정보만을 사용한 방법에 비해 색과 질감 정보를 동시에 사용하는 방법 ($\alpha = 7.0$, $N_B = 192$)의 시각적 화질이 더 우수함을 알 수 있다. 그러나 색과 질감 정보를 사용하는 방법과 비교해 특징점을 사용하는 방법은 보다 빠른 수렴 속도를 얻을 수 있다. 기존의 ICP 및 Color ICP에 비해 제안된 방법들이 수렴 속도가 느린 것처럼 보이지만 제안된 방법들은 2차원 영상 평면에서 대응점을 찾는 반

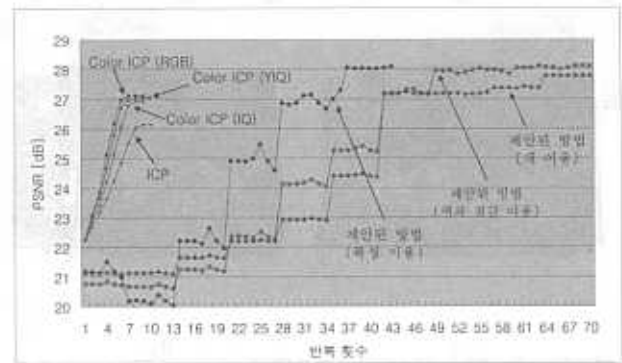


그림 15. 수렴율의 비교

Fig. 15. Comparison of convergence rate.

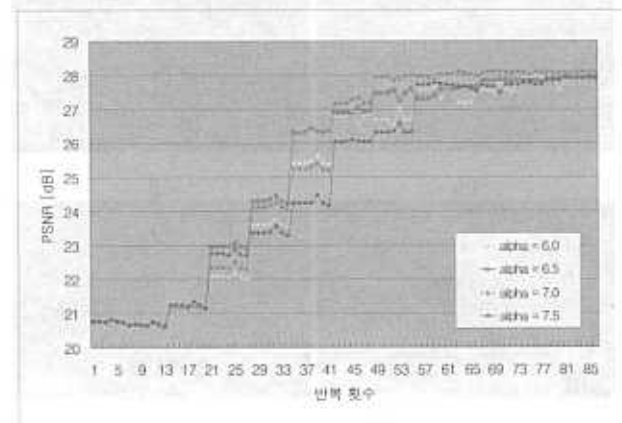


그림 16. α 값에 따른 PSNR

Fig. 16. PSNR according to α .

면 기존의 방법들은 3차원 상에서 찾기 때문에 한 회에 소요되는 시간은 제안된 방법이 적게 소요된다.

그림 16은 식 (13)에서 40×40 인 블록이 $N_B = 192$ 개 일 때, α 에 따른 PSNR의 변화를 나타낸 것이다. 결과에서 보는 것과 같이 α 값이 달라짐에 따라 수렴 속도가 서로 달라질 뿐만 아니라 수렴 가능한 PSNR 역시 서로 다름을 알 수 있다. 결과에서 보이는 것과 같이 α 가 7.0 일 때 수렴 속도가 가장 빠르고 PSNR도 가장 높음을 알 수 있다. 그러나 결과가 가장 우수한 α 값을 찾기 위한 최적화 방법이 요구된다.

IV. 결론 및 추후 과제

본 논문에서는 실내환경에 대한 3차원 복원을 수행하기 위해 다시점 카메라를 사용하여 색과 깊이 정보를 이용한 정합 방법을 제안하였다. 제안된 방법을 사용함으로써 깊이 정보가 정확하게 획득되지 않는 부분이나 잡음에 의해 양안차 추정의 결과가 정확하지 않는 경우 등에 있어 효과적으로 사용될 수 있음을 알 수 있다. 또

한 단순히 복원하고자 하는 환경에 대해 수 장면에 대한 획득된 정보만으로도 효과적인 복원이 가능함을 보였다. 추후에는 정합을 수행하기 소요되는 시간을 단축하고 실내의 모든 면에 대한 복원을 수행하기 위해 전역 최적화 (Global optimization)을 수행할 예정이다. 또한, 복원된 환경 내에 새로운 가상 객체를 효과적으로 삽입하기 위해 방법에 대한 연구를 진행할 예정이다.

참고 문헌

- [1] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. on Pattern Recognition and Machine Intelligence*, vol. 14, no. 2, pp. 239-256, 1992.
- [2] A. Johnson and S. Kang, "Registration and Integration of Textured 3-D Data," Tech. report CRL96/4, Digital Equipment Corporation, Cambridge Research Lab, 1996.
- [3] S. Kang and R. Szeliski, "3-D scene data recovery using omnidirectional multibaseline stereo," Tech. report CRL-95-6, Oct 1995.
- [4] G. Blais and M.D. Levine, "Registering multiview range data to create 3-D computer objects," *IEEE Trans. PAMI*, vol. 17, no. 8, pp. 820-824, 1995.
- [5] M. Levoy, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson, J. Davis, J. Ginsberg, J. Shade, and D. Fulk. "The digital Michelangelo project: 3D scanning of large statues," *SIGGRAPH'00*, pp. 131-144, July 2000.
- [6] T. Kanade and P. Rander, "Virtualized Reality: Constructing Virtual Worlds from Real Scenes," *IEEE MultiMedia*, vol.4, no.1, pp.34-47, Jan. - Mar. 1997,
- [7] K. Nishino and K. Ikeuchi, "Robust Simultaneous Registration of Multiple Range Images Comprising A Large Number of Points," *ACCV2002*, 2002.
- [8] Kari Pulli, *Surface Reconstruction and Display from Range and Color Data*, Ph.D. dissertation, University of Washington, 1997.
- [9] G. C. Sharp, S. W. Lee and D. K. Wehe, "Invariant Features and the Registration of Rigid Bodies," *IEEE Int'l Conf., on Robotics and Automation*, pp. 932-937, 1999.
- [10] VIDERE DESIGN, <http://www.videredesign.com/>, 2004.
- [11] R. 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. 3, no. 4, pp. 323-344, 1987.
- [12] K. Kim and W. Woo, "Multi-view Camera Tracking for Modeling of Indoor Environment," *Lecture Notes in Computer Science 3331*, pp. 288-297, 2004.
- [13] S. Kim, K. Kim and W. Woo, "Projection-based Registration using Color and Texture Information for Virtual Environment Generation," *Lecture Notes in Computer Science 3331*, pp. 434-443, 2004.
- [14] B. S. Manjunath and W. Y. Ma. "Texture features for browsing and retrieval of large image data," *IEEE Trans on PAMI*, vol. 18(8), pp. 837-842, 1996.
- [15] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, March 2004.

저 자 소 개



김 세 환(학생회원)
 1998년 서울 시립대 전자공학과
 학사 졸업.
 2000년 광주과학기술원 정보통신
 공학과 석사 졸업.
 2000년~현재 광주과학기술원
 정보통신공학과
 박사 과정.

<주관심분야: Virtual/Mixed Reality, 3D Vision>



우 운 택(중신회원)
 1989년 경북대학교 전자공학과
 학사 졸업.
 1991년 포항공과대학교 전기전자
 공학과 석사 졸업.
 1998년 University of Southern
 California, Electrical
 Engineering-System
 박사 졸업.

1991년~1992년 삼성종합기술연구소 연구원
 1999년~2001년 ATR MIC Lab., Japan,
 초빙 연구원
 2001년~현재 광주과학기술원 정보통신공학과
 조교수

<주관심분야: Virtual/Mixed Reality, 3D Vision, 3D Display, HCI, Networked VR, Ubiquitous/Wearable Computing>