

# 영상기반 3-D 모델링 및 합성 기법 구현

송문재\* \*\*, 권용무\*, 임성규\*\*

\*한국과학기술연구원 영상미디어연구센터, \*\*단국대학교 전자공학과

## Implementation of image-based 3-D modeling and synthesis technique

Munjae Song\*, Yong-Moo Kwon\*, Sungkyu Lim\*\*

\*IMRC, Korea Institute of Science and Technology, \*\*Dept of Electronic Engineering, Dankook Univ

### 요 약

실제감을 가지는 가상의 환경을 만들고, 사용자의 interaction에 맞추어 가상 환경과 오브젝트를 조정하는 본 이는 최근에 들어 많은 관심을 끌고 있다. 본 논문에서는 실제 영상을 이용하여 3차원 오브젝트 모델링 및 합성하는 방법에 대한 논의를 하였다. 실제 영상에서 깊이 정보를 얻어내는 방법으로는 다해상도 다기선 스테레오 정합 기법(multi-resolution multiple-baseline stereo matching technique MR-MBS)을 이용하였으며, 실제 영상에서 생성된 모델을 Open Inventor 환경에서 모델을 재구성하고 performer를 이용하여, 최종적인 합성 영상을 만들어 내었다. 합성된 임의시점에서의 영상에서 사용자 조작에 따라 각각의 오브젝트를 조정할 때 결과 영상을 실시간으로 얻을 수 있었고, 실제 영상의 텍스처를 그대로 사용함으로써 실제감을 높일 수 있었다.

## 1. 서 론

가상의 임의 시점영상을 만들어 내는 기술, 혹은 실제로 존재하지 않는 가상 환경을 만들어 내는 기술은 여러 가지 응용분야에서 가까운 미래에 매우 중요한 위치를 차지하게 될 것이다[1]. 응용 분야를 살펴보면, 3-D 원격 회의, interactive service, 원격 진료, 수술 등의 의료 기기 분야, 가상의 오케스트라 단원을 배경으로 하는 노래방 [2], 3-D 환경을 이용하는 방송 또는 영화 산업 등에 사용 될 수 있다. 예를 들면, 몇 대의 카메라를 이용하여 실제의 영상을 받아들이고 이를 이용하여 3차원 깊이 정보를 추출해 낸다. 추출된 깊이 정보를 이용하여 장면을 각각의 오브젝트로 나누어주어 이용할 수 있다. 예를 들어 배경과 주 관심 물체는 또 다른 배경, 혹은 물체 오브젝트와 합성 될 수 있고, 이런 방법으로 새로운 영상을 만들어 낼 수 있게 된다. 깊이 정보를 가지고 있는 상태이기 때문에 이를 이용하여 마치 사람의 두 눈으로 보는 것과 같은 두 장의 영상을 만들어 낼 수 있다. 또한, 두 영상을 사람의 두 눈에 비추어 줌으로써 입체감 있는 영상을 제공할 수 있다. 본 논문에서는 이러한 가상 환경을 만들어 내기 위하여 다음과 같은 방법을 사용하였다. 깊이 정보를 추출해 내기 위하여 다해상도 다기선 스테레오 정합 기법 ( MR-MBS

multi-resolution multiple-baseline stereo matching technique )을 사용하였고, 추출된 깊이 정보를 사용하여 다수의 오브젝트를 Open Inventor를 이용하여 모델링 하였고 이를 깊이정보를 이용하여 합성하고, performer를 이용하여 표시 해주었다.

## 2. 관련 연구

본 논문의 최종적인 목표는 실제 영상에서 깊이정보를 추출해 내고, 이를 이용하여 모델을 만들어 새로운 영상을 합성하고 사용자의 조작에 따라 자유롭게 조정하는 것이다. 이러한 가상의 환경을 만들어 내고 합성하는 방법에는 여러 가지 접근 방법이 오랫동안 연구되어 있었으나, 다음과 같이 크게 두 가지로 나눌 수 있다[3].

