

# 보조적 다시점 카메라를 이용한 깊이 맵 정확도 개선 기법에 관한 연구

엄기문\* · 허남호\* · 김진웅\* · 이수인\*

## 1. 서 론

최근 3 차원 TV 및 실감형 디지털 방송에 대한 관심이 높아지고 있는 가운데, 실감 방송 기술에 대한 연구가 국내외에서 활발하게 이뤄지고 있다. 실감 방송에 중요한 요소 중의 하나인 3차원 비디오는 기존의 하나의 시점에서 시청하던 비디오를 사용자가 원하는 다양한 시점에서 입체감을 제공할 수 있으며, 실사/CG 합성 등 가상 스튜디오 및 증강 현실, 화상회의, 원격 교육, 원격 수술, 홈쇼핑, 광고 등 여러 응용에서 이용될 수 있다. 3차원 비디오를 생성하기 위한 기법으로는 스테레오 또는 다시점 카메라로부터 얻은 스테레오 또는 다시점 비디오를 직접 이용하거나, 3차원 정보를 획득 또는 추출하여 이로부터 가상 시점 영상 생성 기법을 통해 스테레오 또는 다시점 3차원 비디오를 생성하는 기법 등이 있다. 최근에는 컴퓨팅 성능 및 소자기술의 발전으로 영상처리를 통하여 3차원 정보를 얻거나 3차원 정보 획득 장치로부터 얻은 3차원 정보를 통하여 스테레오 또는 다시점 3차원 비디오를 생성하는 기법에 대한 연구가 부각되고 있다. 이러한 기법에 있어서 고화질의 스테레오 또는 다시점 3차원 비디오를 생성하기 위

해서는 정확한 3차원 정보 획득 및 추출이 필수적이다. 또한 3차원 방송의 실현을 위해서는 3차원 정보 추출을 위한 계산 시간의 감소를 위한 노력도 필요하다.

3차원 정보 또는 깊이 정보를 추출하기 위한 기법은 능동적 기법과 수동적 기법의 두 가지 기법으로 나눌 수 있다. 능동적 기법은 레이저 스캐너(laser scanners), 구조광 패턴(structured light pattern beams), 깊이 카메라(depth camera) 등과 같이 직접 깊이 정보를 획득하는 기법으로서 정교하고 정확한 깊이 정보를 얻을 수 있는 반면에, 가격이 비싸고 실내나 정지 영상에 한정되는 등 제한적 환경에서 사용가능하다. 반면에 수동적 기법은 다시점 카메라로부터의 실루엣 정보를 이용한 객체 모델링, 다시점 스테레오 정합, SFM(Shape from Motion) 등 카메라에 의해 취득된 영상의 처리를 통하여 3차원 정보를 획득하기 때문에 상대적으로 가격이 저렴하며 간단하다. 그러나 이러한 수동적 기법은 텍스처가 없는 영역과 차폐영역 등에서의 문제점으로 인해 획득된 3차원 정보의 정확도가 능동적 기법에 비해 떨어지는 문제점을 가지고 있다. 본 논문에서는 이러한 장단점을 가지는 능동적 기법과 수동적 기법을 결합한 기법으로서 깊이 카메라로부터의 깊이 맵(depth map)의 객체 내 정확도를 개선하기 위해 다시점 카메라로

\* 한국전자통신연구원 전파방송연구단  
 ※ 본 연구는 정보통신부의 연구사업인 "지능형 통합정보방송(SmarTV) 기술 개발" 사업 지원으로 수행되었음

부터의 변이 추출을 활용하고, 이러한 변이 추출에 있어서 오정합을 줄이기 위해 깊이 맵을 변이 탐색 범위 설정에 사용하는 깊이 적응 변이 탐색 범위 설정기법을 이용한 깊이 맵 정확도 개선 기법을 제안한다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 관련 연구에 대해 간략하게 소개하고, 3장에서는 제안한 깊이 카메라로부터의 깊이 맵을 이용한 적응적 변이 탐색 범위 설정 및 이를 이용한 변이 추출 기법에 대해 다루며, 4장에서는 ground truth 데이터 및 깊이 카메라로부터 얻은 깊이 맵 데이터에 대해, 제안한 기법과 기존 기법을 적용한 결과를 비교한다. 5장에서 결론 및 추후과제로서 본 논문의 끝을 맺는다.

## 2. 기존 관련 연구

보다 정확하고 신뢰성 있는 3차원 정보 또는 깊이 정보를 획득하기 위한 연구는 전통적인 컴퓨터 비전(computer vision)의 연구 분야로서 많은 연구가 오랫동안 이뤄져 왔다. 그러나 아직까지 임의의 환경에서 자유롭게 3차원 정보를 획득할 수 있는 기법은 아직 개발되지 못하고 있다. 예를 들어 수동적 3차원 정보를 추출기법의 대표적 기법인 영역기반 스테레오 정합기법은 일정한 크기를 가지는 정합 창틀 내에서 기준 영상과 대상 영상 간 색상 유사도를 계산하고 이로부터 유사도가 높은 대응점을 찾는 기법인데, 조밀한 3차원 정보를 제공한다는 장점을 가지는 반면, 차폐영역에서의 오차, 텍스처가 반복되거나 색상이 유사한 부분에서는 오정합(false match)을 많이 발생시키는 단점을 가지고 있다. 또한 깊이 불연속 지점에서 경계가 확장되는 단점도 가지고 있다. 이러한 문제점을 해결하기 위한 방안으로 다중 창틀을 이용한 기법[1], 영상 분할을 통하여 전경(또는 관

