

논문 2007-44SP-1-1

다중 스테레오 카메라를 이용한 3차원 모델링 시스템

(A 3D Modeling System Using Multiple Stereo Cameras)

김 한 성*, 손 광 훈**

(Hansung Kim and Kwanghoon Sohn)

요 약

본 논문에서는 임의 시점에서의 장면을 생성하기 위한 3차원 모델링 및 렌더링 시스템을 제안한다. 제안되는 시스템은 공간 상에 설치된 복수의 스테레오 카메라와 PC들로 구성되며 UDP를 이용해 연결되어 각 카메라에서 획득되고 분석된 영상 데이터 들을 모델링 PC로 전송해 실시간으로 3차원 모델을 생성하고, 이로부터 사용자가 원하는 위치에서의 장면을 생성해 디스플레이 하게 된다. 제안된 알고리즘은 성능 평가 결과 기존의 알고리즘보다 좋은 성능을 보였으며, 구현된 시스템은 실시간으로 사용자에게 원하는 시점에서의 영상을 자연스럽게 제공함을 실험을 통해 검증하였다.

Abstract

In this paper, we propose a new 3D modeling and rendering system using multiple stereo cameras. When target objects are captured by cameras, each capturing PC segments the objects and estimates disparity fields, then they transmit the segmented masks, disparity fields, and color textures of objects to a 3D modeling server. The modeling server generates 3D models of the objects from the gathered masks and disparity fields. Finally, the server generates a video at the designated point of view with the 3D model and texture information from cameras.

Keywords : 3D Modeling, Real-time system, Segmentation, Disparity estimation, Free-view generation

I. 서 론

본 논문에서 우리는 다중 스테레오 카메라 환경에서 원하는 시점에서의 영상을 자유롭게 재현해주는 3차원 영상 시스템을 제안한다. 영상으로부터 3차원 모델을 복원하는 기법은 매우 오래된 연구 주제였다. 대표적인 기술로는 빛과 그림자의 관계로부터 3차원 정보를 복원하는 Shape from shading 기법^[1], 물체 표면 패턴으로부터 복원하는 Shape from texture 기법^{[2][3]}, 윤곽선 정보로부터 복원하는 Shape from silhouette (SFS) 기법

^{[4][5]} 등이 있다. 특히 SFS 기법은 간단하면서도 효율적으로 3차원 모델을 복원할 수 있어 최근 많은 연구가 이루어졌다. 하지만 SFS는 깊이 정보 없이 윤곽선 정보만으로 3차원 공간을 조각하듯 잘라 나가는 방법이므로, 오목한 영역에 대한 복원이 불가능하고 카메라의 수가 적은 경우 조악한 모델이 생성되는 단점이 있다^[6].

이에 본 논문에서는 그림 1과 같이 스테레오 카메라를 사용하여 실시간으로 정밀하게 3차원 모델을 복원하고 렌더링하는 시스템을 제안한다.

전체 시스템은 크게 영상을 획득해서 객체를 분할하고 변이를 추정하는 획득단과 추출된 정보들로부터 모델을 복원해 임의의 시점을 렌더링하는 모델링단으로 구성된다. 설치된 각 스테레오 카메라는 카메라 보정(Camera calibration) 기법에 의해 사전에 내부와 외부 파라미터가 추출되며, 각 PC에서는 실시간으로 배경으로부터 객체를 분할하고 스테레오 기법을 이용해 획득된 영상의 변이를 추정한다. 분할된 객체 정보와 추출

* 정회원, ATR, Japan
(ATR Knowledge Science Lab)

** 정회원, 연세대학교
(Yonsei University)

※ 이 논문은 2005년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임.
(KRF-2005-214-D00356)

접수일자: 2006년7월12일, 수정완료일: 2006년12월26일

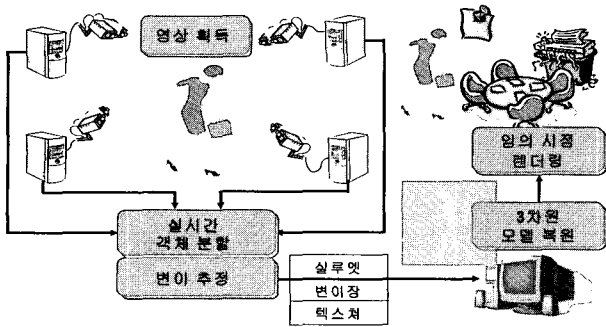


그림 1. 제안 시스템 구성도

Fig. 1. Overall configuration of the proposed system.

된 변이 정보는 네트워크를 통해 텍스처 정보와 함께 모델링 PC로 전송되어 사전에 추출된 카메라 파라미터들을 이용해 3차원 공간상의 완전한 모델로 복원된다.

II. 카메라 보정

카메라 보정은 카메라로부터 획득된 영상으로부터 카메라의 내부와 외부 특성 파라미터를 추출하는 기법으로 일반적으로 그림 2와 같은 특수한 기지 패턴을 이용해 측정된다. 카메라 보정 기법은 패턴으로부터 직접 각 파라미터를 추출하는 방법과 투영 행렬을 먼저 구하고 이로부터 다시 각 파라미터를 구하는 방법이 있다. 제안 시스템에서는 모델 복원에 투영 행렬을 사용하므로 두번째 방법을 사용하여 다음 파라미터들을 추출하였다^[7].