### 1) 영상 기반 방법

영상 기반 방법 중 가장 널리 알려진 방법은 Apple사의 QuickTime VR 일 것이다. 이 방법은 고정된 위치에서 360° 돌아가면서 영상을 입력받고(cylindrical panorama images), 이를 사용자의 조작에 맞추어 적당한 영상을 표시해 주는 방법이다. 다른 방법의

접근은 warping 과 morphing 을 이용하는 것이다[4]

2) 모델 기반 방법

상대적으로 정보량이 적고, 새로운 영상을 합성하기 편리한 모델을 이용하여 가상 환경을 만들어 내는 과정은 다음과 같이 세 단계로 나누어 질 수 있다 첫 번째는 다수의 카메라로부터 영상을 입력받는다 Saed Moezzi 등의 방법[3]에서는 17대의 카메라를 사용하여 역동적으로 움직이는 사람의 3차원 모델을 만들어 내었고, Takeo Kanade 등의 방법[4]에서는 51대의 카메라를 가진 dome 안의 사람을 모델링 하였다 본 논문에서는 5대의 카메라를 사용하여 입력받는다. 다음 단계는 입력된 영상에서 3-D모델을 만들어 내는 과정이고, 마지막 단계는 깊이 정보를 만든어진 모델의 정보를 이용하여 렌더링 하는 과정이다 이 과정에서 실제 영상의 텍스처를 이용함으로써 실제감을 높인다

3. 3차원 오브젝트 모델링 및 합성

3차원 모델을 이끌어 내기 위해서는 조밀하고(dense), 정확한 깊이 정보를 얻는 것이 중요하며 텍스처를 올바르게 모델에 적용하는 문제도 고려해야 한다. 그림 1에서는 이러한 요소들을 고려한 3차원 오브젝트 모델링 시스템 구성을 나타내었다.

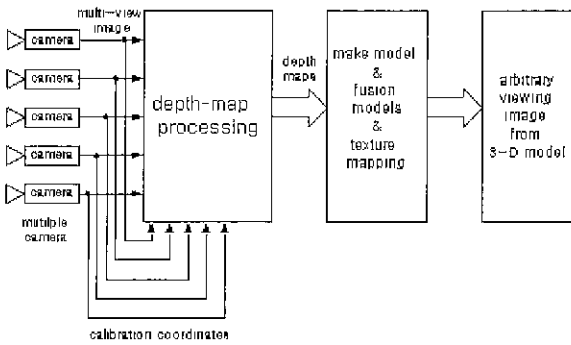


그림 1 오브젝트 모델링 및 합성 시스템

실험은 KIST에서 제작된 영상과 일본 쓰쿠바 대학에서 만들어진 이미지 데이터 베이스로 행하여 졌다 본 논문에서 구현한 실험을 진행하는 과정은 다음과 같다.

1) 카메라 캘리브레이션 다수의 카메라를 사용하는 실험의 특성상 각각의 카메라 특성을 알아내는 작업이 반드시 수반되어야 한다 대표적인 방법으로는 Tsai의 방법[6]이 가장 널리 알려져 있으며 본 실험에서도 역시 Tsai의 방법을 사용하였다

2) 깊이 정보 추출 입력되는 영상에서 깊이 정보를 추출하기 위한 방법으로 MR-MBS방식을 사용하였다 이 방식은 단순한 stereo-matching 방식에 비하여 폐색영역(occlusion-region)에 대해 효과적이며, 경계선 인장 문제를 해결하였고, 개선된 MBS방식에 비하여 처리 시간이 단축되는 장점을 가지고 있다 본 실험에서는 이

와 같은 MR-MBS방식으로 깊이정보를 추출하는 데에 SUN Ultra SPARC 1 기종을 사용하였다 KIST의 LION 영상에 대하여 MR-MBS방식으로 만들어진 depth-map을 그림 2에 나타내었다