심 객체)과 배경을 분리하는 기법[2], 실루엣을 이용한 기법[3], 계층구조를 이용한 기법[4] 등이 제안되었다. 한편, 이러한 기법 외에 여러 가지 3차원 정보 추출 기법을 결합한 복합적 기법들이 최근 많이 시도되고 있다. Dias 등은 스테레오 정합에 의한 변이 맵과 3차원 스캐너에 의해 스캐닝된 데이터를 결합하여 건물의 3차원 복원에 이용하였고[5], 엄기문[6] 등은 기준 시점에 대한 깊이 맵의 정확도를 개선하기 위하여 깊이 카메라로부터 얻은 깊이 맵과 다시점 스테레오 영상으로부터 기준 시점에 대한 변이 추출 기법에 의해 얻어진 변이 맵을 융합하는 기법을 제안하였다. 그러나, 이러한 기법들은 입력으로 사용한 깊이 데이터가 모두 오차를 가질 경우 융합을 하더라도 오차를 그대로 포함하게 되며, 일반적인 장면에서 모두 적용 가능한 융합 규칙을 찾기가 어렵다는 단점이 있다.

## 3. 제안한 깊이 카메라로부터의 깊이 맵을 이용한 깊이 적응 변이 탐색 범위 설정 및 이를 이용한 변이 추출 기법

기존 3차원 정보 추출 기법의 단점을 개선하기 위하여 본 논문에서는 깊이 카메라로부터의 깊이 맵을 이용한 깊이 적응 변이 탐색 범위 설정 및 이를 이용한 변이 추출 기법을 제안한다. (그림 1)은 제안한 기법의 전체 흐름도를 나타낸다. 제안한 기법을 보다 상세하게 살펴보면 다음과 같다.

### 3.1 3시점 영상 및 깊이 맵 획득

본 논문에서는 먼저 (그림 2)와 같은 3시점 및 깊이 카메라 시스템을 이용하여 3시점 영상 및 깊이 맵을 얻었다. 3시점 카메라로서 일본 Ikegami사의 HD(High Definition) 및 SD(Standard Definition)급 해상도가 모두 지원 가능한 CCD와 Fujinon사의 줌 렌즈가 장착된 카메라를 사용하

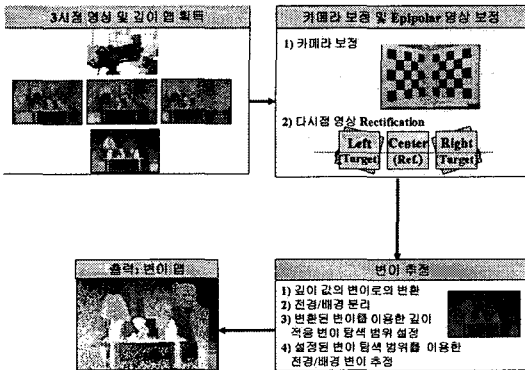


그림 1. 제안한 깊이 적응 변이 탐색범위 설정 기법을 이용한 깊이 맵 개선 기법 전체 흐름도

였고, 깊이 카메라로는 SD급(720 × 486) 해상도와 8 비트 깊이 맵 획득이 가능한 3DV systems사의 ZCam™[7]을 사용하였다.

ZCam™은 컬러 영상과 함께 카메라로부터 장면 내 각 객체까지의 상대적 거리를 8 비트(256 단계)로 양자화하여 깊이 맵을 얻을 수 있다. 또한 깊이 맵을 획득할 때 설정된 깊이 범위에 의해, 깊이 맵의 각 화소의 gray 값은 카메라로부터의 실제 거리로 환산가능하다. (표 1)은 ZCam™의 사양을 나타낸다. 획득된 영상 및 깊이 맵은 프레임별로 저장된다. 만약 깊이 범위가 길면, 깊이 맵은 객체 간 또는 객체와 배경 간의 상대적인 깊이 차이를 잘 표현하는 형태로 얻어진다. 반면에 깊이 범위를 전경(또는 관심객체)만 포함하도록 좁게 설정하면, 이 때 깊이 맵은 객체 내 세밀한

표 1. ZCam™ 사양(7)

깊이 범위	0.5 to 7 m (1.5 - 23 feet)
깊이 해상도	최대 1cm (깊이 범위에 따라 달라짐)
시야각 (Field of View)	40 도
출력 포맷	NTSC or PAL (4:3 비율, SD)
출력 신호	SMPTE 259 M - D1
프레임 율 (Frame Rate)	25-30 fps(60fps 이상도 가능)
조광 (illumination)	ANSI Z136/1-1993 파장 800-900nm

깊이 변화를 잘 표현하는 형태로 얻어진다. 본 논문에서는 장면 전체에 대한 3차원 복원 또는 다시점 3차원 비디오 생성이 목적이므로, 장면 전체에 대해 유효한 깊이 맵을 얻기 위해 깊이 맵의 깊이 범위를 전체 장면의 실제 깊이 범위를 포함하도록 설정한다. 따라서 이 경우에 장면 내 상대적인 깊이 차이는 잘 표현되지만, 객체 내 미세한 깊이 변화는 얻을 수 없다. 본 논문에서는 이러한 문제점을 다시점 카메라로부터 얻어진 다시점 컬러 영상을 이용한 변이 추출에 의해 해결하고자 한다.

