

복합형 카메라 시스템을 이용한 고화질 3차원 비디오 생성 방법

3D Video Generation using Hybrid Camera System

이은경 · 호요성
Eun-Kyung Lee · Yo-Sung Ho

본 논문은 다시점 비디오와 그에 상응하는 다시점 깊이영상을 생성하기 위해 다시점 비디오 카메라와 깊이 카메라를 결합한 복합형 카메라 시스템을 제안한다. 제안한 복합형 카메라 시스템을 이용해 3차원 비디오를 생성하기 위해서는 우선 다시점 카메라와 깊이 카메라를 이용해 다시점 비디오와 깊이영상을 획득한다. 획득한 깊이영상을 이용해 3차원 워핑(warping)을 적용하여 각 다시점 카메라를 위한 초기 깊이 정보를 예측하고 예측한 초기 깊이 정보를 이용해 세그먼트 기반의 스테레오 정합을 수행하여 초기 깊이영상을 생성한다. 3차원 워핑을 이용해 초기 깊이 정보를 예측하는 것은 다시점 카메라의 각 시점에서의 초기 깊이영상을 계산하기 위한 것이다. 본 논문에서는 물체의 외곽선 영역의 깊이영상 정확도를 높이기 위해 비디오 앞 뒤 영상의 색상 차분 정보를 이용해 움직임 영역을 추출한다. 마지막으로 고화질의 다시점 깊이영상을 생성하기 위해서 제안한 비용함수를 이용해 깊이영상을 정제한다. 제안한 복합형 카메라 시스템은 기존의 깊이영상 예측 방법보다 다시점 깊이영상을 포함하는 3차원 비디오를 빠르고 정확하게 생성할 수 있었다. 이는 다양한 3차원 응용 분야를 위해 기존의 방법보다 자연스러운 3차원 영상을 생성할 수 있었다.

주제어: 깊이 비디오 생성, 깊이 카메라, 깊이영상 생성, 다시점 카메라 시스템

In this paper, we present a hybrid camera system combining one time-of-flight depth camera and multiple video cameras to generate multi-view video sequences and their corresponding depth maps. In order to obtain the multi-view video-plus-depth data using the hybrid camera system, we capture multi-view videos using multiple video cameras and a single view depth video with the depth camera. After performing a three-dimensional (3D) warping operation to obtain an initial depth map at each viewpoint, we refine the initial depth map using segment-based stereo matching. In order to reduce mismatched depth values along object boundaries, we detect moving objects using color difference between frames. Finally, we recompute the depth value of each pixel in each segment using pairwise stereo matching with a new cost function. Experimental results show that the proposed hybrid camera system produces multi-view video sequences with accurate depth maps, especially along the boundary of objects. In addition, it is suitable for generating more natural 3D views for 3D TV than previous works.

Keywords: Depth video generation, Depth camera, Depth image generation, Multi-view camera system

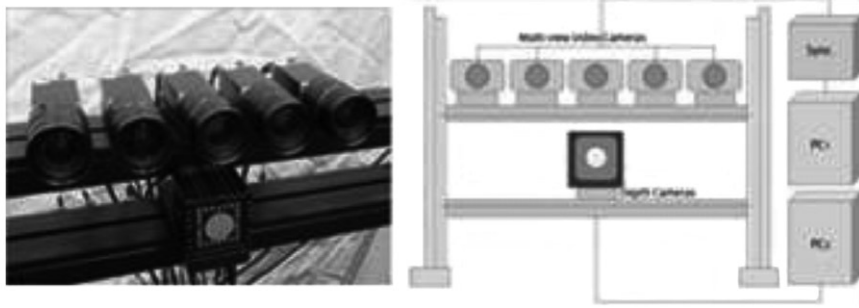


그림1. 제안한 복합형 카메라 시스템

I. 서론

3차원 멀티미디어 응용 분야에 대한 관심이 증가하면서 3차원 비디오를 생성하고자 하는 노력이 많은 분야에서 진행되고 있다. 3차원 TV와 같은 차세대 3차원 응용에서 고화질의 3차원 비디오를 지원하기 위해서는 정확한 깊이 정보를 포함하는 깊이영상을 효과적으로 생성해야 한다[1]. 최근 국제 표준화 그룹인 Moving Picture Experts Group(MPEG)에서는 이러한 자유시점 TV와 3차원 TV에 대한 중요성을 인식하고 3차원 비디오 부호화 기술의 표준화 작업을 활발히 진행하고 있다[2].

머지않은 미래에 시청자들은 많은 응용 분야를 통해 3차원 깊이를 체험할 수 있을 뿐만 아니라 3차원 비디오를 이용해 시청자가 원하는 위치의 시점을 선택하여 볼 수도 있을 것이다[3]. 이미 유럽의 ATTEST 프로젝트는 깊이 카메라를 이용한 3차원 TV 시스템을 선보였으며 광주과학기술원 실감방송연구센터에서도 깊이 카메라를 이용하여 3차원 방송 콘텐츠를 제작했다. 이와 같이 3차원 비디오를 이용해 임의 시점에서 3차원 장면을 생성하기 위해서는 정확한 깊이 정보를 획득하는 것이 무엇보다 중요하다. 오래전부터 컴퓨터 비전 분야에서는 스테레오 정합 기술을 이용한 깊이영상을 생성하는 방법이 연구되어왔다[4]. 그러나 신뢰할만한 깊이영상을 생성하는 것은 여전히 해결해야 할 문제점으로 남아 있다. 깊이영상을 생성하는 방법은 크게 능동적 깊이센서 방식(active depth sensors)과 수동적 깊이센서 방식(passive depth sensors)으로 나누어진다. 능동적 깊이센서 방식은 레이저 센서, 적외선 센서, 패턴센서와 같은 물리적인 센서장치를 이용해 3차원 공간상의 깊이 정보를 직접 획득하는 방식이다 [5],[6]. 능동적 깊이센서 방식은 정확한 깊이 정보를 실시간으로 획득할 수 있지만 저해상도의 깊이영상만을 제공할 뿐만 아니라 큰 비용이 든다는 단점이 있다.

반면에 수동적 깊이센서 방식은 스테레오 및 다시점 카메라에서 획득한 영상을 이용해 스테레오 정합을 수

행하여 3차원 정보를 획득하는 방법이다[7],[8]. 수동적 깊이센서 방식은 고해상도의 깊이영상을 생성할 수 있지만 능동적 깊이센서 방식에 비해 부정확한 깊이영상을 생성한다.

이러한 두 방식의 단점을 해결하고 좀 더 정확한 3차원 깊이 정보를 생성하기 위하여 능동적 깊이센서 방식과 수동적 깊이센서 방식을 결합한 복합형 방법들이 제안되었다[9]~[11]. 한국전자통신원(ETRI)에서는 깊이 카메라를 3시점 카메라 시스템을 결합한 복합형 카메라 시스템을 소개하였다. 그러나 기존의 복합형 카메라 시스템은 깊이 카메라를 이용하여 고품질의 깊이영상을 생성할 수 있지만 저해상도의 깊이 카메라를 기준으로 다시점 카메라의 해상도를 낮추어 깊이영상을 생성하였기 때문에 저해상도의 3차원 비디오를 생성하였다[12].

