

# 다시점 카메라로부터 획득된 깊이 및 컬러 영상을 이용한 실내환경의 파노라믹 3D 복원\*

김세환<sup>1</sup>, 우운택<sup>2</sup>  
광주과학기술원 U-VR 연구실<sup>1,2</sup>  
{skim<sup>1</sup>, wwoo<sup>2</sup>}@gist.ac.kr

## Panoramic 3D Reconstruction of an Indoor Scene Using Depth and Color Images Acquired from A Multi-view Camera

Sehwan Kim<sup>1</sup> and Woontack Woo<sup>2</sup>  
GIST U-VR Lab.<sup>1,2</sup>

### 요약

본 논문에서는 다시점 카메라로부터 획득된 부분적인 3D 점군을 사용하여 실내환경의 3D 복원을 위한 새로운 방법을 제안한다. 지금까지 다양한 양안차 추정 알고리즘이 제안되었으며, 이는 활용 가능한 깊이 영상이 다양함을 의미한다. 따라서, 본 논문에서는 일반화된 다시점 카메라를 이용하여 실내환경을 복원하는 방법을 다룬다. 첫 번째, 3D 점군들의 시간적 특성을 기반으로 변화량이 큰 3D 점들을 제거하고, 공간적 특성을 기반으로 주변의 3D 점을 참조하여 빈 영역을 채움으로써 깊이 영상 정제 과정을 수행한다. 두 번째, 연속된 두 시점에서의 3D 점군을 동일한 영상 평면으로 투영하고, 수정된 KLT (Kanade-Lucas-Tomasi) 특징 추적기를 사용하여 대응점을 찾는다. 그리고 대응점 간의 거리 오차를 최소화함으로써 정밀한 정합을 수행한다. 마지막으로, 여러 시점에서 획득된 3D 점군과 한 쌍의 2D 영상을 동시에 이용하여 3D 점들의 위치를 세밀하게 조절함으로써 최종적인 3D 모델을 생성한다. 제안된 방법은 대응점을 2D 영상 평면에서 찾음으로써 계산의 복잡도를 줄였으며, 3D 데이터의 정밀도가 낮은 경우에도 효과적으로 동작한다. 또한, 다시점 카메라를 이용함으로써 수 시점에서의 깊이 영상과 컬러 영상만으로도 실내환경 3D 복원이 가능하다. 제안된 방법은 네비게이션 뿐만 아니라 상호작용을 위한 3D 모델 생성에 활용될 수 있다.

Keyword : Indoor scene reconstruction, multi-view camera, virtual environment generation, projection-based registration

## 1. 서론

실세계 환경의 영상 기반 3D 복원 (Image-based 3D Reconstruction)은 사용자가 네비게이션하며 상호작용 하도록 함으로써, 시각적인 사실감을 느끼며 가상환경을 체험하는데 중요한 역할을 한다. 복원된 3D 모델의 시각적인 사실감은 사용자의 가상환경과의 적극적인 상호작용에 도움을 준다. 기존의 3D 모델링 툴 또는 능동적 거리 (Active Range) 측정 기법에 기반한 방법에 비해, 영상 기반 3D 모델 복원 방법은 실사를 사용하므로 사실감을 보존할 뿐만 아니라 모델링 과정도 간단하다.

특히, 일반 상용의 다시점 카메라 (Multi-view Camera)는 컬러 영상 뿐만 아니라 깊이 (Depth) 영상도 동시에 제공하므로 실사 기반 모델링을 보다 용이하게 한다. 이러한 카메라를 통해 실세계 환경의 3D 복원을 위해서는 여러 시점에서 획득된 3D 점군 (Point Cloud)을 정교하게 정합 (Registration) 및 병합 (Integration) 방법이 요구된다.

그 동안 실세계 환경의 3D 복원을 위한 다양한 방법들이 제안되었다. Besl 등에 의해 제안된 ICP (Iterative Closest Point) 알고리즘은 현재까지도 널리 사용되며, Johnson 등은 색 정보를 결합한

\* 본 연구는 광주과학기술원 문화기술 연구센터 (CTRC)의 지원에 의한 것임

Color ICP 를 제안하고 이를 통해 실내환경을 복원하였다 [1][2]. 특히, Park 등은 양안차 추정 과정에서 발생하는 오차를 제거하는 방법을 제안하고 실사 기반 3D 모델을 생성하였다 [3]. Pulli 는 데이터 획득 장치를 개발하고 평면의 원근 변형 (Planar Perspective Warping)을 이용한 투영 정합 방법을 제안하여 물체들을 모델링하였다 [4]. 반면, Voxel Carving 이나 Level Set 과 같은 Volumetric 방법들은 3D 공간을 양자화하고 각 복셀의 점유 여부를 판별하는 방식으로 모델을 생성한다 [5][6][7]. 이러한 방법은 임의의 모양에 대한 표현은 가능하지만, 모델의 해상도가 초기 양자화에 의해 결정된다는 단점이 있다. 화소 기반 PDE 접근 방법은 양자화에 의존하지는 않지만, 매 화소마다 연속적인 깊이를 계산해야 한다는 문제점이 있다 [8]. 이와는 달리 메쉬를 이용한 방법들은 세밀한 형상의 표현이 가능하지만, Self-intersections 과 Topological Change 를 다루기가 어렵다는 문제점이 있다 [9]. 그리고 Wide-baseline 스테레오 기법에 기반한 확률적인 접근 방법도 연구되고 있다 [10][11].

본 논문에서는 다시점 카메라로부터 획득된 깊이 및 컬러 영상을 이용하여 실세계 환경을 복원하는 방법을 제안한다. 우선, 동일한 시점에서 다시점 카메라로부터 획득된 여러 프레임의 깊이 영상의 시공간적 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 3D 점군을 동일한 영상 평면에 투영하고 수정된 KLT (Kanade-Lucas-Tomasi) 특징 추적기 (Feature Tracker)를 통해 대응점을 찾는다. 그리고 거리 오차를 최소화함으로써 두 집합의 3D 점군에 대한 정밀한 정합을 수행한다. 마지막으로, 여러 시점에서 획득된 3D 점군들과 각 쌍의 2D 영상을 동시에 이용하여 3D 점들의 위치를 세밀하게 조절함으로써 실세계 환경을 복원한다.