투영 행렬

$$P = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix}$$

외부 파라미터

$$\text{회전 행렬 } R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}$$

$$\text{이동 벡터 } t = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

내부 파라미터

유효 화소 사이즈 $[f_x, f_y]$ 영상 중심 좌표 $[c_x, c_y]$

3차원 공간상의 좌표 (X_i, Y_i, Z_i) 와 이 점이 투영된

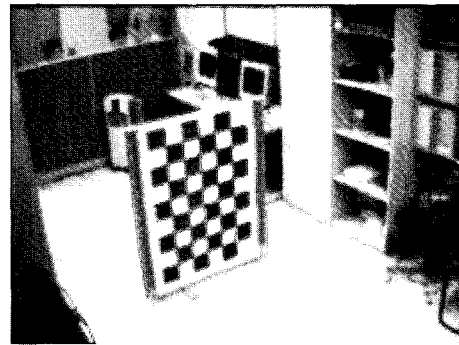


그림 2. 기지패턴에 의한 카메라 보정

Fig. 2. Calibration using a pattern board.

2차원 영상 내에서의 좌표 (x_i, y_i) 사이의 관계는 식 (1)과 식 (2)와 같은 투영행렬로 나타내어진다.

$$\begin{bmatrix} u_i \\ v_i \\ s_i \end{bmatrix} = P \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} \tag{1}$$

$$x = \frac{u_i}{s_i} = \frac{p_{11}X_i + p_{12}Y_i + p_{13}Z_i + p_{14}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}} \tag{2}$$

$$y = \frac{v_i}{s_i} = \frac{p_{21}X_i + p_{22}Y_i + p_{23}Z_i + p_{24}}{p_{31}X_i + p_{32}Y_i + p_{33}Z_i + p_{34}}$$

투영행렬 P 는 랭크 11을 가지므로 6쌍의 3차원 좌표와 2차원 좌표간의 정합점을 알면 각 원소를 계산할 수 있다. 우리는 오차를 줄이기 위해 그림 2와 같은 기지 패턴을 이용해 40개의 좌표 정합을 구하고 최소 자승 기법을 이용해 오차를 최소화하는 투영행렬 P 를 계산하였다. 투영행렬 P 를 구해내면 식 (3)의 관계로부터 내부 파라미터와 외부 파라미터들을 구해낼 수 있다.

$$P = \begin{bmatrix} -f_x r_{11} + c_x r_{31} & -f_x r_{12} + c_x r_{32} & -f_x r_{13} + c_x r_{33} & -f_x t_x + c_x t_z \\ -f_y r_{21} + c_y r_{31} & -f_y r_{22} + c_y r_{32} & -f_y r_{23} + c_y r_{33} & -f_y t_y + c_y t_z \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{3}$$

III. 실시간 객체 분할

실시간 객체 분할 기술은 최종 모델의 정확도와 효율성을 결정하는 매우 중요한 기술이다. 이에 우리는 배경 차분(background difference)과 프레임차분(Frame difference)을 동시에 이용한 전경 분할 기법을 제안한다. 그림 3은 제안된 객체 분할 알고리즘의 블록도이다.

먼저 초기화 단계에서 배경 영상의 각 화소 최소값인 $I_{min}(x,y)$ 와 최대값인 $I_{max}(x,y)$ 를 산출하여 저장한다.

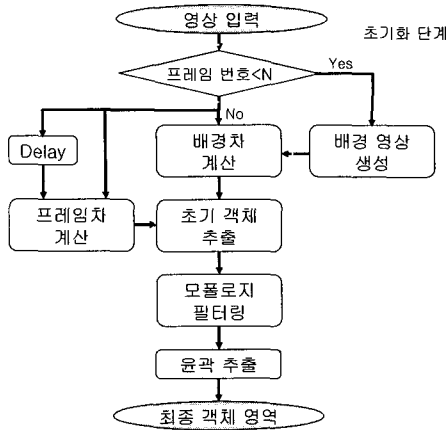


그림 3. 객체 영역 분할 블록도
Fig. 3. Block diagram of object segmentation.

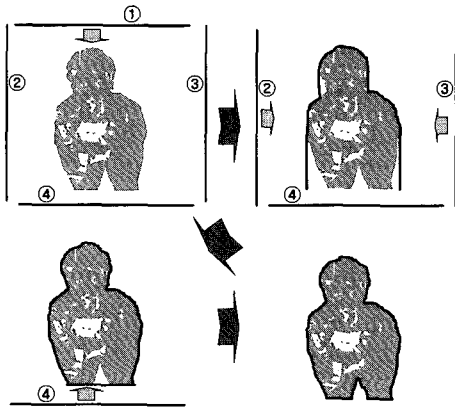


그림 4. 최종 윤곽 추출
Fig. 4. Profile extraction.