최근 정확한 깊이영상을 생성하기 위해 깊이 카메라에서 획득한 깊이 정보를 보정하는 방법들이 제안되었다[13]~[15]. 그러나 이 방법들은 모두 정지된 장면에 대한 깊이영상을 생성하였고 저해상도의 깊이영상만 생성할 수 있다. 미래의 3차원 응용 분야에서는 고품질의 3차원 비디오를 요구하기 때문에 우리는 고화질의 다시점 3차원 비디오를 생성할 필요가 있다.

본 논문에서는 고해상도의 다시점 카메라와 저해상도 깊이 카메라를 결합하여 고해상도의 다시점 영상과 그에 상응하는 깊이영상을 포함하는 3차원 비디오를 생성하는 방법을 제안한다. 제안하는 방법은 깊이 카메라에서 획득한 깊이 정보를 이용해 다시점 카메라를 위한 고품질 3차원 비디오를 생성한다. 본 논문의 장점은 제안한 복합형 카메라 시스템과 그 시스템에 적합한 비용 함수를 이용해 고화질의 3차원 비디오를 기존의 깊이 생성 방법보다 짧은 시간에 더 정확한 다시점 깊이영상을 생성할 수 있다는 것이다.

본 논문의 구성은 다음과 같다. II장에서 제안하는 복합형 카메라의 시스템의 구조와 3차원 비디오 생성 방법에 대해 간략히 소개한다. III장에서는 3차원 비디오를 생성하는 방법을 단계적으로 설명한다. IV장에서 제안한 시스템을 통해 생성한 3차원 비디오 결과를 분

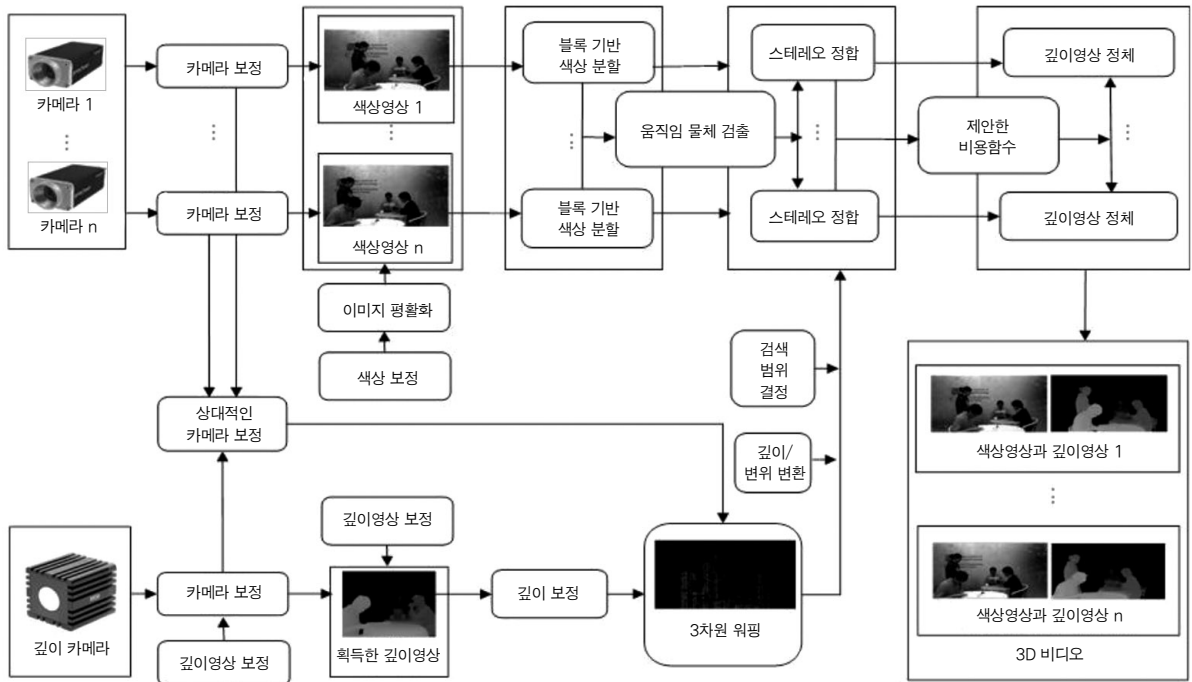


그림 2. 3차원 비디오 생성을 위한 전체 구성도

표 1. 복합형 카메라 시스템 구성

장치	상세정보	
다시점 카메라	출력형식	NTSC/PAL (16:9)
	해상도	1920(h) × 1080(v)
깊이 카메라	깊이범위	0.5~5.0m
	시야각	43.6° (h) × 34.6° (v)
	출력양식	NTSC/PAL(4:3)
동기화 신호	해상도	176(h) × 144(v)
	출력형식	SD/HD 비디오 생성

석하고 마지막으로 V장에서 결론을 맺는다.

III. 복합형 카메라 시스템

그림 1은 제안하는 복합형 카메라 시스템과 그 구성을 보여준다. 제안하는 복합형 카메라 시스템은 5대의 고해상도 다시점 카메라와 1대의 저해상도의 깊이 카메라로 이루어진다. 각 카메라는 비디오 저장 보드가 갖춰진 PC와 연결된다. 또한 동기화 신호 재생기가 모든 카메라에 연결되어 동기화 신호를 연속적으로 보낸다. 제안하는 복합형 카메라 시스템은 프레임마다 다시점 영상과 깊이영상을 동시에 획득할 수 있다. 표 1은 복합형 카메라 시스템의 구성을 보여준다. 복합형 카메

라 시스템은 다시점 카메라, 깊이 카메라, 동기화 세 부분으로 구성된다. 제안하는 복합형 카메라 시스템은 매프레임마다 다시점 카메라에서 획득한 다시점 영상과 깊이 카메라로부터 획득한 깊이영상을 제공한다.

그림 2는 다시점 깊이영상 생성을 위한 전체 구성도이다. 다시점 깊이영상을 생성하기 위해 먼저 모든 카메라에 대한 정보를 알기 위해서 카메라 보정 (camera calibration)을 수행한다[16]. 깊이 카메라에서는 획득한 깊이영상의 노이즈를 제거하고 깊이영상을 각 다시점 카메라로 3차원 워핑을 수행하여 스테레오 정합시 초기 깊이 정보로 사용한다. 다시점 카메라에서는 다시점 영상 평활화(image rectification)와 색상 보정(color correction)을 수행한다. 모든 색상영상은 세그먼트 단위로 나누어 세그먼트 기반의 스테레오 정합을

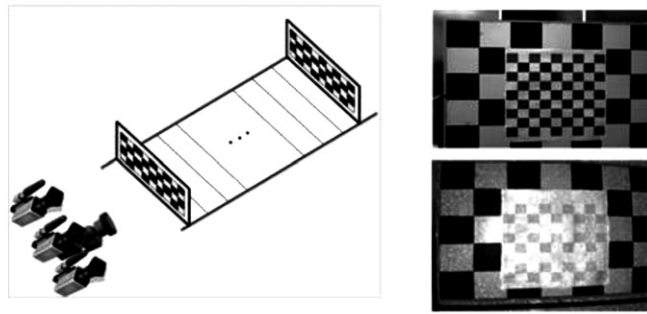


그림 3. 깊이 보정을 위한 보정패턴 획득

수행하여 초기 깊이영상을 생성한다. 이때 앞에서 수행한 위평한 초기 깊이 정보를 이용하여 정합의 검색범위를 결정한다. 세그먼트 단위로 스테레오 정합을 수행해 생성한 초기 깊이영상은 제한한 깊이 정제 함수를 이용해 다시 화소 단위로 정합하여 최종적으로 다시점 깊이영상을 생성한다.