제안된 방법은 정밀도가 상대적으로 낮은 3D 데이터에 대해서도 효과적인 정합이 가능하다. 또한, 3D 모델의 정합 과정을 2D 영상 평면에서 수행함으로써 보다 빠른 정합이 가능하다. 마지막으

로, 다시점 카메라를 실내환경의 임의의 몇 곳에 위치시켜 장면에 대한 3D 복원을 가능하게 함으로써 비교적 용이한 가상환경 생성이 가능하다.

본 논문의 구성은 다음과 같다. 2 장에서는 깊이 영상 정제 과정에 대해 기술하고, 3 장에서는 투영 기반 정합 방법에 대해 설명한다. 그리고 4 장에서는 병합 방법에 대해 다룬다. 5 장에서 실험 결과 및 분석에 대해 기술한 뒤, 6 장에서 결론 및 향후 연구에 대해 언급한다.

## 2. 깊이 영상 정제

깊이 영상 정제 과정은 깊이 영상에 존재하는 오차가 있는 점들을 깊이 값들의 시공간적 특성을 이용하여 보완한다. 본 과정은 두 단계로 나뉘며, 첫 번째 단계에서는 깊이 값들이 시간에 따라 변화하는 시간적 특성을 이용하여 오차를 갖는 점을 제거하며, 두 번째 단계에서는 현재 3D 점이 주변의 점들과 공간적인 상관관계를 갖는다는 공간적 특성을 이용하여 빈 영역 (Hole)을 채운다. 그림 1 은 실세계 환경에 대한 3D 복원의 흐름도이다.

3D 점군의 특성을 분석하기 위해, 동일한 시점에서  $N_f$  장의 깊이 영상을 획득한 후에 각 화소에 대해 평균과 표준편차를 계산한다. 그러나 움직임이 없는 정적인 장면에 대해서조차 동일한 화소에 대응되는 깊이 값이 시간에 따라 큰 변화를 갖는 것들이 존재한다.

깊이 영상에서의 각 화소에 대한 오차의 변화량을 가우시안 분포 (Gaussian Distribution)로 모델링하고,  $i$  번째 화소에 대한 깊이 값의 변화량이 문턱값 (Threshold Value)보다 크면 유효하지 않은 값으로 결정한다.

$$\sigma_i > \alpha Th_i(x_c, y_c, z_c) \quad (1)$$

위 식에서,  $\sigma_i$  는  $i$  번째 화소의 깊이 변화량의 표준 편차를 나타낸다.  $\alpha$  는 비례 인자 (Scale Factor),  $Th_i$  는  $i$  번째 화소에 대한 문턱값을 나타낸다. 그리고  $(x_c, y_c, z_c)$  는 영상 평면의 현재 화소에 대응되는 3D 공간 상의 좌표를 나타낸다. 그러나 문턱값은 카메라로부터의 상대적인 위치에 의해 영향을 받는다. 다시점 카메라의 특성에 따라 카메라로부터 물체가 멀어질수록 양안차 추정시 오차가

증가하기 때문이다. 따라서  $Th_i$  는 카메라로부터의 위치에 따른 함수로 표현되어야 한다.

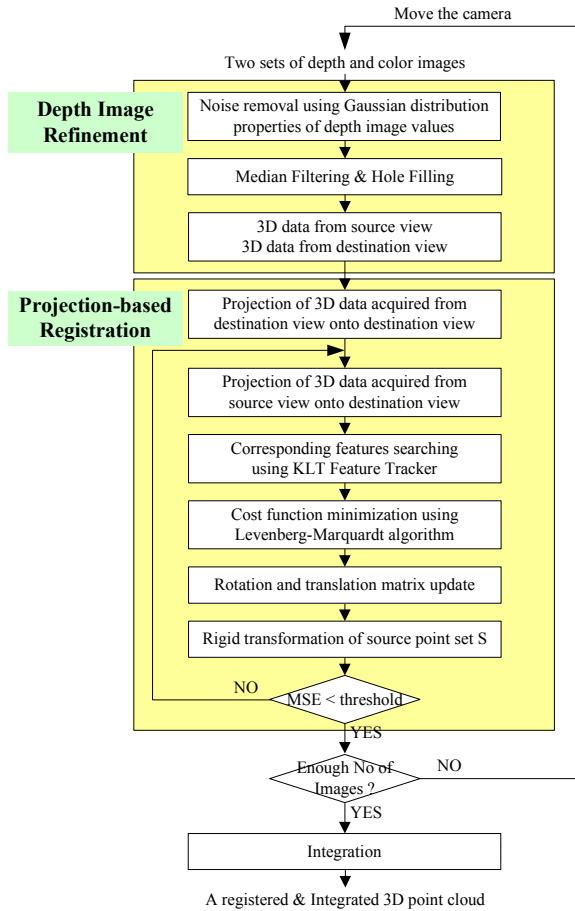


그림 1. 실세계 환경 복원을 위한 흐름도

카메라로부터의 위치를 고려하여  $Th_i$  를 결정하기 위해 카메라의 시선 (Line of Sight: LOS)에 따라 각 축 방향으로의 거리에 따른 오차를 계산한다. 이렇게 3D 공간에서 거리에 따라 적응적으로 변하는 유효한 공간은 3D 공간에서 타원체 모양을 가지며 적응적 불확정 영역 (Adaptive Uncertainty Region)이라 명명한다.

이를 기반으로 다시점 카메라에 적용하면, 거리가 카메라로부터 멀어짐에 따라  $x, y$  방향으로는 선형적으로 증가하고,  $z$  축 방향으로는 곡선에 따라 비례적으로 증가하는 가우시안 분포를 얻는다 [12][13]. 다시점 카메라로 획득된 영상의 각 화소에 대한 3D 점이 주어지며, 이는 3D 공간에서 거리에 따라 변하는 타원체 모양의 불확정 영역을 생성한다. 따라서 거리에 따라 적응적으로 변하는 불확정 영역이 고려된  $Th_i$  가 결정된다. 그러나 위의 모양은 카메라의 광학 중심을 기준으로 화각의 방향에 대해 일정한 각도만큼 회전되어야 한다.