그리고 실제 연산중에는 현재 영상인 I_{cur} 의 값이 배경 영상의 최소값 범위의 일정 영역 Th_{tol} 밖에 존재하거나, 두 연속되는 프레임간의 차영상인 $I_{fd}(x,y)$ 를 계산하여 그 값이 특정 한계치 Th_{fd} 이상이면 초기 객체 영역으로 할당한다. 즉, 현재 프레임의 값이 다음 식 (4) 와 같은 배경차와 프레임차 조건 중 하나를 만족하면 초기 객체 영역으로 분할된다.

$$\begin{aligned} I_{cur}(x,y) &< I_{min}(x,y) - Th_{tol} \\ I_{cur}(x,y) &> I_{max}(x,y) + Th_{tol} \\ I_{fd}(x,y) &> Th_{fd} \end{aligned} \quad (4)$$

그러나 이러한 조건으로 추출된 객체 영역은 많은 잡음과 에러를 포함하고 있다. 이에 우리는 closing 기법과 region growing을 적용한 작은 영역 제거 기법을 통해 잡음과 잘못 추출된 영역을 제거한다^[8]. 마지막으로 전경의 윤곽을 부드럽게 만들면서 영역 내부의 빈 영역들을 제거하기 위해 Kumar가 제안했던 윤곽 추출 기법^[9]을 응용하여 최종 전경 영역을 추출한다. 윤곽 추출은

일정 크기의 탄성을 가진 라인을 영상의 한쪽 끝에서 반대쪽으로 이동시키며 영역을 둘러싸게 되며, 라인의 탄성으로 인해 폭이 일정 크기 이하의 영역에는 침투하지 못하게 된다. 이러한 과정을 4방향에서 모두 수행하여 4개의 결과에 모두 포함된 영역을 최종 객체 영역으로 추출하게 된다. 그림 4는 이러한 최종 객체 영역 추출 과정을 보여준다.

IV. 스테레오 영상의 깊이 복원

깊이 정보 복원에 있어서 가장 중요한 문제는 동일한 장면을 조금 다른 위치에서 찍은 두 장의 영상 내에서 정합점을 찾는 변이추정(disparity estimation)이다. 평행식 스테레오 카메라를 가정할 때 추정된 변이는 스테레오 영상 모델에 의해 다음과 같이 깊이 정보로 변환될 수 있다. 여기서 f 는 카메라의 초점 거리, d 는 추출된 변이, B 는 카메라간의 거리를 의미한다.

$$z = \frac{fB}{d} \quad (5)$$

본 장에서는 배경 정보 등록과 앞 장에서 제안한 전경 분할을 이용한 고속 변이 추정 알고리즘을 제안한다. 제안하는 시스템에서 우리는 영상을 받아들이는 스테레오 카메라는 고정되어 있고, 배경 정보를 추출하기 위한 초기 몇 초간의 초기화 시간 중에는 움직이는 물체가 존재하지 않는다고 가정한다. 이러한 가정 하에서 정밀한 배경 영역의 변이 정보가 사전에 추출되고, 실제 시스템이 동작할 때는 전경 영역만을 추출하여 고속으로 변이를 추정하고, 배경 정보와의 융합을 통해 완

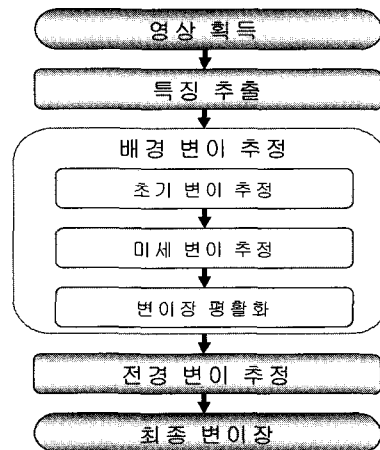


그림 5. 변이 추정 알고리즘 블록도
Fig. 5. Block diagram of disparity estimation.

전한 변이맵을 도출하게 된다. 그림 5는 실시간 변이 추정 알고리즘의 블록도를 보여준다.

먼저 획득된 스테레오 동영상은 전처리 과정으로써 영상내의 잡음을 줄이기 위한 저역 통과 필터와 에피폴라 선을 수평으로 맞추기 위한 정렬 과정을 거친다.

배경 영역의 변이는 초기화 단계에서 수행되기 때문에 실시간 추출이 아니어도 되고, 또한 한 번 추출된 변이가 전체 프로세스에 사용되므로 최대한 정밀한 변이를 추출하여야 한다. 이에 우리는 기존에 제안했던 영역 분할과 변이 평활화 기법을 이용한 계층적 변이 추정을 사용한다^[10].

계층적 변이 추정에서는 먼저 입력된 스테레오 영상을 1/2로 다운 샘플링되어 $B \times B$ 크기의 블록 단위로 변이가 추정되며, 두 번째 단계에서는 이러한 블록 단위의 초기 변이를 기반으로 원영상에서 다시 영역 분할 기법을 사용하여 화소 단위의 미세 변이를 추정한다.