III. 3차원 비디오 생성을 위한 전처리 과정

1. 영상 보정

복합형 카메라 시스템을 이용하여 다시점 영상과 깊이영상을 획득한 후 전처리 과정을 거친다. 전처리 과정은 크게 카메라 보정, 영상 평활화, 다시점 색상 보정으로 나뉜다. 카메라 보정은 스테레오 카메라와 깊이 카메라의 내부 및 외부 인자를 찾는 과정이다. 카메라 보정을 통해 깊이 카메라와 다시점 카메라의 내부 행렬 K_s, K_h , 회전 행렬 R_s, R_h , 이동 행렬 t_s, t_h 을 추정한 후, 각 카메라에 대한 투영 행렬인 P_s, P_h 를 계산한다. 식 (1)과 식(2)는 계산한 투영 행렬을 보여준다[17].

$$P_s = K_s[R_s | t_s] \tag{1}$$

$$P_h = K_h[R_h | t_h] \tag{2}$$

카메라 보정 후 영상 평활화를 통해서 다시점 영상간의 수직위치가 일치하도록 교정한다. 다시점 영상을 이용해 스테레오 정합을 하기 위해서는 먼저 모든 카메라의 수직위치가 동일하다는 가정하에 수행한다. 그러나 실제 영상 획득 과정에서는 모든 카메라는 사람에 의해 수동적으로 정렬되기 때문에 수직방향에도 변이가 존재하게 된다. 본 논문에서는 다시점 영상간에 존재하

는 수직 변이를 제거하기 위해 최적의 기준선을 정의하고 2차원 변환을 통해 카메라의 방향을 보정하는 영상 평활화 방법을 적용하였다[18].

모든 다시점 영상을 평활화 한 후 다시점 영상 보정을 통해 다시점 영상간의 색상 특성을 균일하게 한다. 동일한 기종의 카메라로 다시점 영상을 획득한다 하더라도 각 카메라의 특성과 주변 환경에 따라서 각각 다른 특성을 가지게 되기 때문에 획득한 시점간의 색상 불일치가 발생한다. 이러한 다시점 영상간의 색상 불일치를 해결하기 위해 본 논문에서는 보정의 기본이 되는 색상 참조표를 만들고 이 참조표에 획득한 다시점 영상의 색상을 매핑하는 방법을 통해 다시점 영상의 색상을 보정하는 방법을 적용하였다[19].

2. 깊이영상 보정

일반적으로 깊이 카메라는 획득하고자 하는 장면의 주변 환경에 매우 민감하다. 즉 물체의 움직임이나 색에 따라 같은 거리에서 깊이 정보를 획득한다 하더라도 획득되는 값이 다르다. 이러한 문제를 해결하기 위해 본 논문에서는 획득하는 환경에 적합하게 일정 거리마다 깊이 정보를 획득하여 획득한 깊이 정보와 실제 깊이 정보간의 관계를 정의하는 보정곡선을 생성하고 깊이 카메라의 깊이 측정 특성을 분석하여 깊이영상을 보정한다. 그림 3의 왼쪽은 깊이영상 보정을 위한 보정패턴 영상 획득 방법을 보여주고 그림 3의 오른쪽 그림은 실제 획득한 보정패턴과 깊이영상이다. 일정 거리마다 보정패턴 영상을 획득하고 실제 거리와 깊이 카메라를 통해 획득한 거리 정보를 분석하여 깊이 카메라가 가지고 있는 깊이 왜곡 특성을 분석하고 이를 보정한다.

깊이 보정을 수행하기 위해 먼저 획득한 모든 패턴에서 상응하는 교차점의 위치를 추출한다. 모든 패턴영상은 이미 영상 평활화를 거친 영상이기 때문에 모든 교차점은 수평방향으로만 변이(disparity)를 갖는다. 시점이 다른 두대의 카메라에서 획득한 패턴의 교차점 사이의 변이 정보는 식 (3)을 통해서 실제 깊이 정보로 변

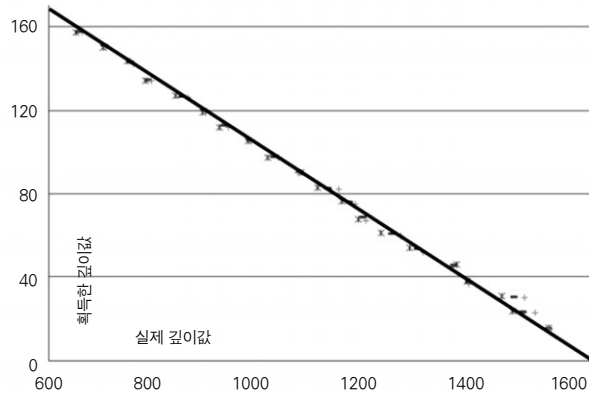


그림 4. 생성한 룩업 테이블을 이용한 보정결과

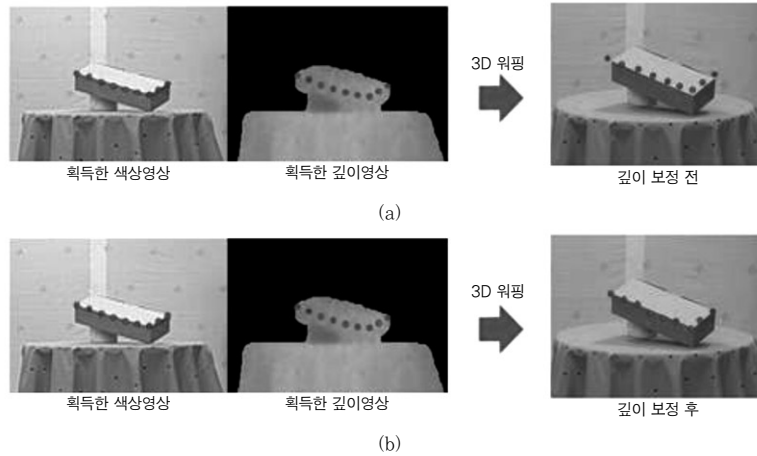


그림 5. 보정한 깊이영상

환이 가능하다.

$$Z(p_x, p_y) = \frac{f \cdot B}{D_h(p_x, p_y)} \quad (3)$$

여기서 f 는 카메라 초점거리, B 는 카메라간의 거리, $D(p_x, p_y)$ 는 교차점 간의 변이, 그리고 $Z(p_x, p_y)$ 는 구하고자 하는 실제 깊이 정보이다. 이 깊이 정보가 실제 획득한 깊이영상에서는 어떻게 표현되었는지 확인할 수 있다. 이상적으로 계산을 통해 구한 실제 깊이 정보와 깊이 카메라를 통해 획득한 깊이 정보는 선형관계를 가져야 한다. 그러나 실제 획득한 환경에 따라 획득한 깊이와 실제 깊이 사이에는 약간의 왜곡이 존재한다. 이러한 왜곡을 보정하기 위해 우리는 각 깊이영상의 x, y

좌표와 획득한 깊이값 그리고 실제 계산을 통해 획득한 깊이값을 이용해 4차원의 룩업 테이블을 생성한다[14]. 그림 4는 보정패턴의 임의의 세점을 위치를 이동시키며 획득했을 때 획득한 깊이값과 룩업 테이블을 통해 보정된 실제 깊이값 사이의 관계를 그림으로 나타낸 것이다.