### 3.2 카메라 보정(camera calibration) 및 epipolar 영상 보정(epipolar image rectification)

(그림 2)의 카메라 시스템에 의해 3시점 영상 및 깊이 맵을 획득한 다음, 각 카메라의 내부 및 외부 파라미터를 계산하기 위하여 3시점 카메라에 대한 보정을 수행한다. 카메라 보정 기법으로서 본 논문에서는 Tsai 등이 제안한 기법[8]을 사용하였으며, 구해진 각 카메라의 초점 거리 및 카메라 간격(baseline length)은 다음 절에서 기술

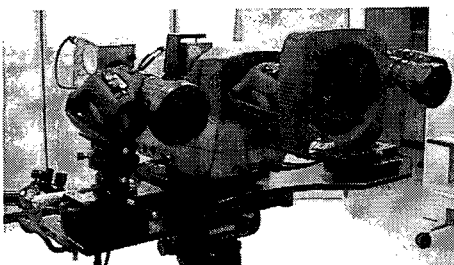


그림 2. 3시점 및 깊이 카메라 시스템

할, 가운데 시점에 대한 깊이 맵의 가운데 시점을 기준으로 한 좌측과 우측 스테레오 영상 쌍들에 대한 변이 맵으로의 변환 과정에 사용된다. 한편, 다시점 영상 간에 색상 보정(color correction)을 수행하여 변이 추출 과정에서 생길 수 있는 색상 차이로 인한 오정합을 방지하고자 하였다.

다음 단계로서 가운데-좌측 및 가운데-우측 스테레오 영상들 간 epipolar 영상 보정(rectification)을 수행한다. Epipolar 영상 보정은 각 스테레오 영상 쌍의 epipolar 선이 영상의 주사선(scan line)과 일치하도록 두 영상을 재투영(warping)하는 과정을 의미한다. 이러한 epipolar 영상 보정을 수행함으로써 변이 추출 문제는 2차원 탐색 문제에서 1차원 탐색 문제로 단순화 될 수 있다.

3.3 깊이 적응 변이 탐색 범위 설정 기법을 이용한 변이 추출

본 절에서는 앞의 단계들에서 추출된 카메라 파라미터, epipolar 보정된 3시점 영상 및 깊이 맵을 이용하여 가운데-좌측 및 가운데-우측 스테레오 영상 쌍에 대한 변이를 추정하는 기법에 대해 설명하기로 한다.

먼저 변이 추출을 위해서는 각 스테레오 영상 간의 변이 탐색 범위를 설정하여야 한다. 기존의 변이 추출 기법들에서는 대부분 사용자에게 변이 탐색 범위가 전체 장면의 변이 범위를 포함하도록 설정되며, 모든 화소에 대해 동일한 탐색 범위를 사용하여 대응점을 찾는다. 그러나 이러한 기법들은 탐색 범위 내의 epipolar 선상에 유사한 색상 분포를 가지는 점들이 여러 개 존재할 경우 오정합 발생 확률이 높다. 따라서 적절하게 변이 탐색 범위를 제한하는 것이 필요하다. 이를 위해 본 논문에서는 깊이 카메라로부터의 깊이 맵을 이용하여 적응적으로 변이 탐색 범위를 결정한다.

3.3.1 깊이 맵으로부터 변이 맵으로의 변환

변이 탐색 범위 설정을 위하여 먼저 깊이 맵으로부터 변이 맵으로의 변환을 수행한다. 먼저 깊이 카메라로부터 얻은 깊이 맵의 각 화소의 밝기 값에 대해 미리 설정된 깊이 값 범위를 이용하여 실제 깊이(카메라 광축 중심에서 장면 내 3차원 공간상의 점까지의 거리)로 변환을 수행한다. (그림 3)은 이 두 값의 상호 관계를 기술하고 있다. (그림 3)에서 실제 최소 및 최대 깊이(거리)는 사용자에게 의해 깊이 맵 획득 시 사전에 설정된 최대 및 최소 깊이 값에 해당한다. 만약 깊이 맵의 한 화소에 대한 밝기 값을  $g$ 라고 한다면, 밝기 값  $g$ 는 카메라 광축 중심으로부터 장면 내에서 대응되는 3차원 점까지의 거리  $D$ 에 해당된다. 이러한  $g$ 에서  $D$ 로의 변환은 다음 식 (1)에 의해 이뤄진다.

$$D = (256 - g) \times \frac{(D_{Max} - D_{Min})}{256} + D_{Min} \tag{1}$$

여기서,  $D_{Max}$ 과  $D_{Min}$ 은 사용자에게 의해 설정된 최대 및 최소 깊이를 의미한다.