그러나 이러한 계층적 추정 방식은 과정에서 변이장내의 공간적 상관도가 고려되지 않았기 때문에 일부 잘못된 변이가 할당될 수 있다. 또한 변이의 값이 정수값으로 제한되었기 때문에 그 정확도가 떨어진다. 이에 우리는 변이 평활화 기법을 적용하여 더욱 정밀한 변이장을 추출한다^[10]. 변이 평활화 과정에서는 다음과 같이 변이의 정확도와 변이의 평활화도를 함께 고려한 에너지 함수를 사용하여 이를 최소화함으로써 최종 변이를 결정하게 된다.

$$E(d(x, y)) = \int_{\Omega} (I_l(x, y) - I_r(x + d(x, y), y))^2 dx dy + \lambda \int_{\Omega} \Psi(\nabla d(x, y), \nabla I_l(x, y)) dx dy \quad (6)$$

위 식에서 Ω 는 이미지 영역을 의미하고, λ 는 평활화부의 가중치를 나타내며, $\Psi(\nabla d, \nabla I_l)$ 는 그 기울기 (gradient)가 다음 식 (7)로 나타나는 변이장과 영상의 기울기에 관한 함수로서 객체의 경계 영역에서는 평활화가 일어나지 않고 경계가 보존된다.

$$\nabla(\Psi(\nabla d, \nabla I_l)) = \frac{1}{(1 + |\nabla I_l|^2)^2} \nabla d \quad (7)$$

식 (7)의 에너지 함수를 최소화하는 변이장 $d(x, y)$ 를 찾는 문제는 대응되는 Euler-Lagrange 방정식의 해를 찾는 문제로 변형되며, 이는 다음과 같은 포물선형 (parabolic) 편미분 방정식에서 t 가 무한대(∞)로 접근할 때의 수렴해와 동일하다.

$$\frac{\partial d}{\partial t} = \lambda \operatorname{div} \left(\frac{1}{(1 + |\nabla I_l|^2)^2} \nabla d(x, y) \right) + (I_l(x, y) - I_r(x + d, y)) \frac{\partial I_r(x + d, y)}{\partial x} \quad (8)$$

위 편미분 방정식은 보조항이 있는 비선형 확산 방정식과 동일한 꼴이며^[11] 우리는 유한 차분을 이용해 이 연속적인 시스템을 이산화하였다. 모든 공간적인 미분은 전방 차분(forward difference)을 이용해 근사화되었으며, 비선형적인 시스템 내의 함수는 Taylor 전개를 한 후, 1차항까지만 사용함으로써 선형 시스템으로 근사화하였다. 최종적인 변이장은 선형 시스템의 반복적 갱신법에 의해 구해진다.

반면, 전경 영역의 변이는 각 프레임마다 매번 추출되어야 하므로 전경 영역 변이 추정에서 가장 중요한 요소는 속도라 할 수 있다. 이에 배경 영역의 변이 추정에서 사용되었던 계층적 변이 추정을 동일하게 전경 영역에도 적용하지만, 시간 소요가 많은 변이 평활화 과정은 거치지 않는다. 또한 전경 영역은 언제나 배경보다 거리상 앞에 존재하기 때문에 변이 추정에서 검색 범위는 인접한 배경 영역의 변이에 의해서도 제한된다. 식 (9)는 배경 변이에 따른 전경 영역 최대 검색 범위 SR_{Max} 와 최소 검색 범위 SR_{Min} 를 보여준다. 식에서 d_{ln} 과 d_{rn} 는 동일한 행 내에서 추정 위치에서 각각 오른쪽과 왼쪽 방향으로 인접해 있는 배경 영역의 변이를 의미한다.

좌→우 변이

$$SR_{Max} = \operatorname{Min}(d_{ln}, d_{rn}) \quad (9)$$

우→좌 변이

$$SR_{Min} = \operatorname{Max}(d_{ln}, d_{rn})$$

그러나 정합 과정에서 기존의 정사각형 정합창은 주변부의 큰 특징점에 강하게 영향을 받으므로 객체의 경계 부분 영역에서는 오류가 발생하여 정확한 경계가 검출되지 않는 단점이 있다. 배경 영역의 변이 추정에서는 이러한 오류가 평활화 과정에서 보정되지만, 전경 영역에서는 평활화를 수행하지 않으므로 이러한 오류가 계속 남아있게 된다. 이에 우리는 전경 영역과 배경 영역의 경계를 포함하고 있는 블록의 화소들에 대해서는 다시한번 변이의 신뢰도를 비교하여 전경 변이와 배경 변이 중 높은 신뢰도의 변이를 할당하는 과정을 포함한

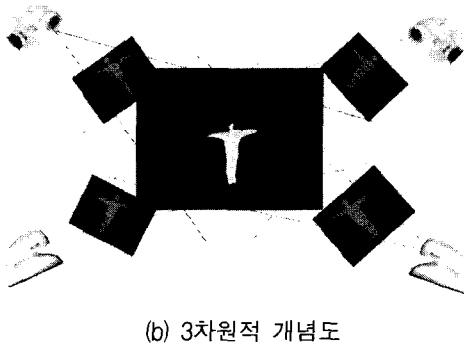
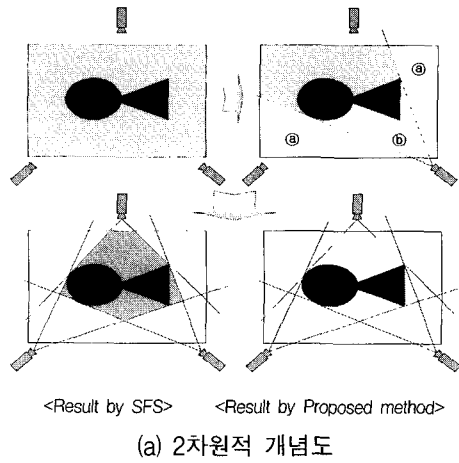


그림 6. 윤곽과 변이 정보에 의한 모델링
Fig. 6. Modeling from silhouette and disparity.