그림 5는 보정한 깊이영상의 정확도를 확인하기 위해 3차원 워핑을 통해 물체의 윤곽선 부분에 보정한 깊이값을 투영해 본 결과이다. 그림의 결과에서 확인할 수 있듯이 깊이 보정을 통해 획득한 깊이값의 정확도가 향상된 것을 확인할 수 있었다. 그러나 그림 6에서 보인 것처럼 깊이 카메라의 특성 때문에 색상이 어두운 영역에서는 획득한 깊이에 많은 에러가 포함되어 있는 것을 확인할 수 있다. 두 점 A와 B는 실제 비슷한 깊이값을 가지는 영역이지만 그림 6에서 보이는 것처럼 실제 획득한 깊이값은 큰 차이를 가지는 것을 알 수 있다. 따라서 정확한 깊이영상 생성을 위해서는 획득한 깊이

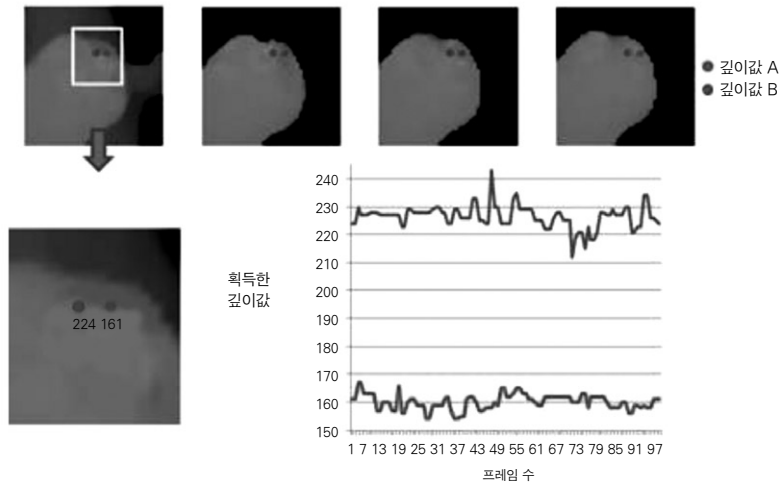


그림 6. 정지 영역에 대한 깊이영상



그림 7. 3차원 워핑한 결과

영상을 기반으로 다시 정합하는 과정이 필요하다.

$$R'_{ori} = R_s \cdot R_s^{-1} = I \tag{4}$$

IV. 3차원 비디오 생성

$$t'_{ori} = t_s - t_s = O \tag{5}$$

1. 3차원 워핑

본 논문은 깊이 카메라로 획득한 깊이 정보를 다시 점 영상의 초기 깊이 정보로 간주한다. 다시점 영상의 색상 정보와 상응하는 깊이 정보를 정합하기 위해 다시점 카메라와 깊이 카메라를 각각 독립적으로 카메라 보정 (camera calibration)한다. 결과적으로 깊이 카메라와 다시점 카메라의 투영 행렬을 다음과 같이 얻는다. 여기서 K_s, R_s, t_s 는 깊이 카메라의 내부 및 외부 인수 행렬이고 K_h, R_h, t_h 는 다시점 카메라의 내부 및 외부 인수 행렬이다. 깊이 카메라와 다시점 카메라의 상대적인 위치관계를 계산하기 위해 깊이 카메라의 위치를 기준으로 다시점 카메라의 위치를 결정한다. 먼저 식 (4)와 식 (5)를 이용하여 깊이 카메라의 회전 행렬 R_s 를 단위 행렬 I 로 이동 행렬 t_s 를 영행렬 O 으로 변환한다.

그런 다음, 식 (6)과 같이 깊이 카메라의 회전 행렬 R_s 의 역행렬 R_s^{-1} 를 각 다시점 카메라의 회전 행렬 R_h 에 곱하고 식 (7)을 이용해 깊이 카메라의 이동 행렬 t_s 를 각 다시점 카메라의 이동 행렬 t_h 에서 빼준다. 결과적으로 깊이 카메라에 대한 다시점 카메라의 상대적 위치인 회전 행렬 R'_h 과 이동 행렬 t'_h 를 결정한다.

$$R'_h = R_h \cdot R_s^{-1} \tag{6}$$

$$t'_h = t_h - t_s \tag{7}$$

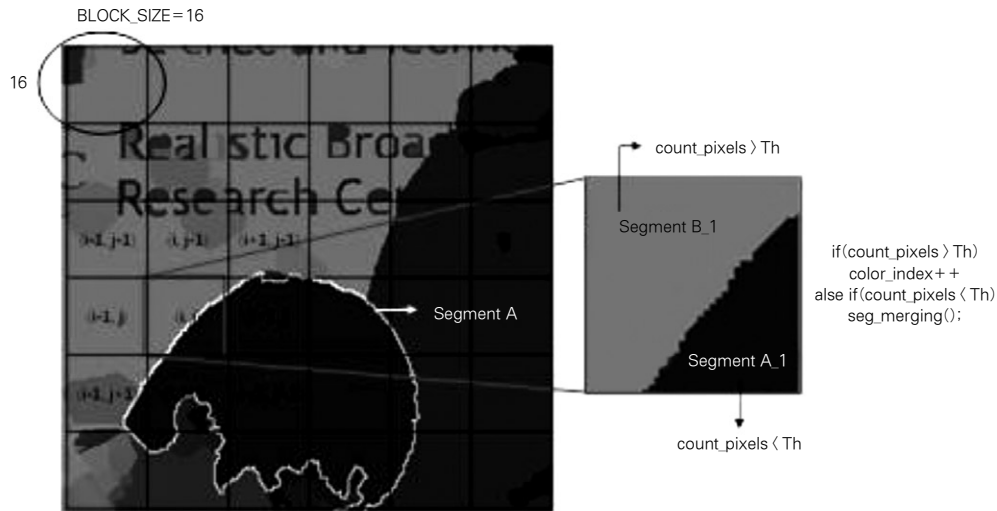


그림 8. 세그먼트 재정의

마지막으로 깊이 카메라의 색상영상의 화소위치 (p_{sx}, p_{sy}) 와 상응하는 깊이영상의 화소값이 $D_s(p_{sx}, p_{sy})$ 일 때 3차원 화소 $p_s = (p_{sx}, p_{sy}, D_s(p_{sx}, p_{sy}))$ 를 식 (6)과 식 (7)을 이용해 계산한 투영 행렬 P'_h 을 이용하여 식 (8)와 같이 3차원 워핑을 수행한다.