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

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

$$R_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & z_c/d & y_c/d \\ 0 & -y_c/d & z_c/d \end{pmatrix} \quad (3)$$

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

위 식에서,  $(x_c \ y_c \ z_c)^T$  는 영상 평면의 각 화소에 대응되는 3D 공간 상의 좌표로서 타원체 모양의 중심이 카메라의 광학 중심 (Optical Center)에서 이동한 이동 벡터 (Translation Vector)를 나타낸다.  $(x' \ y' \ z')^T$  는  $(x_c \ y_c \ z_c)^T$  를 고려하여 카메라의 광학 중심을 기준으로 분포하게 될 카메라로부터의 위치에 따른 오차 분포에 대한 최종적인 적응적 불확정 영역을 나타낸다.  $\Delta x, \Delta y$  그리고  $\Delta z$  는 각 축으로의 불확정 거리 (Uncertainty Distance)를 나타낸다. 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 채움과정이 요구된다. 즉, 현재 3D 점에 대해 주변 점들과의 공간적인 상관관계를 이용하여 위의 과정을 수행한다 [12][13].

### 3. 투영 기반 정합

수동적 기법 (Passive Technique)으로 획득된 깊이 영상은 오차가 크기 때문에 ICP 를 이용한 정합 과정이 용이하지 않다. 이는 3D 점군이 큰 오차를 포함하므로 최단 거리에 기반한 ICP (또는 Color ICP)는 많은 오류를 야기시키기 때문이다 [1][2]. 따라서 원본 시점 (Source Viewpoint)의 3D 데이터의 대응점을 대상 시점 (Destination Viewpoint)의 3D 데이터에서 효과적으로 찾기 위해 투영 기반 정합 방법을 제안한다. 카메라는 실내 공간의 벽을 따라 움직이며 부분적인 3D 점군을 획득한다. 대상 및 원본 시점은 카메라의 이전 및 현재 위치에서의 카메라 시점을 의미한다.

초기 정합 (Initial Registration) 단계에서는 다시

점 카메라를 움직이면서 추적되는 특징점에 대한 3D 정보를 기반으로 강체 변환 (Rigid-body Transformation)을 수행한다 [14][12]. 또한 반수동 방법과 같은 임의의 방법이 사용될 수도 있다 [4].

정밀한 정합 (Fine Registration) 단계에서는 겹치는 부분에서의 대응되는 특징점을 이용한다. 대상 시점을 기준으로 원본시점의 외부 파라미터를 반복적으로 조정함으로써 두 개의 부분적인 3D 점군을 정합한다. 즉, 원본 표면 (Source Surface)에 유클리드 변환 (Euclidean Transformation)  $T: \mathbb{R}^3 \rightarrow \mathbb{R}^3$  을 적용한다. 대상 표면 (Destination Surface)  $S_{Dst}$  은 대상 시점의 영상 평면으로 투영되며, 특징점  $f_{Dst}$  가 추출된다. 반면, 매 반복 회수마다 원본 표면  $S_{Src}$  는 대상 시점의 영상 평면으로 투영되며 대응되는 특징점  $f_{Src}$  '을 찾는다. 이를 그림 2에 나타내었다.

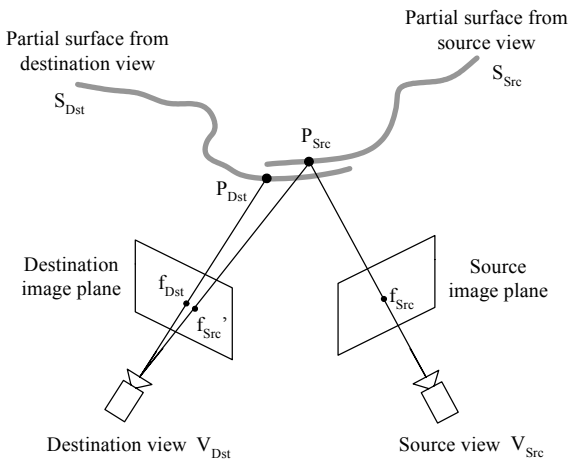


그림 2. 대응되는 특징점의 추출

대상 영상의 각  $f_{Dst}$  에 대해 원본 영상에서의  $f_{Src}$  '은 수정된 KLT 특징 추적기를 사용하여  $f_{Dst}$  의 주변에서 검색한다 [15].  $P_{Dst}$  및  $P_{Src}$  는  $f_{Dst}$  및  $f_{Src}$  의 3D 점들이다. 그리고  $c_{Dst}$  및  $c_{Src}$  는  $f_{Dst}$  및  $f_{Src}$  의 컬러를 나타낸다.

우선, 수정된 KLT 특징 추적기를 이용하여 대상 영상의 겹치는 영역  $\Omega$ 에서 특징점들을 찾는다. 그리고  $S_{Dst}$  의 투영에 사용된 동일한 카메라 파라미터를 사용하여  $S_{Src}$  를 대상 영상 평면으로 투영한다. 대응점들은 하위 화소 (sub-pixel) 단위로 투영된 원본 영상에서 찾는다. 이 때, 대응되는 특징점을 찾는데 발생할 수 있는 오류, 즉 이상치 (Outlier)를 제거하고 올바른 쌍만을 추출한다.