다. 이러한 경계 영역에서의 최종 변이는 다음과 같은 조건에 의하여 결정된다. 식 (10)에서 $d_{back}(x,y)$ 는 (x,y) 위치의 배경 변이장의 변이를 의미하며, $d_{fore}(x,y)$ 는 동일한 위치에서의 전경 영역의 변이장을 의미한다.

$$\begin{aligned} & \text{if } (|I_r(x,y) - I_r(x+d_{fore},y)| < |I_r(x,y) - I_r(x+d_{back},y)|) \\ & \quad d_{final}(x,y) = d_{fore}(x,y) \\ & \text{else} \\ & \quad d_{final}(x,y) = d_{back}(x,y) \end{aligned} \quad (10)$$

V. 3차원 모델링 및 렌더링

1. 3차원 모델링

3장에서 추출된 객체 정보와 IV장에서 추정된 변이 정보를 이용해 3차원 공간상에 객체의 모델을 생성한다. 그림 6은 제안하는 윤곽 정보와 변이 정보를 이용한 3차원 복원 알고리즘의 개념을 보여준다. 그림 6(a)에서 보듯 기존의 SFS 방식은 윤곽선 정보만을 이용하므로 객체 정보를 투사할 때 ㉠ 영역만이 제거되고 ㉡ 영역은 제거되지 않는다. 결과적으로 SFS 방식은 카메

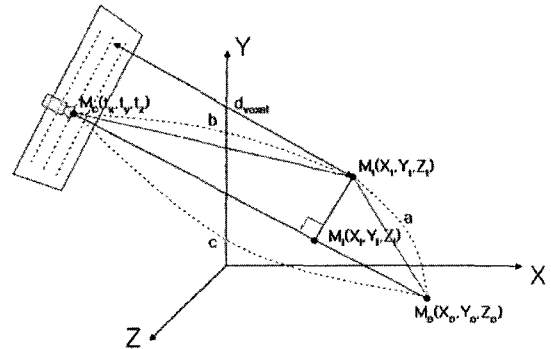


그림 7. 카메라로부터 3차원 공간까지의 거리
Fig. 7. Distance from a camera to a 3D point.

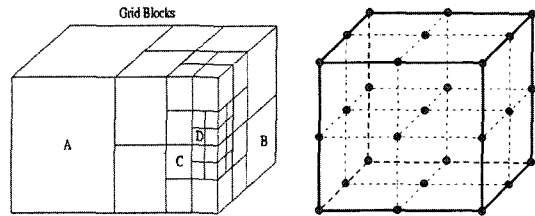


그림 8. 8진 트리 구조
Fig. 8. Octree structure.

라의 수가 적은 경우 실제 모델보다 큰 조각한 모델이 생성되며, 카메라 수에 관계없이 오목한 영역에 대한 복원이 불가능하다. 그러나 제안하는 방식은 각 카메라에서 추출된 객체의 윤곽과 변이 정보가 각 카메라의 파라미터에 의해 3차원 공간상으로 투사되며 투사된 정보들은 객체의 윤곽과 깊이 이외의 영역을 제거하여 최종적으로 객체 영역만 남기게 된다.

실제적인 모델 형성은 3차원 공간을 구성하는 요소인 복셀(voxel)을 각 카메라에 투영하여 이루어진다. 3차원 공간상의 점 $M(X, Y, X)$ 은 각 카메라의 영상 I_n 에 카메라의 투영행렬에 의해 식 (11)과 같이 투영된다.

$$[u, v, s]^T = P_n [X, Y, Z, 1]^T \quad (11)$$

만약 M 이 객체의 윤곽 내부에 투영된다면 다음으로 M 이 투영점의 변이 정보로부터 계산된 깊이보다 더 뒤쪽에 존재하는지 검사한다. 카메라로부터 M 까지의 거리 d_{voxel} 은 그림 7에 기반하여 식 (12)로 계산되며, 변이로부터 계산되는 거리 d_{surf} 는 식 (5)로 계산된다.

$$\begin{aligned} d_{voxel} &= \frac{-a^2 + b^2 + c^2}{2c} \\ &= \frac{\left\{ \begin{aligned} & -(X_t - X_o)^2 - (Y_t - Y_o)^2 - (Z_t - Z_o)^2 \\ & + (X_t - t_x)^2 + (Y_t - t_y)^2 + (Z_t - t_z)^2 \\ & + (X_o - t_x)^2 + (Y_o - t_y)^2 + (Z_o - t_z)^2 \end{aligned} \right\}}{2\sqrt{(X_o - t_x)^2 + (Y_o - t_y)^2 + (Z_o - t_z)^2}} \end{aligned} \quad (12)$$