$$P'_h = K_h [R'_h | t'_h] \quad (8)$$

$$p_h = P'_h \cdot P_s^{-1} \cdot p_s \quad (9)$$

여기서 $p_h = (p_{hx}, p_{hy}, 1)$ 는 p_s 에 상응하는 다시점 영상의 위치 정보 (p_{tx}, p_{ty}) 를 나타낸다. 또한 p_t 에서의 깊이 정보 $D_h(p_{hx}, p_{hy})$ 는 식 (10)와 같다. 그림 7은 3차원 워핑을 통해 얻어진 초기 깊이영상을 보여준다.

$$D_h(p_{hx}, p_{hy}) = (t_{hz} - t_{sz}) + D_s(p_{sx}, p_{sy}) \quad (10)$$

2. 세그먼트 기반 스테레오 정합

본 논문은 스테레오 정합을 수행하는 동안 세그먼트 단위로 3차원 워핑을 하고 생성한 초기 깊이 정보를 고려하여 다시점 영상의 정확한 깊이영상을 생성한다. 본 논문에서는 mean-shift 기반의 색상 분할 방법을 이용하여 다시점 영상을 색상 분할한다[20]. 제한된 방법은 3차원 워핑된 세그먼트를 기반으로한 스테레오 정합

을 수행하여 깊이영상을 생성한다. 이때 세그먼트의 크기가 너무 클 경우 하나의 세그먼트가 하나의 깊이 값으로 정의되기 때문에 본 논문에서는 최소 128화소 이상, 최대 256 화소 이내의 크기로 세그먼트 크기를 다시 정의한다. 그림 7은 세그먼트를 다시 정의하는 방법을 보인 것이다.

모든 영상을 먼저 16×16 블록으로 나눈다. 그리고 하나의 블록안에 각 세그먼트를 하나의 세그먼트 단위로 정의한다. 그 다음 각 세그먼트의 크기가 미리 정의한 최소 세그먼트 크기보다 작은 경우는 주변 블록을 탐색하여 블록을 정합한다. 정합을 위한 블록 탐색 순서는 대각선 방향을 포함하여 오른쪽, 아래, 왼쪽, 위쪽 방향으로 진행한다. 왜냐하면 아직 정합되지 않은 영역을 먼저 정합하고 이후 정합이 이루어지지 않은 이전 블록을 탐색하기 위해서이다. 이러한 과정을 앞서 정의한 최대 블록크기를 만족할 때까지 수행한다.

예를들어 그림 8에서 보는 것처럼 현재 블록안에는 2개의 세그먼트가 존재한다. 세그먼트 B_1은 임계값을 만족하지만 세그먼트 A_1은 최소 세그먼트 크기를 만족하지 않는다. 새롭게 세그먼트를 정의하기 위해 가장 순서가 빠른 위치인 오른쪽 블록에서 같은 세그먼트를 찾아 하나의 세그먼트로 정의한다. 이 경우는 한 번의 주변 탐색을 통해 조건을 만족하였다. 그러나 경우에 따라 조건을 만족하지 못하는 경우 탐색순서에 따라서 이와 같은 과정을 반복한다.

색상 분할에 기반으로한 스테레오 정합에서 일반 자연영상은 비슷한 변이를 가지는 영역으로 분할되고 세그먼트는 하나의 변이를 가진다고 가정한다. 각 세그먼트에 초기 변이를 설정하기 위해서 우선 III장에 소개한 변이와 깊이 변환 식 (3)을 이용해 초기 깊이영상의 깊이 정보를 변환한다.

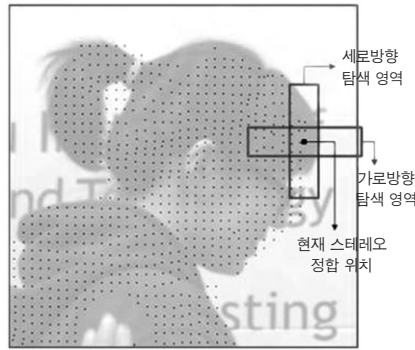


그림 9. 탐색 영역 설정

각 세그먼트의 변이를 결정하기 위해 우리는 초기 깊이값을 위한 탐색영역을 설정한다. 그림 8에서 보는 것처럼 3차원 워핑을 통해 얻은 초기 깊이영상은 실제 색상영상과 경계영역에서 잘 일치하지 않는다. 이 때문에 경우에 따라서는 깊이값이 잘못된 영역에 위치하는 문제가 발생한다. 이렇게 깊이값이 잘못 설정될 경우 잘못된 정합결과를 가져올 수 있기 때문에 본 논문에서는 초기 깊이값 설정을 위한 영역을 그림 9처럼 가로, 세로방향으로 확대하여 설정한다. 그리고 설정한 영역 안의 초기 깊이값을 이용해 재정의한 세그먼트를 기반으로 스테레오 정합을 수행한다. 초기 깊이값을 기반으로 MAD(mean absolute difference)값이 가장 작은 영역을 정합 위치로 결정하여 초기 깊이영상을 생성한다.

각 세그먼트의 변이를 결정하기 위해 우리는 먼저 전경영역에 대한 변이를 계산한다. 식 (11)은 전경영역의 변이를 계산하기 위한 식이다. 앞에서 정의한 가로방향과 세로방향 초기 변이값을 이용해 최소의 MAD 값을 가지는 변이를 계산한다. i 는 현재 정합할 세그먼트 번호이고 a 와 b 는 가로방향과 세로방향의 초기 깊이값이다. 본 논문에서는 초기 깊이값을 이용해 정합범위를 줄였다. 이는 색상 특성이 없는 영역에서의 잘못 정합되는 문제를 줄일 수 있다. 정합을 위한 검색범위는 초기 깊이값을 기준으로 -5와 +5 범위의 값을 계산하였다.

$$FG-d_i(Initdisp) = \min(\min_{j=0}^a MAD(j), \min_{k=0}^b MAD(k)) \quad (11)$$

획득한 깊이영상을 통해 알 수 있듯이 획득한 깊이 영상은 전경에 대한 깊이값만 가지고 있다. 따라서 초

기 깊이 정보가 없는 배경영역에 대한 깊이영상을 생성하기 위해서는 배경영역에 대한 정합을 따로 처리해 주어야 한다. 식 (12)는 배경영역을 위한 정합식이다. 대부분의 스테레오 정합에서와 같은 방법으로 배경영역에 대한 최소/최대 변이범위를 정해주고 최소 MAD를 갖는 변이를 찾는다.

$$BG-d_i(Initdisp) = \min(\min_{i=\min Disp}^{\max Disp} MAD(i)) \quad (12)$$

마지막으로 식(13)을 이용해 각 세그먼트의 최종 변이를 결정한다. 이미 계산한 전경영역과 배경영역에 대한 MAD 중에서 다시 최소의 MAD를 가지는 변이를 현재 세그먼트의 최종 변이로 선택한다.

$$d_i(Initdisp) = \min(FG-d_i(InitDisp), BG-d_i(InitDisp)) \quad (13)$$

색상영역 기반의 스테레오 정합 방법을 통해 깊이영상을 생성하는 경우 영역이 얼마나 정확하게 나누어졌는지가 최종 깊이영상에 영향을 미칠 수 있다. 특히 이러한 문제점은 물체의 경계부분이나 비슷한 색상의 물체가 존재하는 경우 많이 발생한다. 본 논문에서는 이러한 영역 기반 스테레오 정합의 단점을 해결하기 위해 일차적으로 세그먼트를 적합한 크기로 다시 나누어 정의하였으며 다음 단계에서 진행되어질 화소 단위의 스테레오 정합을 통해 잘못된 영역 분할로 발생하는 부정확한 깊이값 문제를 보완할 수 있다.