식 (1)에 의해  $D$ 를 계산한 후에, (그림 4)의 카메라 중심  $O$ , CCD 평면 상의 점  $p$ , 그리고 3차원 공간상의 점  $P$  간의 상호 위치 관계, 가운데와 좌우측 카메라 간의 카메라 간격(baseline length)  $B$  및 가운데 카메라의 초점 거리  $f$  등을 이용하여 화소 단위의 변이  $disp_{CL}(x, y)$ 은 식 (2),(3),(4)에 의해 표현될 수 있다.

$$d_{CL}(x, y)(pixels) = \frac{disp_{CL}(x, y)}{c} \tag{2}$$

$$disp_{CL}(x, y) = \frac{f_c \times B_{CL}}{Z} \approx \frac{f_c \times B_{CL}}{D} \tag{3}$$

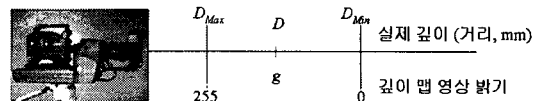


그림 3. 실제 깊이(거리)와 깊이 맵 내 밝기 값의 상호 관계

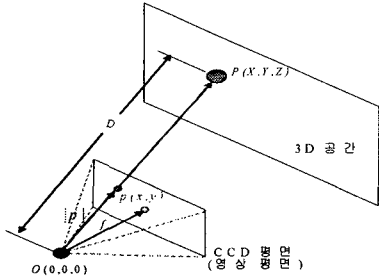


그림 4. 카메라 중심 O, CCD 평면 상의 점 p, 그리고 3차원 공간상의 점 P 간의 상호 위치 관계

$$c = \frac{W}{720}, \quad W = \frac{2}{3}(\text{인치}) \times 25.4(\text{mm}) \times \tan^{-1}\left(\frac{486}{720}\right) \quad (4)$$

여기서,  $disp_{CL}(x, y)$ 는 가운데-좌측 스테레오 영상 쌍에 대해 변환된 점  $p(x,y)$ 의 변이이며,  $f_c$ 는 가운데 카메라의 초점거리,  $B_{CL}$ 은 가운데-좌측 카메라 간격,  $Z$ 는 깊이 카메라로부터 획득된 깊이 맵 평면상의 점  $p$ 로부터 3차원 공간상으로 투영된 3차원 공간상의 점  $P(X,Y,Z)$ 까지의 z축 방향 거리를 나타낸다. 또한 CCD 크기는 2/3인치이고, 영상 크기는  $720 \times 486$ 으로 가정한다.  $W$ 는 CCD 너비이다.

본 논문에서는 지면관계상 자세한 설명은 생략한다.

이렇게 계산된 변이는 변이추출 시에 변이 탐색 영역을 설정하는데 사용된다. 그러나 실제 이 변환 과정의 구현에 있어서, 카메라 보정 오차 등에 의해 변환 오차를 포함할 수 있다. 본 논문에서는 이러한 문제점을 고려하여 적응적 변이 탐색 범위 설정 기법을 사용하였다.

### 3.3.2 적응적 변이 탐색 범위 설정

앞에서 변환된 변이  $d_{CL}(x,y)$ 를 이용하여 변이 탐색 범위를 설정하는 과정은 다음과 같다. 먼저 변이 임계값을 설정하여 기준 시점 영상에서 전경(또는 관심 객체)과 배경을 분리하고, 각 화소가 전경 또는 배경 여부에 따라 서로 다른 변이 탐색

범위를 사용한다. 이 때, 전경과 배경 분리에 사용되는 변이 임계값은 장면에 따라 경험적으로 결정된다. 만약 정합하려는 화소가 전경에 속하는 화소이며, 변환된 변이 정보가 존재할 경우에는 (그림 5)와 같이 변환된 변이에 의해 변이 탐색 범위를 설정한다. 반면에 정합하려는 화소가 배경에 속하고, 깊이 정보가 존재하지 않을 경우에는 전경에서의 최소 변이에서 전체 장면에서의 최소 변이까지의 범위를 변이 탐색 범위로 설정한다. (그림 5)에서  $p(x,y)$ 는 가운데 영상 내 정합할 화소의 영상좌표이며,  $p'(x,y)$ 는 좌측 영상에서의 대응점 후보 중의 한 화소에 대한 영상좌표,  $d_{CL}(x,y)$ 은 깊이 맵으로부터 변환된 변이를 각각 나타낸다. (그림 5)에서 보는 바와 같이 본 논문에서는 전경(또는 관심 객체) 화소에 대해 식(5)와 같이 변이 탐색 범위를 설정한다.

$$d_{CL}(x,y) - 0.2 \times d_{CL}(x,y) < d < d_{CL}(x,y) + 0.2 \times d_{CL}(x,y) \quad (5)$$

이 때, 0.2는 변환된 변이의 오차 비율로서 변환 오차를 고려하기 위한 것이다. 이와 같이 오차 비율에 의해 변이 탐색 범위를 설정함으로써, 큰 변이를 가지는 가까운 객체에 대해서는 긴 탐색 범위를, 작은 변이를 가지는 먼 객체에 대해서는 짧은 탐색 범위를 가지도록 설정하는 효과가 있다. 또한 변환된 변이에 의해 구해진 대응점 후보점 주위에서만 변이 탐색을 수행함으로써 오정합 확률을 줄일 수 있다.