만약 점 M 이 모든 카메라에 대해 윤곽 내부에 존재하며 동시에 표면까지의 거리 뒤쪽에 존재하면 점 M 은 객체 내부에 속해 있는 복셀로 판단된다. 그러나 모델링할 공간상의 모든 복셀을 각 카메라에 투영해 보는 것은 매우 소모적인 일이다. 따라서 우리는 복셀의 크기를 계층적으로 줄여가면서 모델링하는 그림 8의 8진 트리구조 (Octree structure)를 사용한다. 8진 트리 구조에서 각 복셀의 꼭짓점과 모서리의 중심점, 면의 중심점, 복셀의 중심점을 포함하는 총 27개의 점이 체크된다. 만약 27개의 점이 모두 객체 내부에 포함되거나 모두 객체 밖에 있다고 판단되면 그 복셀은 각각 객체 내부 혹은 외부 복셀로 지정된다. 그렇지 않을 경우에는 그 복셀은 하위단계의 8개의 복셀로 분할되어 동일한 과정이 반복된다. 이러한 8진 트리 구조를 사용하면 계산 시간뿐 아니라 데이터의 양도 크게 줄어들게 된다.

2. 임의 시점 렌더링

3차원 모델이 완성되면 서버는 모델과 전송 받은 텍스처 정보를 기반으로 사용자가 요구하는 시점에서의 장면을 렌더링하여 화면에 디스플레이 해준다. 배경 공간은 시스템이 작동되기 전에 사전에 모델링되며, 실시간 처리에서는 객체 영역만 업데이트되어 렌더링된다. 우리는 시스템을 단순화 하고 렌더링 속도를 높이기 위해 OpenGL에서 제공하는 투영 텍스처 매핑 (Projective texture mapping) 함수를 사용하였다^[12]. 투영 텍스처 매핑은 그림 9와 같이 마치 슬라이드 프로젝터가 영상을 공간상에 투사하듯 카메라 위치로부터 텍스처를 모델 위에 투사한다.

VI. 실험 결과

우리는 그림 10과 같은 실험 공간을 구축하였다. 스테레오 카메라는 각각 2대의 Digiclops와 Bumblebee를 사용하였으며^[13], 각 카메라에는 Pentium IV 3.0GHz PC가 한 대씩 연결되어 영상을 획득과 동시에 객체를 분할하고 변이를 추정하여 텍스처 정보와 함께 UDP를 통해 모델링 PC로 전송한다. 모델링에는 Quadro FX1300 그래픽 카드가 장착된 Pentium IV PC를 사용하였다. 각 카메라의 해상도는 320×240이며, 모델링되는 공간은 최소 1cm×1cm×1cm 크기의 복셀로 256×256×256 해상도로 구성된다.

그림 11은 4대의 카메라로 동시에 획득한 영상과 이에 대한 객체 분할 정보 및 변이 추정 결과를 보여준다.

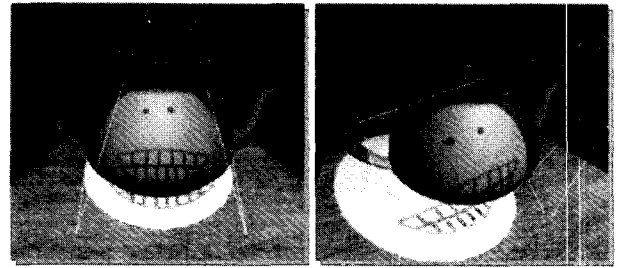


그림 9. 투영 텍스처 매핑^[12]

Fig. 9. Projective texture mapping.

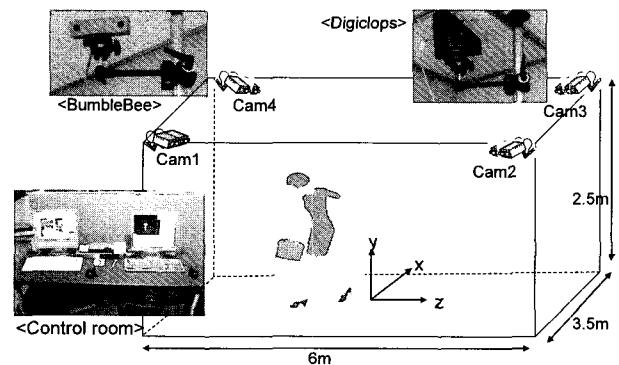


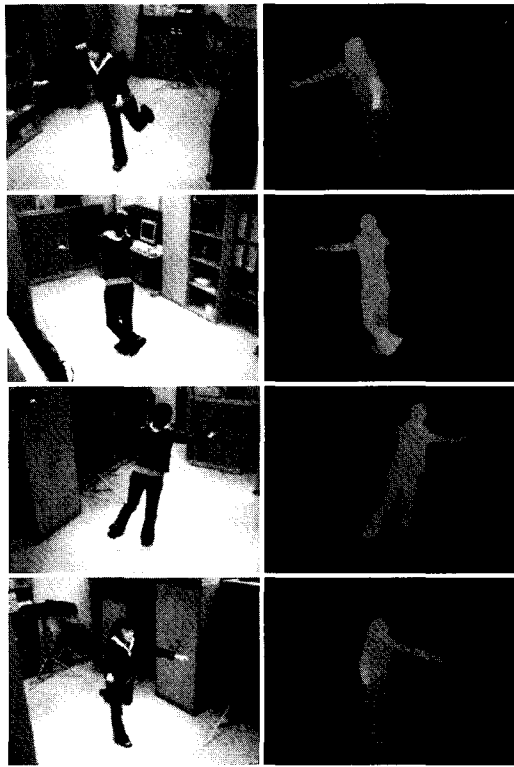
그림 10. 실험 공간

Fig. 10. Experimental room.