그림 10. 움직임 영역 추출

3. 움직임 영역 분리

일반적으로 스테레오 정합을 통해 생성된 초기 깊이 영상은 전경영역과 배경영역의 경계부분에서 많은 불일치를 가진다. 본 논문에서는 깊이영상의 정확도를 높이기 위해 움직임 정보를 찾고 이를 이용해 전경영역을 추출한다.

그림 10은 이전 프레임과 현재 프레임의 색상 차이를 이용해 현재 프레임에서의 움직임 영역을 추출한 결과이다. 움직임이 없는 영역은 이전 프레임과 현재 프레임이 차이가 없는 영역으로 이전 프레임에서 생성한 깊이 정보를 참조할 수 있다. 차분영상은 식 (14)를 이용하여 RGB 차분의 합이 임계값 이상일 경우 움직임 영역으로 판단한다. 본 논문에서는 색상 차이값이 10 이상인 경우를 움직임 영역으로 결정하였다.

$$E_n(x,y) =$$

$$\sqrt{(R_{n-1}(x,y)-R_n(x,y))^2 + (G_{n-1}(x,y)-G_n(x,y))^2 + (B_{n-1}(x,y)-B_n(x,y))^2} \quad (14)$$

4. 깊이영상 정제

색상 분할에 기반한 스테레오 정합 방법으로 깊이 정보를 획득할 경우 하나의 세그먼트는 하나의 변이를 갖는다. 그러나 실제 환경은 연속적인 부드러운 형태의

깊이 정보로 이루어져 있다. 이러한 문제점을 해결하기 위해 일반적으로 다시 화소 단위의 깊이영상 정제를 수행한다. 그러나 대부분 현재 제안된 기존의 스테레오 정합 방법의 경우 반복적인 계산을 통해 깊이 정제를 수행하기 때문에 많은 시간과 메모리를 요구한다. 그러나 많은 3차원 응용 분야에서는 고화질과 고해상도의 콘텐츠를 요구한다. 이러한 응용에 맞는 고화질의 콘텐츠를 생성하기 위해서 계산시간과 메모리 사용이 문제가 될 수 있다. 제안한 정제단계에서는 반복적인 계산 과정이 없어 수행시간이 빨라질 뿐만 아니라 메모리 사용도 줄일 수 있는 비용함수를 사용하였다. 식 (15)는 세부분으로 구성되어 있다.

$$E(x,y,d) =$$

$$\begin{cases} w_s f_s(x,y,d_s(x,y)) + w_d f_d(x,y,d_d(x,y)) & \text{if } obj\text{-}mov = 1 \\ w_s f_s(x,y,d_s(x,y)) + w_d f_d(x,y,d_d(x,y)) + w_r f_r(x,y,d_r(x,y)) & \text{if } obj\text{-}mov = 0 \end{cases} \quad (15)$$

f_s 는 깊이영상의 자연스러운 깊이 변화를 주기 위해 깊이 카메라에서 획득한 깊이영상에서 현재 위치를 기준으로 이미 계산한 주변 화소간의 편미분값을 이용해 자연스러운 깊이영상 생성을 고려한 값이다. 그림 11은 f_s 를 계산하기 위해 주변 깊이값을 참조하는 방법을 설명하고 있다. 깊이영상은 비슷한 색상을 가지는 영역에서 대부분 자연스럽게 깊이 변화하는 특징을 가지고 있

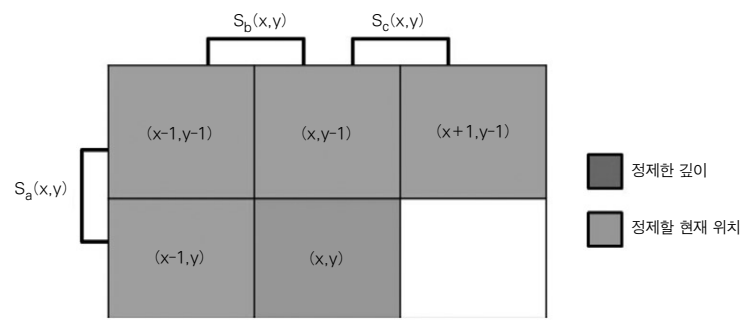
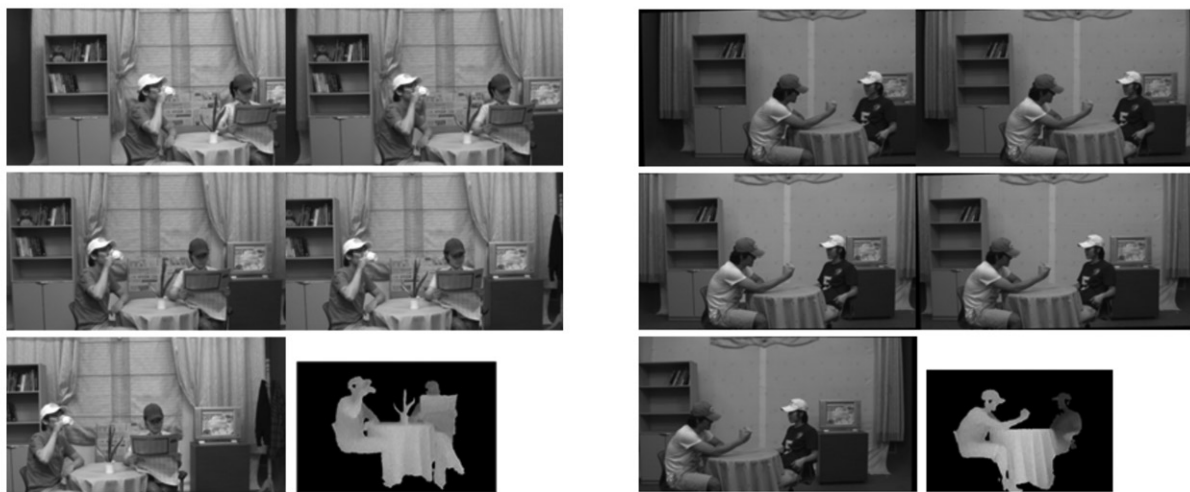