$S_{Src}$  를 대상 시점으로 투영하면 영상  $I_{Src}$  '가 생성된다. 따라서  $I_{Src}$  '와 대상 영상  $I_{Dst}$  사이의 불일치를 측정하는 비용함수를 다음과 같이 정의한다.

$$L = \sum_{i=1}^{N_{feat}} \kappa_1 \left\{ \left( 1 - \frac{\|f_{Dst,i} - f_{Src,i}\|}{Dist_{ff}} \right) \|f_{Dst,i} - f_{Src,i}\|^2 + \kappa_2 \|c_{Dst,i} - c_{Src,i}\|^2 \right\} \quad (4)$$

위 식에서,  $\|\cdot\|$  및  $Dist_{ff}$  는 놈 (Norm),  $f_{Dst}$  와  $f_{Src}$  ' 사이의 거리를 고려하여 미리 정의된 값이다.  $\kappa_2$  는 컬러 정보를 위한 가중치,  $N_{feat}$  는 특징점의 개수를 나타낸다. 특히,  $\kappa_1$  는 3D 공간에서 각 쌍의 거리가 미리 정의된 문턱값  $Th$  을 넘는 경우 이를 제거하기 위해 다음과 같이 정의된다. 즉,  $P_{Dst}$  와  $P_{Src}$  사이의 거리가 큰 경우에는 비용함수를 계산할 때 제외된다. 그렇지 않으면, 각 쌍의 거리에 따라 가중치가 결정된다.

$$\kappa_1 = \begin{cases} \frac{\|P_{Dst} - P_{Src}\|}{Dist_{PP}} & \text{if } \|P_{Dst} - P_{Src}\| < Th \\ 0 & \text{o/w} \end{cases} \quad (5)$$

위 식에서,  $Dist_{PP}$  는  $P_{Dst}$  와  $P_{Src}$  사이의 거리를 고려하여 미리 정의된 값이다.

영상 평면의 겹치는 부분에 대해 비용 함수를 정의하기 위해 특징점을 찾고 대응되는 특징점들 사이의 유클리드 거리 (Euclidean Distance)를 사용하여 카메라의 자세를 추정한다. 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 추정된다. 이는 다음과 같이 표현된다.

$$\begin{aligned} & \text{Given two sets of corresponding points,} \\ & \text{Find } \{R_{Src}, T_{Src}\} \text{ w.r.t } \{R_{Dst}, T_{Dst}\} \\ & \text{such that } \arg \min_{\{R_{Src}, T_{Src}\}} L \end{aligned} \quad (6)$$

이를 최소화하기 위해 Levenberg-Marquardt 알고리즘을 사용한다.

#### 4. 병합

정합 수행 후, 실세계 공간에서는 물리적으로 동일한 점인데도 복원된 3D 공간에서는 양안차 추정 오차 때문에 동일한 점을 점유하지 못한다. 따라서 물리적으로 동일한 3D 점들은 동일한 3D 좌표를 갖도록 3D 점들을 조정하는 과정이 요구된다. 이러한 과정의 수행 후, 카메라의 위치에 따

라 서로 다른 조명 조건의 변화를 고려하기 위해 컬러 조정 (Color Adjustment) 과정이 요구된다.

우선, 복원된 3D 공간에서 서로 인접하게 위치하도록 3D 점들을 조정한다. 이를 위해 여러 시점에서 획득된 3D 점군들과 함께 각 쌍의 영상을 동시에 이용하여 깊이 영상 향상 (Depth Image Improvement) 과정을 수행한다. 즉, 여러 시점에서 획득된 영상의 쌍들과 함께 미리 획득된 3D 점군들을 동시에 사용함으로써 복원되는 3D 모델의 질을 향상시킨다.

일반적으로 영상 획득을 위한 측정 (Measurement) 과정은 언제나 잡음에 의해 왜곡된다. 3D 점군의 불확실성 (Uncertainty)은 복원되는 물체의 국부적인 특성 뿐만 아니라 전체적인 구조에도 영향을 미친다. 따라서 복원되는 구조에 대한 불확실성의 영향을 줄이고 여러 시점에서 획득된 점군들을 병합하여 하나의 3D 모델을 생성해야 한다. 이를 위해 3D 공간 뿐만 아니라 2D 공간도 동시에 고려하는데, 이 두 공간이 깊이 값을 결정하는데 서로 보완적인 역할을 한다는 사실에 기반한다.

그림 3을 통해 각 카메라 시점에서의 데이터 획득에 대한 한 예를 볼 수 있다. 여기서 각 카메라는 다시점 카메라를 나타내며, 비록 잡음은 포함되지만 각 시점에서 3D 점군과 함께 한 쌍의 좌우 영상을 얻는다. 최종 목적은 깊이 오차가 존재하는 3D 점군으로부터 이미 보정되어 카메라의 내부/외부 파라미터를 알고 있는 3D 점군과 영상의 쌍들을 이용하여 실세계 환경에 대한 정확한 3D 표현 방법을 찾아내는 것이다. 따라서, 입력 영상의 각 화소에 대응되는 깊이 값을 추정하고자 한다. 이러한 깊이 영상들은 하나의 3D 모델로 병합된다. 최종적인 깊이 추정을 위해 다음과 같은 비용함수를 정의한다.

$$d^* = \arg \min_d E(d). \quad (7)$$

$$E(d) = E_{Data}(d) + \lambda_1 E_{Smoothness}(d)$$

$$E_{Data}(d^1) = \sum_{(x,y)} |I_L^1(x-d^1, y) + I_R^1(x, y)| + \alpha \sum_{j=2}^{N_V} |P_{L,d^1}^1 - P_{L,d^j}^j|, \quad P_{L,d^j}^j \in C_{L,d^j}^{1,0} \quad (8)$$

$$E_{Smoothness}(d^1) = \sum_{\eta} |d^1 - d_{\eta}^1| + \beta \sum_{i=1}^{N_C} \sum_{j=2}^{N_V} |C_{L,d^1}^{1,0} - C_{L,d^j}^{j,j}| \quad (9)$$

위 식에서,  $N_V$  는 뷰의 개수,  $N_C$  는 주변 육면체 (Neighboring Cube)의 개수를 나타낸다.  $P_{L,d^j}^j$  는 각 3D 점에 대한 컬러 정보를 의미한다.  $j$  는  $j$  번째 뷰,  $L/R$  은 왼쪽/오른쪽 영상을 의미하는데 사용된다. 그리고  $d^j$  는  $j$  번째 뷰에 대한 양안차 값을 의미한다. 반면,  $\#C_{L,d^j}^{j,j}$  는  $j$  번째 뷰에 대해  $i$  번째 육면체에 포함되는 3D 점들의 개수를 의미한다.

$d^*$ 는  $E(d)$ 를 최소화함으로써 구하고자 하는 최종적인 양안차 값이다. 비용 함수는 다음과 같이 두 가지로 구성된다: (i) Data Part  $E_{Data}$  그리고 (ii) Smoothness Part  $E_{Smoothness}$ . Data Part 는 양안차 값  $d$  가 여러 쌍들의 영상들 및 3D 점군들과 얼마나 잘 조화가 되는지를 측정한다. 반면, Smoothness Part 는 알고리즘에 의해 가정되는 부드러움의 정도를 내포한다.