단, 이 실험에서는 객체 분할 정보만을 사용하는 SFS 기법과의 성능 비교를 위해 객체는 반수동으로 분할하였고 변이는 평활화 과정까지 수행하였다. 그림 12는 그림 11을 기반으로 생성된 3차원 모델을 보여주며 이를 기존 방식인 SFS의 결과와 비교하였다. 그림에서 보듯 기존의 SFS 방식은 윤곽선 정보만을 이용하므로 실제 모델에 비해 더 크고 조악한 모델이 생성되었고, 이로 인해 모델에 텍스처를 매핑할 경우 배경 영역이 모델 내부에 붙는 등 매우 어색한 결과를 초래하게 된다. 이에 비해 제안된 윤곽과 변이 정보를 함께 고려한 모델링 기법은 더 실제 모델에 가까운 결과를 보여준다.

그림 13은 실시간으로 전체 시스템이 동작될 때의 최종 렌더링 결과를 보여준다. 획득 영상의 해상도가 낮고 모델을 복셀 기반의 구조를 사용하였기에 줌 인을 많이 하면 해상도가 떨어지는 문제가 있으나 전반적으로 지정된 시점에서 자연스러운 영상을 제공한다. 또한 3차원 모델로부터 직접 렌더링을 하므로 그림 13 (c) 와 같이 두 번의 렌더링을 통해 스테레오 영상을 생성하여 입체 영상을 만들어 낼 수도 있다.

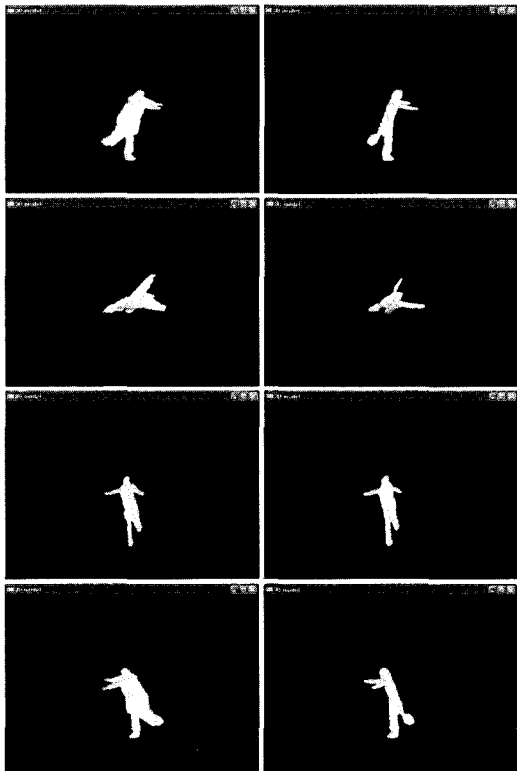
표 1은 제작된 시스템의 속도 분석 결과를 보여준다. 측정된 시간은 모델링 공간 내에 한 명의 모델이 움직이고 있을 때 평균 처리 시간이다. 카메라 측에서의 영상 분석은 평균 약 15 frames/sec의 속도를 보였으며



(a) 획득 영상 (b) 윤곽 및 변이 정보

그림 11. 영상 획득 및 분석 결과

Fig. 11. Captured and analyzed images.



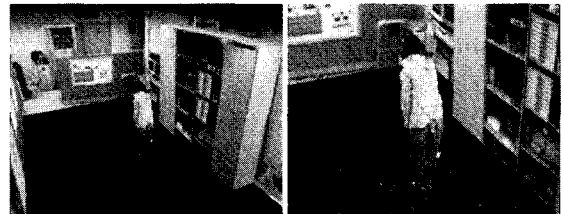
(a) 기존 방식 (SFS) (b) 제안 알고리즘

그림 12. 모델링 결과

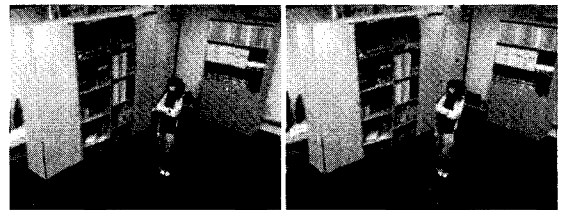
Fig. 12. Results of 3D modeling.



(a) 카메라 회전 모드



(b) 줌 아웃/인 모드



(c) 스테레오 모드

그림 13. 최종 렌더링 결과

Fig. 13. Finally rendered scenes.

표 1. 시스템 속도 분석

Table 1. Processing speed analysis for 3D modeling system.