그림 11. 깊이영상 정제를 위한 비용함수 정의



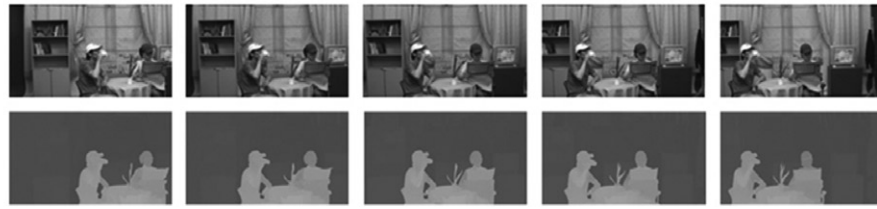
(a) Newspaper

(b) Delivery

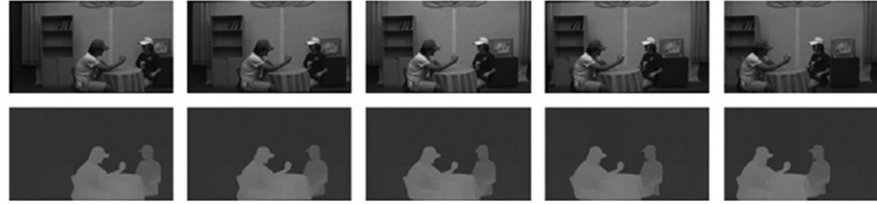


(c) cafe

그림 12. 입력영상



(a) Newspaper



(b) Delivery



(c) cafe

그림 13. 3D 비디오 생성결과

다. 따라서 이미 계산한 깊이값간의 변화되는 정도를 참조할 수 있다. 식 (16)은 그러한 주변 깊이값들의 변화율을 계산하여 선택하는 방법을 보여준다. $s_a(x,y)$, $s_b(x,y)$, $s_c(x,y)$ 는 각각 두 인접 깊이값의 차분에 대한 절대값을 계산한 것이다. 혹시 발생할 수 있는 에러를 고려하여 세값에서 중간값을 선택한다.

$$f_s(x,y, d_s(x,y)) = \text{med}(s_a(x,y), s_b(x,y), s_c(x,y)) \quad (16)$$

f_d 는 세그먼트 기반에서 한번 정제한 깊이 정보를 다시 화소 단위로 스테레오 정합을 수행하여 최소 MAD를 갖는 변이를 찾는다. 마지막으로 f_t 는 이전 프레임에서 생성한 깊이 정보로 이전 단계에서 계산한 움직임 영역결과에서 움직임없는 영역 즉 $obj-mov$ 값이 0인 위치에서만 이전 프레임의 깊이 정보를 참조한다. 이 세가지 정보를 하나의 함수로 정의하여 비용함수값이 가장 작은 값을 가지는 변이를 최종 깊이값으로 결정한다. 이 비용함수의 경우 이전 스테레오 정합의 정제함수에

서 항상 존재하는 반복적인 계산이 없기 때문에 다시점 비디오의 깊이영상을 생성하는 시간을 크게 단축시킬 수 있는 장점이 있다.

V. 실험결과 및 분석

제안한 깊이영상 생성 방법을 평가하기 위해, 5대의 다시점 카메라와 깊이 카메라를 이용하여 복합형 카메라 시스템을 구성하였다. 그림 12는 제안한 복합형 카메라 시스템을 이용해 획득한 입력영상들이다. 5개의 다시점 색상영상과 1장의 깊이영상을 획득하였다.

그림 13은 제안하는 방법을 통해 생성한 다시점 영상과 그에 상응하는 깊이영상 결과들을 보여준다. 생성한 깊이영상은 머리카락과 같이 복잡하고 미세한 외곽선 정보를 가지는 영역과 같이 비슷한 색상 정보가 반복적으로 존재하는 영역에서도 좋은 결과를 보였다. 그림 14는 제안한 방법으로 생성한 깊이영상과 기존의 알고리즘을 이용해 생성한 깊이영상 결과를 비교한 결과이다. 기존 방법은 DERS 프로그램을 이용했다

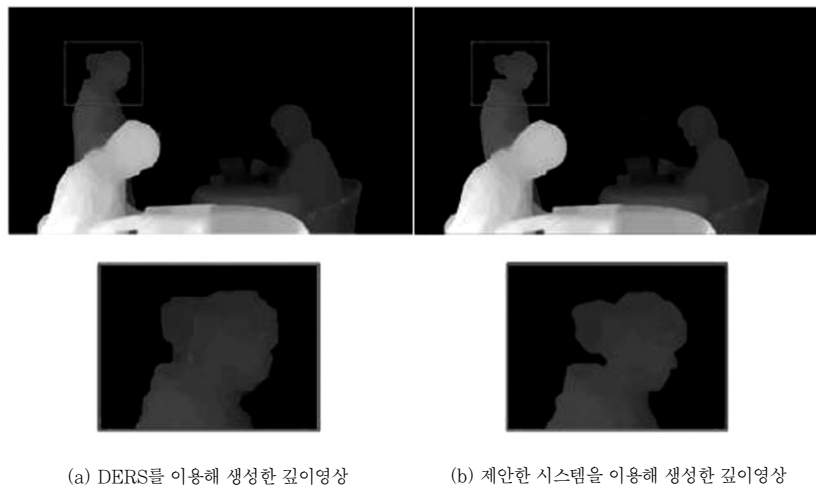


그림 14. 깊이영상 생성결과 비교

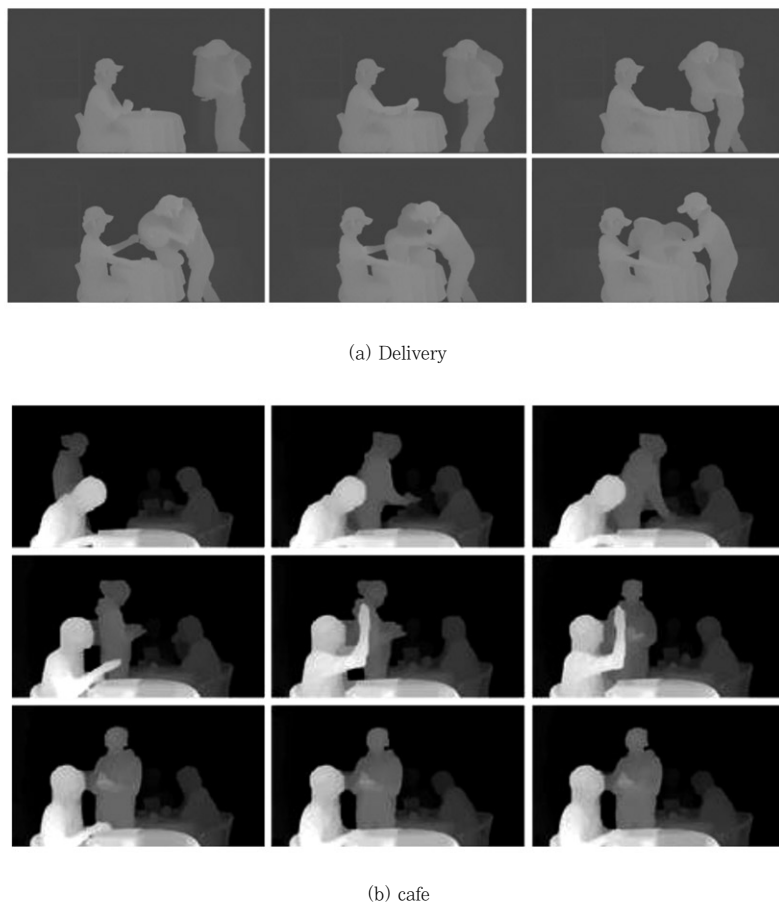


그림 15. 깊이 비디오 생성결과



그림 16. 중간영상 합성결과

표 2. PSNR 비교 (dB)

실험영상	기존 방법	제안한 방법
Newspaper	29.89	31.96
Delivery	29.91	31.70
Cafe	34.89	35.18

표 3. 계산시간 비교 (sec.)

실험영상	기존 방법	제안한 방법
Newspaper	826.17	364.90
Delivery	845.68	350.59
Cafe	836.26	337.21

[21]. 결과에서 확인할 수 있듯이 기존의 방법은 몇몇 영역에서 잘못된 정합결과를 포함하고 있으나 제안한 방법은 정확한 깊이영상을 생성했다.