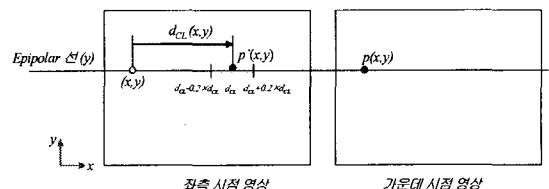


그림 5. 제한한 깊이 맵 기반 깊이 적응 변이 탐색 범위 설정 기법

3.3.3 적응적 변이 탐색 범위를 이용한 대응점 탐색 및 변이 추출

앞 절에서 기술한 적응적 변이 탐색 범위 설정 기법에 따라 설정된 변이 탐색 범위 내에서 기준 시점(여기서는 가운데) 영상의 각 화소에 대한 정합할 영상(좌측 또는 우측 영상) 내 대응점을 찾고 변이를 추출하게 된다. 대응점 탐색 기법에는 여러 가지가 있으나 본 논문에서는 전통적인 정사각형 창틀 및 식 (6)의 SSD(Sum of Squared Difference) 유사함수를 사용하였다. 대응점 결정은 SSD를 변이 탐색 범위 내에서 각 대응점 후보에 대해 계산하여 식 (7)와 (8)에 나타난 바와 같이 SSD가 최소로 되는 화소를 대응점으로 선택하고, 이때의 변이를 기준 영상의 해당 화소에 대한 변이로 기록한다.

$$C_{SSD}(x, y, d) = \sum_{n=y-L/2}^{y+L/2} \sum_{m=x-K/2}^{x+K/2} [I_{target}(x+m+d, y+n) - I_{ref}(x+m, y+n)]^2 \quad (6)$$

$$d(x, y) = \operatorname{argmin}_{d \in S} C_{SSD}(x, y, d) \quad (7)$$

$$S = \{d \mid \delta_{min} \leq d \leq \delta_{max}\} \quad (8)$$

여기서, K와 L은 SSD 계산에 사용되는 정합 창틀의 열과 행 크기에 해당하는 값이며, m 과 n은 각각 정합 창틀의 열과 행 지수(index)이다. 한편,  $\delta_{min}$  과  $\delta_{max}$ 는 각각 탐색하는 최소 및 최대 변이 값이다.

4. 실험 결과 및 토의

제안한 기법에 의한 변이 추출 성능을 검증하기 위해 Middlebury College에서 제공되는 스테레오 영상 쌍[9] (그림 6 참조) 및 (그림 2)의 카메라 시스템에 의해 얻어진 다시점 영상 및 깊이 맵에 제안한 기법을 적용하였다.



(a) 좌측 영상 (b) 우측 영상



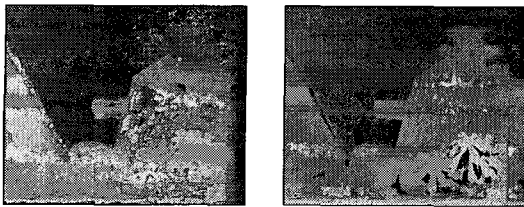
(c) 우측 영상의 Ground Truth 데이터

그림 6. 'Teddy' 스테레오 영상 및 우측 영상의 Ground Truth 데이터[10]

4.1 Ground truth 데이터를 가진 스테레오 영상에 대한 결과

(그림 6)은 'Teddy' 영상으로서 (a),(b),(c)는 좌측 영상, 우측 영상, 그리고, 우측 영상에 대한 ground truth 데이터를 나타내고 있다. 영상 크기는  $450 \times 375$ 이며, 깊이 카메라에 의한 깊이 맵 대신 구조광 패턴을 이용하여 얻어진 ground truth 데이터를 깊이 맵으로 간주하여 제안한 깊이 적응 탐색 변이 범위 설정기법을 적용하였다. 또한 SSD 유사함수와  $5 \times 5$  크기 창틀을 사용하여 화소 단위 변이 추출을 수행하였다. (그림 7)는 (그림 6(b))의 우측 영상에 대해 계산된 좌측 영상의 변이 맵들을 나타낸다.

(그림 7(a)), (그림 7(b)), (그림 7(c))는 각각 ground truth 데이터를 이용하지 않고, 고정 변이 탐색 영역을 사용한 결과, ground truth 데이터를 변이 탐색 범위 결정에 사용하되, 변이 크기에 무관하게 고정된 변이 탐색 범위를 사용한 결과, ground truth 데이터를 변이 탐색 범위 결정에 사용하고, 전경(또는 관심 객체)/배경 여부에 따라



(a) Ground truth 미사용/고정 탐색 영역 (b) Ground truth 사용/고정 탐색 영역



(c) Ground truth 사용/ 적응적 탐색 영역(제안 기법)

그림 7. 여러 가지 변이 탐색 범위 설정 기법을 이용한 'Teddy' 영상의 우측 영상에 대한 좌측 영상의 변이 추출결과

적응적으로 변이 탐색 범위를 사용한 기법에 의한 결과이다. (그림 7(b))의 경우 깊이 맵 변환 오차를 고려하여 변환된 변이 위치 좌우 양쪽으로 5 화소씩의 추가 탐색 범위를 설정하였다. (그림 7(c))의 경우, 변환된 변이 위치에서 좌우 양쪽으로 20%의 추가 탐색 범위를 설정하였다. (그림 7(a))와 (그림 7(b))를 비교하였을 때, ground truth 데이터(깊이 맵 데이터로 가정된)로부터 변환된 대응점 위치를 변이 탐색 범위의 중심점으로 선택함으로써 변이 맵 정확도가 향상됨을 볼 수 있다. 또한 (그림 7(b))와 (그림 7(c))를 비교함으로써, 제안한 적응적 변이 탐색 범위 설정에 의해 정확도가 추가적으로 개선됨을 볼 수 있다.