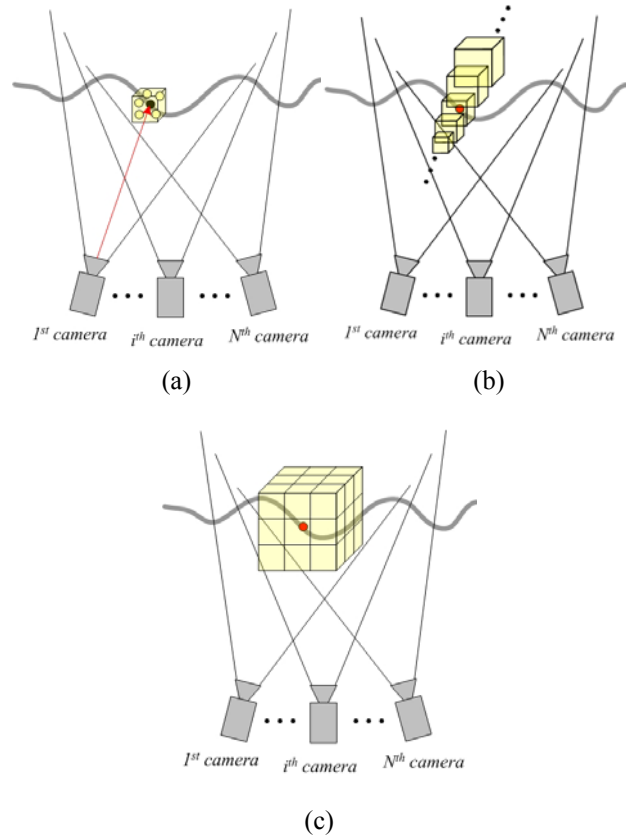


그림 3. 육면체 표현 (a) 기준 카메라의 현재 화소에 대한 육면체 (b) 양안차 값에 따른 가변 육면체 (c) 중심 육면체 및 주변 (비중심) 육면체

각 Part 의 첫 항은 일반적인 양안차 추정 알고리즘의 Data 및 Smoothness 에 대한 조건과 유사하

다. 다만, 양안차의 범위가 입력 데이터로 제공되는 양안차 값에 의해 결정된다는 사실만 다르다.

Data Part의 두 번째 항목은 3D 공간에서 3D 점들의 분포 (Constellation)를 분석한다. 이는 3D 점들의 상대적인 컬러 뿐만 아니라 상대적인 3D 위치까지도 동시에 비교한다. 그림 3(a)는 기준 카메라의 3D 점에 대해 생성된 육면체를 보인 것으로, 3D 공간에서 점들간의 공간적인 관계를 내포한다. 그림 3(a)가 기준 카메라의 현재 화소에 대한 육면체를 보인 반면, 그림 3(b)는 변화 가능한 양안차 값에 따른 가변적인 육면체를 보인 것이다.

Smoothness Part의 두 번째 항목은 주변 육면체에 포함되는 3D 점들을 고려하며, 각 육면체는 카메라의 중심으로부터 기준 시점의 현재 3D 점까지의 거리에 따라 가변된다. 그림 3(c)는 Smoothness Part에 대해 기준 카메라의 현재 화소에 대응되는 중심의 육면체 뿐만 아니라 주변 육면체 (비중심 육면체)를 동시에 나타낸 것이다.

## 5. 실험 결과 및 분석

본 실험은 일반적인 실내환경의 정상적인 조명 하에서 수행되었으며, Digiclops를 사용하였다 [16]. Digiclops는 IEEE 1394 방식의 다시점 카메라로서 동일한 평면 위에 있는 3개의 렌즈를 이용하여 양안차 추정을 통해 3D 정보를 획득한다. 카메라의 초점 거리  $f=6\text{ mm}$ , 기본선 (Baseline)  $B=10\text{ cm}$ 이다. 그리고  $7\times 5$ 개의 격자점, 격자간의 간격은  $10.6\text{ cm}$ 인 평면 패턴을 사용하였다.

### 5-1 가상 장면

우선, 깊이 영상 정제 과정 적용 전후의 오차 비교를 위해 ground truth로 사용될 3D 모델을 생성했다. 이상적인 데이터와 추정된 데이터를 비교하기 위해서는 적합한 가상 카메라의 생성이 필수적이다. 즉, 초점 거리, 주점 (Principal Point), 종횡비 (Aspect Ratio) 그리고 영상의 크기를 고려하여 가상 카메라를 생성해야 한다. 본 실험에서는 가상 뷰를 얻기 위해 그림 4에 보인 스크립트 파일이 사용되었다. 모든 파라미터는 실험에 사용된 Digiclops™와 동일하게 설정되었다.

```
#FOCAL LENGTH 843.859
#PRINCIPAL POINT 343.534 268.394
#ASPECT 1.0
#IMAGE WIDTH 640.0
#IMAGE HEIGHT 480.0
```

그림 4. 가상 카메라 생성을 위한 스크립트 파일

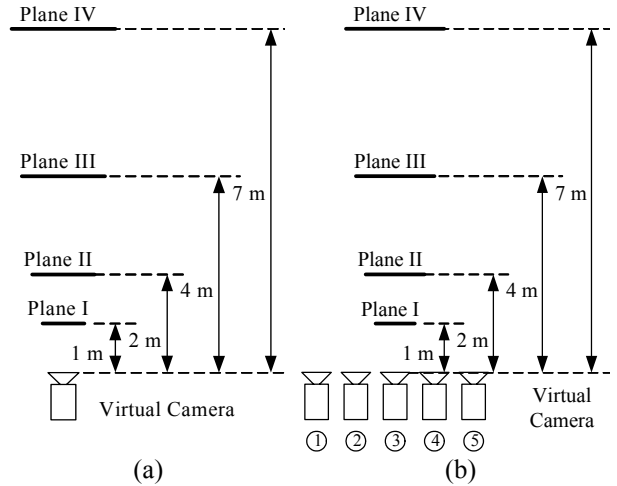


그림 5. 가상 장면 생성을 위한 실험 설정 (a) 깊이 영상 정제 과정 (b) 깊이 영상 향상 과정

그림 5는 특별히 디자인된 3D 모델에 대한 상위 뷰를 나타낸 것이다. 가상 장면은 네 개의 평면으로 구성된다. 평면간의 거리는 각각  $1\text{ m}$ ,  $2\text{ m}$ , 그리고  $3\text{ m}$ 이며, 평면의 크기는 카메라로부터 평면까지의 거리에 따라 증가한다.