카메라		모델링 서버	
기능	시간 (ms)	기능	시간 (ms)
획득 및 교정	28.26	데이터 수신	1.85
초기 분할	9.69	초기화	10.15
모폴로지	3.44	모델링	23.59
최종 윤곽 추출	6.40	레이블링	0.16
변이 추정	17.81	트래킹	0.08
전송	1.25	렌더링	26.30
합계	66.85 ms	합계	62.13 ms
속도	14.96 f/s	속도	16.10 f/s

모델링 서버는 독립적으로 수행될 때 평균 16 frames/sec의 속도를 보였다. 따라서 이상적으로 전체 시스템은 약 15 frames/sec의 속도로 동작할 수 있으며, 이는 영상을 끊김없이 자연스럽게 재생할 수 있는 속도

이다. 하지만 실제적으로는 각 카메라간의 비동기, 그리고 서버 측에서 데이터가 모두 도착할 때까지 대기하는 시간으로 인해 전체 시스템은 약 7~8 frames/sec의 속도로 동작하였다. 동작 속도는 영상의 해상도 및 모델의 복잡도에 따라 변한다.

시스템의 속도를 타 시스템과 동일한 환경에서 비교하기는 어렵지만 기존 SFS 방식에 대하여 참고문헌에 따르면 Matsuyama의 시스템은 6대의 PC를 사용하여 비슷한 공간 크기와 객체에 대하여 2.7 frames/sec 정도의 속도를 보여주었으며, 이를 24대의 PC를 통한 분산처리를 통해 약 6 frames/sec 까지 향상시켰다^[5]. 또한 Cheung과 Kanade의 시스템은 조금 낮은 사양의 PC 및 카메라 환경에서 약 5 frames/sec의 속도를 보여주었다^[6]. 이러한 결과들과 비교할 때 제안된 시스템은 속도면에 있어서도 매우 우수한 성능을 보여주고 있음을 볼 수 있다.

VII. 결 론

본 논문에서 우리는 임의 시점에서의 장면을 생성하기 위한 3차원 모델링 시스템을 제안하였다. 제안된 시스템은 깊이 추정을 위한 능동 장비 없이 스테레오 카메라와 일반 PC만으로 구현되었으며 평균 7~8 frames/sec의 속도로 만족할만한 영상을 재현해내었다. 하지만 아직 제안 시스템은 각 구성 요소간의 동기화 문제, 그림자나 광원의 변화에 대해 취약한 점, 그리고 복셀 기반으로 인한 매끄럽지 못한 모델 생성 등 극복해야 할 문제들을 가지고 있다. 그러나 제안 시스템은 차후 여러 3차원 복원 기술 및 응용 분야 개발에 있어 기반 시스템으로 활용될 수 있을 것이다.

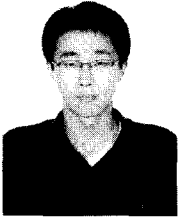
참 고 문 헌

- [1] R. Zhang, P. Tsai, J.E. Cryer and M. Shah, "Shape from Shading: A Survey," *IEEE Trans. PAMI*, vol.21, no.8, pp.690-706, 1999.
- [2] Y. Aloimonos and M.J. Swain, "Shape from Patterns: Regularization," *IJCV*, vol.2, no.2, pp.171-187, 1988.
- [3] J. Garding, "Direct Estimation of Shape from Texture," *IEEE Trans. PAMI*, vol.15, no.11, pp.1202-1208, 1993.
- [4] C. Buehler, W. Matusik, L. McMillan and S. Gortler, "Creating and Rendering Image-based Visual Hulls," *Technical Report*

MIT-LCS-TR-780, MIT, 1999.

- [5] T. Matsuyama, X. Wu, T. Takai and T. Wada, "Real-Time Dynamic 3-D Object Shape Reconstruction and High-Fidelity Texture Mapping for 3-D Video," *IEEE Trans. CSVT*, vol.14, no.3, pp.357-369, 2004.
- [6] G.K.M. Cheung, S. Baker and T. Kanade, "Visual Hull Alignment and Refinement Across Time: A 3D Reconstruction Algorithm Combining Shape-From-Silhouette with Stereo," *Proc. CVPR*, vol.2, pp.375-382, 2003.
- [7] E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*, New Jersey: Prentice Hall, 1998.
- [8] R. C. Gonzalez and R. E. Wood, *Digital Image Processing*, Addison-Wesley publishing company, 1993.
- [9] P. Kumar, K. Sengupta and S. Ranganath, "Real Time Detection and Recognition of Human Profiles using Inexpensive Desktop Cameras," *Proc. ICPR*, pp.1096-1099 2000.
- [10] 김한성, 손광훈, "영역분할 기법과 경계보존 변이 평활화를 이용한 스테레오 영상의 변이추정," 대한 전자공학회 논문지, 제41권 SP편, 제6호, 25-32쪽, 2004년
- [11] J. Weickert, "A Review of Nonlinear Diffusion Filtering," *LNCS*, vol.1252, pp.3 -28, 1997
- [12] C. Everitt, "Projective Texture Mapping," *NVIDIA SDK White Paper*, 2001.
- [13] <http://www.ptgrey.com/>

저 자 소 개



김 한 성(정회원)
1998년 연세대학교
전파공학과 학사
2001년 연세대학교
전기전자공학과 석사
2005년 연세대학교
전기전자공학과 박사

현재 일본 ATR Knowledge Science Lab 연구원
<주관심 분야: 3차원 영상 신호 처리, 혼합 현실>



손 광 훈(정회원)
대한전자공학회 논문지
제 41권 SP편 6호 참조

<주관심분야 : 통신, 컴퓨터, 신호
처리, 반도체>