다음으로, RMSE(Root Mean Square Error) 및 계산 시간 측면에서 정량적인 비교 분석을 수행하였다. 이 때, 그림 6(c)의 ground truth 데이터를 참조 데이터(reference data)로 사용하였다. 각 기법에 의해 추정된 각 화소의 변이  $d_e(x,y)$  와 ground truth 데이터의 변이  $d_g(x,y)$ 에 대한 RMSE는 다음

식 (9)에 의해 계산된다.

$$RMSE = \sqrt{\left\{ \frac{1}{N} \sum_{(x,y)} (d_e(x,y) - d_g(x,y))^2 \right\}} \quad (9)$$

여기서  $N$ 은 RMSE 계산에 사용된 샘플 데이터의 개수이다.

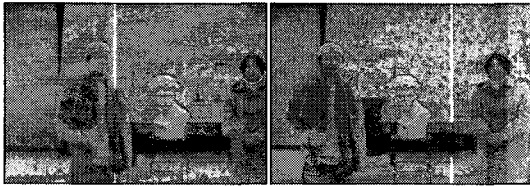
(표 2)는 각 기법의 정량적 비교 결과를 나타내고 있는데, 제안 기법의 RMSE가 다른 두 변이 탐색 범위 설정 기법들의 RMSE에 비해 작음을 볼 수 있다. 또한 계산 시간 측면에서도 제안 기법은 전경과 배경 정보 및 변환된 변이에 대한 비율에 따라 적응적으로 변이 탐색 범위를 설정함으로써 다른 두 기법에 비해 단축됨을 볼 수 있다.

#### 4.2 3시점 영상 및 깊이 맵 데이터에 대한 결과

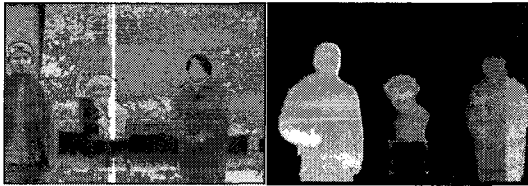
(그림 8)은 'People' 영상으로서 (a)부터 (d)까지 좌측 영상, 가운데 영상, 우측 영상, 가운데 깊이 맵을 나타내며, 깊이 맵의 범위는 1.6m~2.8m로 설정되었다. 이는 그림 8(d)에서 보는 바와 같이 전경에 해당하는 깊이 값이다. 따라서 전체 장면(scene)의 최대 및 최소 변이 외에 전경(또는 관심 객체)과 배경을 구분하기 위한 변이 임계값의 선택이 필요하며, 선택된 변이 임계값은 전경(또는 관심 객체)에 대해서는 최소 탐색 변이로,

표 2. 여러 변이 탐색 범위 설정 기법에 의한 변이 추정 결과의 정량적 비교('Teddy' 영상)

변이 탐색 범위 설정 기법	성능평가기준 RMSE (화소수)	계산 시간 (초)
깊이 맵 미사용, 고정 탐색 범위	35.67	5.156
깊이 맵 기반 고정 탐색 범위 (5화소)	20.72	1.153
깊이 맵 기반 적응 탐색 범위 (20%, 제안 기법)	19.81	0.625



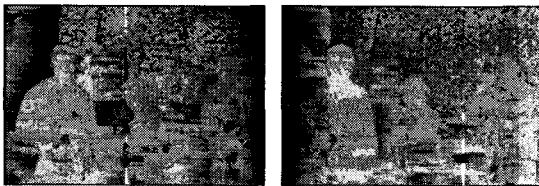
(a) 좌측 시점 영상 (b) 가운데 시점 영상



(c) 우측 시점 영상 (d) 깊이 카메라에 의해 취득된 가운데 시점 깊이 맵

그림 8. 'People' 영상 데이터 집합

배경에 대해서는 최대 탐색 변이로 사용된다. (그림 9(a))와 (그림 9(c))는 각각 (그림 8(a)) 와 (그림 8(b)), (그림 8(b))와 (그림 8(c))에 대한 변이 맵을 기존 깊이 맵을 사용하지 않은 고정 변이 탐색 범위 설정 기법과 제안한 깊이 적응적 변이



(a) 가운데-좌측 시점 변이 맵 (고정 변이 탐색 범위 사용) (b) 가운데-우측 시점 변이 맵 (고정 변이 탐색 범위 사용)

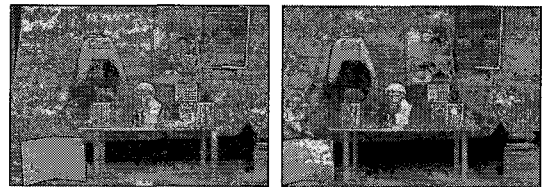


(c) 가운데-좌측 시점 변이 맵 (적응 변이 탐색 범위 사용) (d) 가운데-우측 시점 변이 맵 (적응 변이 탐색 범위 사용)

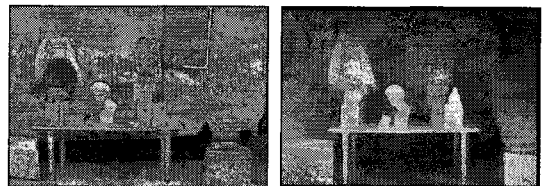
그림 9. 여러 변이 탐색 범위 설정 기법을 이용한, 'People' 영상의 두 스테레오 영상 쌍으로부터 얻어진 가운데 시점 영상에 대한 변이 추출 결과