표 1에서의 깊이 값은 그림 5에 보인 각 평면에 대해 평가되어 실제 깊이와 추정된 깊이가 거리에 따라 비교된다. 추정된 깊이는 기본선  $10\text{ cm}$ 를 갖는 가상 다시점 카메라를 통해 획득된 좌우 영상에 Triclops 라이브러리를 적용함으로써 계산된다 [16]. 공정한 비교를 위해, 각 평면에서 그 표본이 5,000개가 되도록 선택하였다.

표 1. 여러 평면에 대한 깊이 비교 (단위: m)

	True Depth $d_T$	Mean $d_E$	Std
Plane I	1.0	0.99625	0.00448
Plane II	2.0	1.99387	0.01274
Plane III	4.0	3.98951	0.02946
Plane IV	7.0	6.97237	0.08414

세 번째 열은 각 평면에 대해 추정된 깊이 값의 평균, 네 번째 열은 추정된 깊이 값들에 대한 표준 편차를 나타낸다. 추정된 깊이 값이 가상 카메라와 멀수록 부정확하다는 것을 관찰할 수 있다. 그리고 가상 카메라와 평면간의 거리가 멀수록 양안차 추정에 의해 발생하는 깊이 오차는 급격하게

증가한다. 각 평면에 대해 추정된 표준 편차는 가우시안 잡음을 갖는 가상 데이터를 생성하는데 사용된다.

깊이 영상 정제 과정의 성능을 증명하기 위해, 이상적 모델에 추정된 표준편차 (표 1)를 거리에 따라 적용함으로써 가상 장면을 생성하였다. 표 2에 그 결과 데이터를 보였다. 획득된 표준편차가 표 1에 보인 값들과 거의 동일함을 알 수 있다.

**표 2.** 깊이 영상 정제 과정의 성능을 보이기 위한 가상 깊이 영상의 평균과 표준 편차 (단위: m)

	Mean of Original Depth $d_M$	Std of Original Depth
Plane I	1.000042	0.004491
Plane II	2.000147	0.012672
Plane III	3.999663	0.029438
Plane IV	7.002396	0.084712

표 3과 표 4는 획득된 깊이 영상이 각각 다섯 프레임과 열 프레임일 때, 깊이 영상 정제 과정을 적용한 후에 정제된 깊이 영상의 평균과 표준 편차를 보인 것이다. 즉, 서로 다른 가우시안 잡음 분포를 갖는 다섯 프레임과 열 프레임의 가상 깊이 영상을 생성하고 깊이 영상 정제 과정을 수행하였다. 그림 6에 거리에 따른 결과의 표준 편차를 보였다. 본 과정을 수행한 후에 표준 편차의 상당한 개선이 있음을 알 수 있으며, 사용되는 영상의 개수가 많은 경우에 보다 개선되었음을 알 수 있다.

**표 3.** 다섯 프레임을 이용하여 깊이 영상 정제 과정을 수행한 후의 평균과 표준 편차 (단위: m)

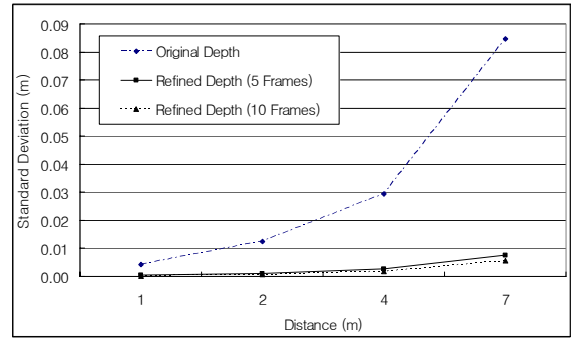
	Mean $d'_M$	Std
Plane I	0.999972	0.000491
Plane II	2.000100	0.001168
Plane III	4.000193	0.002611
Plane IV	7.001695	0.007560

**표 4.** 열 프레임을 이용하여 깊이 영상 정제 과정을 수행한 후의 평균과 표준 편차 (단위: m)

	Mean $d'_M$	Std
Plane I	1.000003	0.000373
Plane II	2.000200	0.000834
Plane III	4.000098	0.001972
Plane IV	7.001378	0.005658

정합 과정의 성능 증명을 위해 가상의 실내환경을 생성하고 텍스처 매핑을 적용하였다.

연속된 두 뷰 사이의 거리 및 벽과의 거리는 0.1 m와 3.5 m로 설정하였다. 회전 및 이동 벡터에 대한 오차의 표준 편차를 각 방향으로 0.3° 및 0.02 m로 설정하였다. 본 카메라 배치에서, pitch, yaw 또는 roll 방향으로의 회전 오차  $\pm 0.3^\circ$ 는  $\pm 0.018 \text{ m}$  ( $\approx \pm 3.5 \text{ m} \times \tan(0.3^\circ)$ )의 변화를 야기시킨다. 이 경우, 대응되는 특징점의 개수  $N_{feat} = 216$ 이다.



**그림 6.** 가상 깊이 영상에 대한 깊이 영상 정제 결과의 거리에 따른 표준 편차

표 5는 정합 후의 평균 회전 및 이동 오차를 보인 것이다. 카메라의 오차 범위를 고려하여 오차를 충분히 크게 설정하였음에도 정합 결과를 통해 제안된 방법이 효과적임을 알 수 있다.