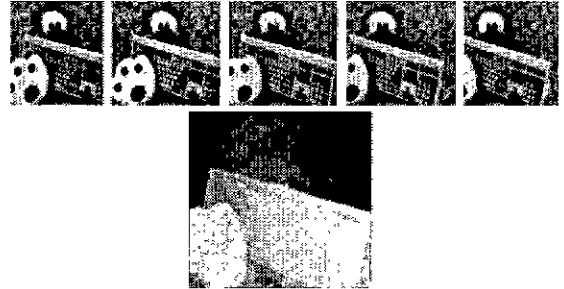


그림 2. 위 - 입력된 5개의 영상(치민 왼쪽으로 직각 표시함)  
아래 - 5개의 영상으로부터 구하여진 depth-map

3) 모델링과 텍스처 매핑 실험 영상은 배경과 오브젝트가 익힌의 거리를 두고 떨어져 있다는 조건이 필요하다 이러한 조건에서 구하여진 depth-map을 히스토그램 방식을 이용하여 배경과 오브젝트를 분리 해 낸다 그림 3은 LION영상의 디스패리티 히스토그램을 나타내고 있다.

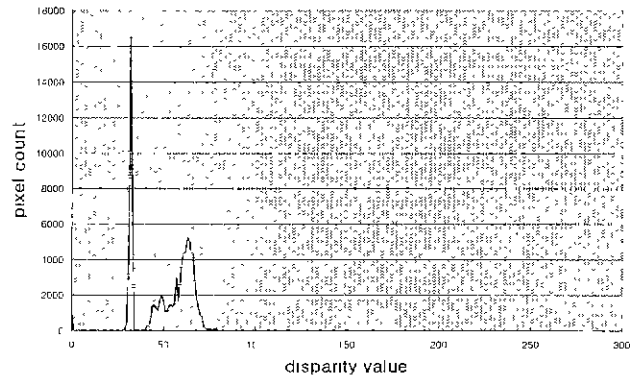


그림 3 LION 영상의 디스패리티 히스토그램

히스토그램에서 두 개의 영역을 볼 수 있는데 낮은 디스패리티와 차지하는 영역이 좁은 부분은 배경으로 볼 수 있고, 높은 디스패리티 값을 가지고 넓은 영역을 가지는 부분을 오브젝트로 볼 수 있다. 오브젝트 영역을 구하게 되면 깊이 정보를 이용하여 일정한 간격으로 vertex를 취하고, 각 vertex를 연결한 삼각형으로 만들어 오브젝트의 표면을 표현하게 된다. 최종적으로 5개의 입력 영상 중 깊이 정보와 동일한 위치인 가운데 영상을 사용하여 텍스처 매핑을 한다 이러한 일련의 작업과정을 그림 4에 나타내었고, 실험 대상을 모델링 하기 위해 필요한 vertex수와 삼각형 수를 표 1에 나타내었다.

4) 새로운 영상 합성과 각각의 모델 객체의 조정 · 오브젝트 모델을 이용하여 만들어진 결과는 사용자 조작에 따라 실시간으로 오브젝

트의 XYZ축 이동, XYZ축을 중심으로 한 회전 그리고 오브젝트의 크기를 자유롭게 변화시킬 수 있었다. 최종적인 결과는 performer를 이용하여 렌더링 하였으며 오브젝트 모델링부터 렌더링까지의 작업은 SGI OCTANE 기종에서 실험하였다. 두 개의 오브젝트를 합성한 결과를 그림 5에 나타내었다.

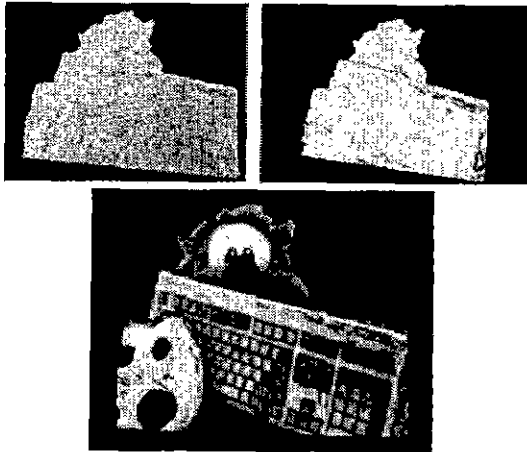


그림 4. 모델링된 오브젝트 - wire-frame, field, texture mapping(왼쪽 위부터 시계방향으로)