탐색 범위 설정 기법을 사용하여 계산한 결과이다. 결과를 보면, 변이 맵의 전체적인 정확도는 이전보다 전체적인 정확도가 저하되었음을 볼 수 있다. 이러한 결과의 원인은 깊이 맵에서 변이 정보로 변환하는 과정에서 카메라 보정 오류에 의해 생긴 오차, 깊이 맵 자체의 오차, 변이 추출 과정에서의 오차 등에 의한 것으로 볼 수 있다. 그러나, 깊이 맵의 사용에 의해 변이 맵의 정확도가 변이의 연속성 반영 정도 및 정확도 측면에서 개선되었음을 볼 수 있다. 한편, 수행시간에서도 기존 기법에 비해 약 40% 단축된 것으로 나타났다. 따라서 제안한 깊이 적응 변이 탐색 영역 설정 기법이 정확도와 계산 시간 측면에서 장점을 확인할 수 있었다. 한편, (그림 8(d))와 (그림 9(c))를 비교하였을 때, 이 경우에는 전경(또는 관심 객체)에 대해서만 깊이 값이 얻어지도록 깊이 맵을 획득하였으므로 두 깊이 맵의 정확도가 거의 유사하다고 판단된다.

(그림 10(a)), (그림 10(b)) 및 (그림 10(c))는 'Background' 영상 데이터 집합을 나타내며, 이 영상의 경우 모든 화소에 대해 깊이 값이 존재하



(a) 좌측 시점 영상 (b) 가운데 시점 영상



(c) 우측 시점 영상 (d) 깊이 카메라에 의해 취득된 가운데 시점 깊이 맵

그림 10. 'Background' 영상 데이터 집합



며, 다른 변이 탐색 범위(주로 배경)가 사용되었다. (그림 11)은 각각 기존 고정 변이 탐색 범위 설정 기법을 사용한 지역 탐색(local search) 기법, 고정 변이 탐색 범위 설정 기법을 이용한 Zitnick 등[10]의 협력 스테레오 정합(cooperative stereo matching) 기법 및 제안한 깊이 적응 변이 탐색 범위 설정 기법을 이용한 지역 탐색 기법을 (그림 10)의 가운데-좌측 스테레오 영상 쌍에 적용한 결과이다. 세 기법의 결과를 보면, 제안한 기법을 이용한 결과가 나머지 기존 기법에 의한 결과에 비해 더 연속성을 가지면서 전체적으로 우수한 성능을 보여주고 있다. 그러나 일부 영역에서 발생한 오정합 화소들에 대해서는 개선이 필요하다. 한편, (그림 10(d))와 (그림 11(c))를 비교하면, 전체적으로 두 결과가 거의 유사하거나 일부 영역에서 (그림 10(d))의 결과가 더 우수한 것으로 나타났다. 다만 그림 (그림 10(d))의 일부에서 잘 구분되지 않는 깊이의 변화가 (그림 11(c))의 결과에서 약간 더 명확하게 구분되는 경향을 볼 수 있었다.

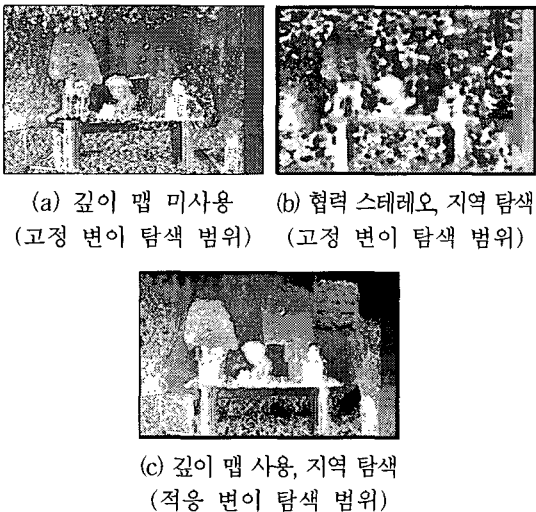


그림 11. 여러 변이 탐색 범위 설정 기법과 지역 탐색 및 협력 스테레오 기법을 사용한 'Background' 영상 데이터의 가운데-좌측 시점 스테레오 영상 쌍에 대한 변이 추출 결과

### 5. 결론 및 추후 과제

본 논문에서는 보조적 다시점 카메라의 사용으로 깊이 카메라로부터 얻어진 깊이 맵의 정확도를 높이면서, 변이 추출 시 깊이 맵을 변이 탐색 범위 설정에 이용함으로써 오정합을 줄일 수 있는 깊이 적응 변이 탐색 범위 설정 기법을 이용한 깊이 맵 정확도 개선 기법을 제안하였다.

제안한 기법은 깊이 카메라로부터 얻어진 깊이 맵을 변이 정보로 변환하여, 이 변이 값에 따라 전경(또는 관심 객체)과 배경으로 구분한 다음, 변이 크기와 위치에 따라 적응적으로 변이 탐색 범위를 설정하고, 대응점을 찾게 된다. 제안한 기법의 ground truth 및 실제 깊이 카메라로부터 얻은 깊이 맵을 사용한 모의실험 결과, 제안한 깊이 맵을 사용한 변이 탐색 범위 설정 및 깊이 적응 변이 탐색 범위 설정 기법이 기존의 깊이 맵을 사용하지 않고, 고정 변이 탐색 범위 설정 기법을 사용한 기법에 비해 변이의 정확도가 전체적인 측면에서 우수함을 볼 수 있었으며, 특히 연속적 변이 변화를 잘 반영하였다. 또한 변이 탐색 범위를 제한함으로써 계산 시간을 줄일 수 있었다. 다만 제안한 기법도 기존의 고정 창틀을 사용하여 차폐영역 문제를 현재 구현에서는 고려하지 못하였기 때문에 깊이 카메라로부터 얻은 깊이 맵에 비해 정확도가 뚜렷이 향상되는 결과는 얻지 못하였다. 따라서 깊이 카메라로부터 얻은 깊이 맵 자체에 오차가 있을 경우를 보정하기 위한 기법과 변이 추출 단계에서 적응 창틀[4] 또는 다중 창틀의 사용[1], 전역최적화기법(global optimization)[12]을 통한 변이 정확도 개선에 대한 연구가 필요하다. 한편, 이렇게 얻어진 변이 맵은 3차원 장면 복원(3D scene reconstruction)이나 다시점 비디오생성(multi-view video synthesis) 등에 이용될 수 있다.