**표 5.** 정합 후의 평균 회전 및 이동 오차

Rotation (degrees)	Pitch	Yaw	Roll
	-0.057912	-0.036512	-0.004023
Translation (meters)	x	Y	Z
	-0.000195	-0.0024234	+0.0008412

깊이 영상 향상 과정의 성능을 평가하기 위해 그림 5(b)에 보인 것과 같이 가상 카메라를 움직였다. 표 6과 표 7은 사용되는 뷰가 각각 세 개와 다섯 개일 때 향상된 깊이 영상의 평균과 표준 편차를 보인 것이며, 세 번째 카메라에 대해 평가되었다. 세 개의 뷰에 대해서는 두 번째, 세 번째 그리고 네 번째 카메라를 사용했다. 반면, 다섯 뷰에 대해서는 모든 가상 카메라를 사용하였다.

**표 6.** 세 뷰를 사용하여 깊이 영상 향상 과정을 수행한 후의 평균과 표준 편차 (단위: m)

	Mean $d''_M$	Std
Plane I	0.998469	0.004256
Plane II	1.996712	0.010300
Plane III	4.007860	0.018064
Plane IV	7.006391	0.030071



표 7. 다섯 뷰를 사용하여 깊이 영상 향상 과정을 수행한 후의 평균과 표준 편차 (단위: m)

	Mean $d''_M$	Std
Plane I	0.997560	0.003571
Plane II	1.994816	0.008200
Plane III	4.004060	0.013073
Plane IV	7.001172	0.024634

가상 깊이 영상에 대해 거리에 따른 깊이 영상 향상 결과의 표준 편차를 그림 7에 보였다. 본 과정을 수행한 후에 표준 편차의 상당한 개선이 있음을 알 수 있으며, 사용되는 카메라의 개수가 많은 경우에 보다 개선되었음을 알 수 있다.

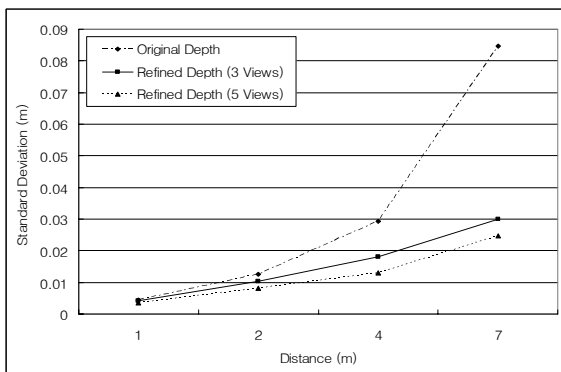


그림 7. 가상 깊이 영상에 대한 깊이 영상 향상 결과의 거리에 따른 표준 편차

### 5-2 실제 장면

제안된 방법은  $N_f = 20$ ,  $Th_{dd} = 0.15$  로 설정하여 실제 장면에 적용되었다. 그림 8은 깊이 영상 정제 과정의 결과를 보인 것이다.

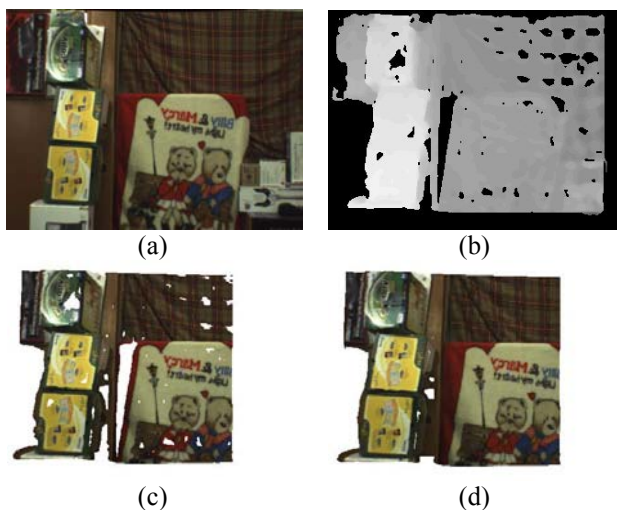


그림 8. 깊이 영상 정제 과정 결과 (a) 원 영상 (b) 깊이 영상 (c) 깊이 영상 정제 과정 전의 3D 점군 (d) 깊이 영상 정제 과정 후의 3D 점군

그림 8(a) 및 그림 8(b)는 원 영상과 깊이 영상을 나타낸다. 대응되는 3D 점군과 깊이 영상 정제 과정 후의 결과를 그림 8(c)와 그림 8(d)에 나타내었다. 본 예에서는 영상 오른쪽의 3D 오차 범위가 매우 커서 오른쪽 부분의 일부는 제외하였다. 결과에 보인 것과 같이 유효하지 않은 영역의 3D 점들을 제거하고 그 후에 빈 영역을 채움으로써 양안차 추정으로 인한 오차 부분을 효과적으로 제거하고 평면으로 추정되는 부분을 채웠음을 알 수 있다.

정합 결과를 그림 9에 보였으며, 제안된 방법의 정합 결과의 시각적 성능이 ICP보다 우수함을 알 수 있다. 그림 9(a) 및 그림 9(b)는 왼쪽과 오른쪽 영상을 보여준다. 초기 정합 후, 그림 9(c)에 보인 것과 같은 결과를 얻는데, 곰의 얼굴 부분 및 글자들이 번져있는 것을 알 수 있다. 그림 9(d) 및 그림 9(e)에 ICP와 제안된 방법의 정합 결과를 보였다. 최단거리 오차는 ICP가 제안된 방법보다 우수하다. 그러나 제안된 방법의 시각적인 성능이 ICP보다 우수하다는 것을 관찰할 수 있는데, 이는 ICP가 3D 데이터 자체가 아닌 최단 거리만을 고려하기 때문이다.