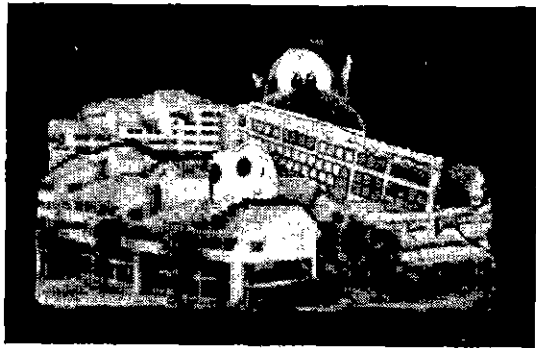


그림 5. LION 오브젝트와 CITY오브젝트를 합성한 영상

표 1. 오브젝트 모델링 결과

	영상 크기(pixel)	vertex 수(개)	triangle 수(개)
LION	320 x 307	64,872	21,624
CITY	320 x 240	23,523	7,841

#### 4. 결 론

실제영상을 이용하여 3차원 모델을 만들어 내는 작업은 그 응용분야가 매우 넓은 뿐 아니라 모델이 가지는 장점인 정보량이 상대적으로 적고 기하학적인 간단한 인산변으로 새로운 영상을 구현 할 수 있으며, 합성이 쉽다는 장점 때문에 지속적인 연구가 뒤따라야 할 것으로 생각된다. 본 논문에서는 실제 영상에 기반한 모델의 생성 및 생성된 모델들의 용합을 통한 새로운 영상의 합성을 보였다 또한 합성된 영상에서 각 모델 객체에 대한 조각이 가능함을 보였다. 본 연구의 추후과제는 on-line 상으로 오브젝트를 모델링 할 수 있는 시스템을 구현하는 것이다 이러한 시스템 구성도를 그림 6에 나타내었다. DSP 보드를 사용하여 실시간에 가깝게 depth-map을 구하고 이를 이용하여 오브젝트 모델링을하고 텍스처 매핑 및 렌더링 작업을 거쳐 렌더링 시스템 또는 LCD 안경을 이용하는 방법으로 입체영상을 표시하게 되며, 각 부분들은 네트워크로 묶여 전체 시스템을 구성하게 된다.

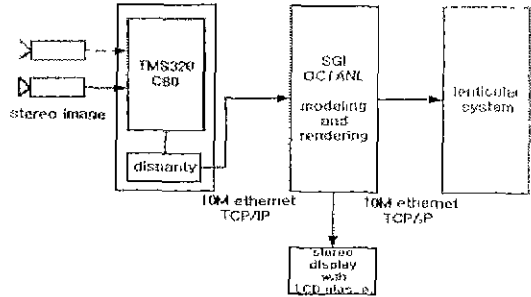


그림 6 on-line 오브젝트 모델링 시스템 구성도

#### 참 고 문 헌

[1] Ebroul Izquierdo M and Silko Kruse, "Image Analysis for 3D Modeling, Rendering, and Virtual View Generation," COMPUTER VISION AND IMAGE UNDERSTANDING, Vol. 71, No. 2, pp 231-253, Aug. 1998

[2] ChangWhan Sul, KeeChang Lee, and Kwangyun Wahn. "Virtual Stage A location-Based Karaoke System," IEEE MultiMedia 5, pp 42-52, 1998

[3] Saied Moezzi, Li-Cheng Tai, and Philippe Gerard, "Virtual View Generation for 3D Digital Video," IEEE MultiMedia 4, pp 18-26, 1997

[4] M Hirose, "Image-based virtual world generation," IEEE MultiMedia 1, pp 27-33, 1997

[5] Takeo Kanade, and Peter Tander, "Virtualized Reality Constructing Virtual Worlds from Real Scenes," IEEE MultiMedia 4, pp 34-37, 1997.

[6] Roger Y Tsai, "A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Camera and Lenses." IEEE J Robotics And Automation, Vol 3, No 4, pp 323-344, Aug. 1987.