## 참 고 문 헌

- [1] A. Fusiello, V. Roberto, and E. Trucco, "Symmetric Stereo with Multiple Windowing," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 14, no. 8, pp. 1053-1066, Dec. 2000.
- [2] C. Fehn, and P. Kauff, "Interactive Virtual View Video (IVVV) - The Bridge Between 3D-TV and Immersive TV," *Proceedings of ITCOM 2002: 3D-TV, Video & Display*, Boston, Massachusetts, Aug. 2002.
- [3] Soon-Yong Park and M. Subbarao, "A Range Image Refinement Technique for Multi-View 3D Model Reconstruction," *Proceedings of 3-D Digital Imaging and Modeling(3DIM)'03*, pp. 147-154, Banff, 2003.
- [4] Jong-Il Park, Gi Mun Um, Chunghyun Ahn, and Chieteuk Ahn, "Virtual Control of Optical Axis of the 3DTV Camera for Reducing Visual Fatigue in Stereoscopic 3DTV," *ETRI Journal*, vol. 26, no. 6, pp. 597-604, Dec. 2004.
- [5] P. Dias, V. Sequeira, F. Vaz, and J. G. M. Gonçalves, "Registration and Fusion of Intensity and Range Data for 3D Modeling of Real World Scenes," *Proceedings of 3-D [10] Digital Imaging and Modeling(3DIM)'03*, pp. 418-426, Banff, 2003.
- [6] G.-M. Um, K. Y. Kim, C. Ahn, and K. H. Lee, "There-dimensional Scene Reconstruction using Multiview Images, and Depth Camera," *Proceedings of Electronic Imaging 2005*, vol. 5664, pp. 271-280, 2005.
- [7] <http://www.3dvsystems.com>
- [8] 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.
- [9] <http://cat.middlebury.edu/stereo/>
- [10] C. L. Zitnick and T. Kanade, "A Cooperative Algorithm for Stereo Matching and Occlusion Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-22, no. 7, pp. 675-684, July 2000.
- [11] S. O. Chen, Y. P. Wong, J. K. Daniel, "Dense Stereo Correspondence Based on Recursive Adaptive Size Multi-Windowing," *Proceedings of Image and Image and Vision Computing 2003*, pp. 256-259, 2003.
- [12] V. Kolmogorov and R. Zabih, "What Energy Functions Can be Minimized via Graph Cuts?" *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. PAMI-26, no. 2, pp. 147-160, 2004.



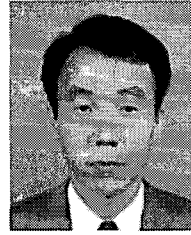
엄 기 문

- 1991년 서강대학교 전자공학과(공학사)
- 1993년 서강대학교 전자공학과(공학석사)
- 1998년 서강대학교 전자공학과(공학박사)
- 1998년~2000년 한국전자통신연구원 박사후 연수연구원
- 2000년~현재 한국전자통신연구원 선임연구원
- 2001년~2002년 캐나다 CRC 방문연구원
- 관심분야 : 3DTV, 휴먼 팩터, 비디오기반 렌더링, 컴퓨터 비전



허 남 호

- 1992년 포항공과대학교 전자전기공학과(공학사)
- 1994 포항공과대학교 전자전기공학과(공학석사)
- 2000년 포항공과대학교 전자전기공학과(공학박사)
- 2000~현재 한국전자통신연구원 선임연구원,  
3DTV시스템연구팀 팀장
- 관심분야 : 3DTV, 모터제어, 휴먼팩터



이 수 인

- 1985년 경북대학교 전자공학과(공학사)
- 1989년 경북대학교 전자공학과(공학석사)
- 1996년 경북대학교 전자공학과(공학박사)
- 1990년~현재 한국전자통신연구원 책임연구원, 방송시스  
템연구그룹 그룹장
- 관심분야 : 지상파 DTV/DMB/DCATV 방송시스템, 3DTV  
전송 시스템



김 진 응

- 1981년 서울대학교 전자공학과(공학사)
- 1983년 서울대학교 전자공학과(공학석사)
- 1993년 미국 Texas A&M University(공학박사)
- 1983년~현재 한국전자통신연구원 책임연구원
- 1983년~1984년 스웨덴 LM Ericsson사 방문연구원
- 1995년~1996년 미국 Texas A&M University 연구교수
- 2004년~2005년 미국 Columbia University 방문연구원
- 관심분야 : 신호처리, 대화형 방송시스템, 3DTV, 3D DMB  
시스템