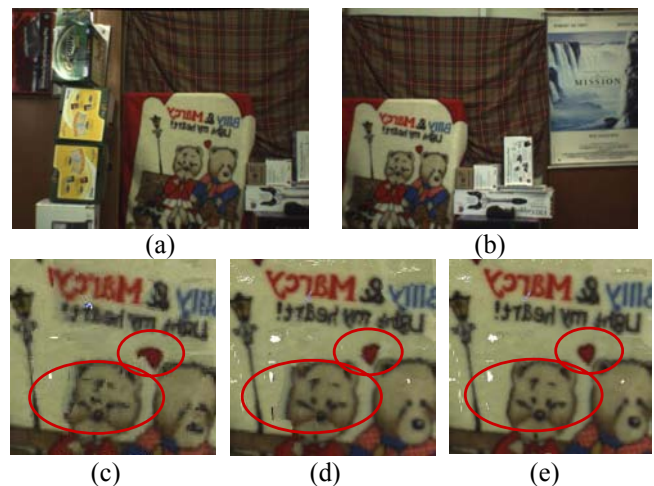


그림 9. 정합의 시각적인 성능 비교 (a) 좌 영상 (b) 우 영상 (c) 초기 정합 (d) ICP (e) 제안된 방법

그림 10은 다른 방법들과의 성능 비교를 보인 것이다. 본 실험에서 대상 시점과 원본 시점에서 획득된 3D 점들은 145,870 개와 189,341 개이다. 제안된 방법이 성능 뿐만 아니라 속도 면에서도 ICP with kd-tree나 Color ICP with kd-tree보다



우수함을 알 수 있다. 그리고 표 8은 정합의 정확도 및 수행 시간을 나타낸 것이다.

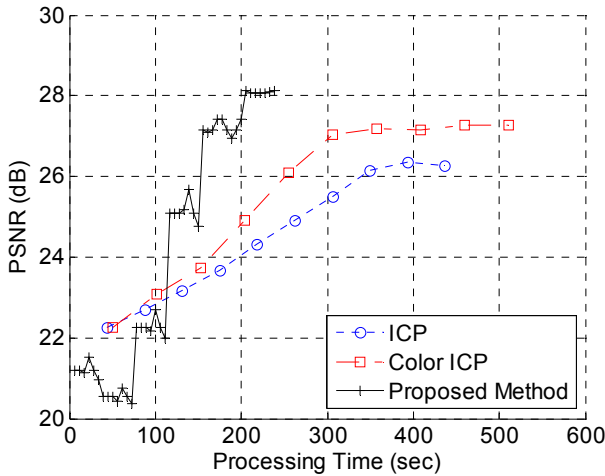


그림 10. 성능 비교

표 8. 정합의 정확도 및 수행 시간

Method	ICP with kd-tree	Color ICP with kd-tree	Proposed
Final PSNR (dB)	26.2556	27.2856	28.1307
Time (sec)	436.531 (10 Itr's)	511.516 (10 Itr's)	238.8590 (43 Itr's)
Time/Itr (sec)	43.6531	51.1516	5.55486

두 벽면에 대한 정합 및 모델링 결과를 그림 11에 보였다. 이를 얻기 위해 다시점 카메라를 두 벽면 주위로 움직이고, 획득된 데이터에 대해 제안된 방법을 적용하였다. 왼쪽 벽면에는 소파, 화분, 탁자, TV, 액자, 인형, 오른쪽 벽면에는 화분, 소파, 책장, 창문 등이 있다.



그림 11. 실내환경 복원 결과

## 6. 결론 및 향후 연구

본 논문에서는 다시점 카메라를 통해 획득된 부분적인 3D 점군을 이용하여 실내환경을 3D 복원하는 방법을 제안하였다. 제안된 방법은 깊이 정보가 정확하지 않거나 잡음에 의해 양안차 추정이 어려운 경우에도 효과적으로 사용될 수

있다. 또한 정합을 위해 요구되는 시간이 ICP (또는 Color ICP) with kd-tree보다 적게 소요된다. 그리고 한 쌍의 2D 영상을 3D 점군과 함께 사용함으로써 보다 개선된 복원이 가능함을 보였다. 추후에는 실내의 모든 면에 대한 복원을 위해 전역 최적화를 수행할 예정이다. 가상 물체를 복원된 실내환경에 자연스럽게 증강하기 위한 정확한 광원 추정 및 가상 환경의 조명 조건과 일치시키기 위한 분석이 요구된다. 또한, 세밀한 양안차 추정을 함으로써 보다 개선된 복원 결과를 얻을 있을 것이다.

## 참고 문헌

- [1] P. J. Besl and N. D. McKay, "A Method for Registration of 3-D Shapes," *IEEE Trans. on PAMI*, 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 Corp., Cambridge Research Lab, Oct., 1996.
- [3] S. Park and M. Subbarao, "A Range Image Refinement Technique for Multi-view 3D Model Reconstruction," *3-DIM*, pp. 147-154, 2003.
- [4] K. Pulli, *Surface Reconstruction and Display from Range and Color Data*, Ph.D. dissertation, UW, 1997.
- [5] K. Kutulakos and S. Seitz, "A Theory of shape by space carving," *IJCV*, vol. 38(3), pp. 197-216, 2000.
- [6] A. Broadhurst, T. W. Drummond, R. Cipolla, "A probabilistic framework for space carving," *ICCV*, 2001.
- [7] O. Faugeras and R. Keriven, "Complete dense stereovision using level set methods," *ECCV*, 1998.
- [8] C. Strecha, T. Tuytelaars and L. Van Gool, "Dense matching of multiple wide-baseline views," *ICCV*, pp. 1194-1201, 2003.
- [9] G. Vogiatzis, P. Torr and R. Cipolla, "Bayesian stochastic mesh optimization for 3d reconstruction," *BMVC*, 2003.
- [10] P. Gargallo, P. Sturm, "Bayesian 3D Modeling from Images Using Multiple Depth Maps," *CVPR'05*, vol. 2, pp. 885-891, 2005.
- [11] C. Strecha, R. Fransens and L. Van Gool, "Wide-baseline Stereo from Multiple Views: a Probabilistic Account," *CVPR'04*, vol. 2, pp. 552-559, 2004.
- [12] S. Kim and W. Woo, "Projection-based Registration Using a Multi-view Camera for Indoor Scene Reconstruction," *3-DIM*, pp. 484-491, June. 13-17, 2005.
- [13] S. Kim and W. Woo, "Registration of partial 3D point clouds acquired from a multi-view camera for indoor scene reconstruction," *IEICE*, 2006. (accepted)
- [14] 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.
- [15] KLT: Kanade-Lucas-Tomasi Feature Tracker, <http://www.ces.clemson.edu/~stb/klt/>, 2005.
- [16] Point Grey Research Inc., <http://www.ptgrey.com>, 2002.