표 2는 각 실험영상 별로 제안한 방법과 기존 방법을 통해 생성한 깊이영상을 이용해 중간시점 영상을 생성하고 원본 영상과의 PSNR을 구한 결과이다. 결과에서도 확인할 수 있듯이 제안한 방법으로 생성한 중간영상이 더 정확하게 생성할 수 있음을 확인했다. 표 3은 깊이영상을 생성하는데 걸리는 시간을 각각 비교한 것이다. 제안한 방법은 초기 깊이값을 이용해 제한된 영역내에서 정합을 수행해서 계산시간을 크게 감소시킬 수 있었다. 그림 15는 카메라 3에서의 생성한 깊이영상을 30프레임마다 보여준 결과이다.

그림 16은 카메라 2와 카메라 4의 색상영상과 그에 상응하는 깊이영상을 이용해 중간시점 영상을 생성한 결과이다. 그림 16에 보인 것처럼 깊이영상을 이용해 생성한 중간영상은 자연스러운 3차원 장면영상을 생성했다.

VI. 결 론

본 논문은 복합형 카메라 시스템을 이용해 고화질 깊이영상을 생성하는 방법을 제안했다. 기존의 깊이영상 생성 방법이 시점 변화에 의해 정합시 다른 시점에서 보이지 않아 정확한 계산이 어려운 영역과 비슷한 색상 정보를 가지는 영역에 대해 잘못된 깊이 정보를 생성하는 문제점을 해결하였다. 영상내에 움직이는 물체를 구분하여 깊이영상 경계영역에서도 기존 방법과 비교했을 때 정확한 깊이영상을 생성할 수 있었다. 또한 깊이영상 정제를 위해 획득한 깊이영상, 스테레오 정합 그리고 이전 프레임의 깊이 정보를 이용해 새로운 비용함수를 정의하였고 이를 이용해 고화질의 다시점 깊이영상을 생성할 수 있었다. 이를 통해 실험결과에서도 확인할 수 있듯이 기존의 깊이영상 생성 방법보다 향상된 화질의 깊이영상을 좀더 빠른 시간안에 얻었다. 최종적으로 복합형 카메라 시스템과 제안한 깊이 정제함수를 이용해 고해상도 3차원 비디오를 생성하는 방법을 제시했다

[참고문헌]

- [1] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic, and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing Image Communication*, Vol. 22, No. 2, Feb. 2007, pp. 217-234.
- [2] C. Fehn, R. Barre, and S. Pastoor, "Interactive 3-DTV-concepts and key technologies," *Proceedings of the IEEE*, Vol. 94, No. 3, Mar. 2006, pp. 524-538.
- [3] ISO/IEC JTC1/SC29/WG11 N8944, "Preliminary FTV model and requirements," Apr. 2007.
- [4] A. Smolic and D. McCutchen, "3DAV exploration of video-based rendering technology in MPEG," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, No. 3, Mar. 2004, pp. 348-356.
- [5] S. Nayar and Y. Nakagawa, "Shape from Focus," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 16, No. 8, Aug. 1994, pp. 824-831.
- [6] C. Zitnick, S. Kang, M. Uyttendaele, S. Winder, and R. Szeliski, "High-quality video view interpolation using a layered representation," *Proc. of ACM SIGGRAPH*, Aug. 2004, pp. 600-608.
- [7] D. Scharstein, and R. Szeliski, "High-accuracy stereo depth maps using structured light," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 1, Jun. 2003, pp. 195-202.
- [8] G. Iddan and G. Yahav, "3D imaging in the studio and elsewhere," *Proc. of SPIE Vidometrics and Optical Methods for 3D Shape Measurements*, Jan. 2001, pp. 48-55.
- [9] M. Kawakita, T. Kurita, H. Kikuchi, and S. Inoue, "HDTV axi-vision camera," *Proc. of International Broadcasting Conference*, Sep. 2002, pp. 397-404.
- [10] S. Gokturk, H. Yalcin, C. Bamji, "A time-of-flight depth sensor - system description, issues and solutions," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshop*, Vol. 3, Jun. 2004, pp. 35.
- [11] SwissRanger™ SR4000. <http://www.mesa-imaging.ch/>.
- [12] G. Um, K. Kim, C. Ahn, and K. Lee, "Three-dimensional scene reconstruction using multi-view images and depth camera," *Proc. of SPIE Stereoscopic Displays and Virtual Reality Systems XII*, Vol. 5664, Jan. 2005, pp. 271-280.
- [13] J. Diebel and S. Thrun, "An application of Markov random fields to range sensing," *Proc. of Advances in Neural Information Processing systems*, pp. 291-298, Dec. 2005.
- [14] J. Zhu, L. Wang, R. Yang, and J. Davis, "Fusion of time-of-flight depth and stereo for high accuracy depth maps," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2008, pp. 231-236.
- [15] E. Lee, Y. Kang, Y. Jung, Y. Ho, "3-D video generation using hybrid camera system," *Proc. of International Conference on Immersive Telecommunications (IMMERSCOM)*, Jun. 2009, pp. T5(1-6).
- [16] Camera calibration toolbox program for Matlab provided by Caltech, http://www.vision.caltech.edu/bouguetj/calib_doc/
- [17] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 22, Nov. 2000, pp. 1330-1334.
- [18] Y. Kang, and Y. Ho, "Geometrical compensation for multi-view video in multiple camera array," *Proc. of International Symposium ELMAR*, Vol. 1, Sep. 2008, pp. 83-86.
- [19] N. Joshi, B. Wilburn, V. Vaish, M. Levoy, and M. Horowitz, "Automatic color calibration for large camera arrays," in *UCSD CSE Technical Report*, May 2005, CS2005-0821.
- [20] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 4, May 2002, pp. 603-619.
- [21] ISO/IEC JTC1/SC29/WG11 M15377, "Reference softwares for depth estimation and view synthesis," Apr. 2008.



이 은 경
(Eun-Kyung Lee)

2004. 2: 호남대학교 컴퓨터공학과 학사
 2006. 2: 호남대학교 컴퓨터공학과 석사
 2011. 8: 광주과학기술원 정보통신공학부 박사
 관심분야: 다시점 깊이영상 획득 및 처리,
 3차원 비디오 생성, 3차원 TV, 실감방송
 E-mail: eklee78@gist.ac.kr
 Tel: +82-62-715-2258
 Fax: +82-62-715-3164



호 요 성
(Yo-Sung Ho)

1981. 2: 서울대학교 전자공학과 학사
 1983. 2: 서울대학교 전자공학과 석사
 1989. 12: Univ. of California, Santa Barbara,
 Department of Electrical and
 Computer Engineering, 박사
 1983. 3~1995. 9: 한국전자통신연구소 선임연구원
 1990. 1~1993. 5: 미국 Philips 연구소,
 Senior Research Member
 1995. 9~현재: 광주과학기술원 정보통신공학부 교수
 관심분야: 디지털 신호처리, 영상신호 처리 및 압축,
 멀티미디어 시스템, 디지털 TV와 고선명 TV,
 MPEG 표준, 3차원 TV, 실감방송
 E-mail: hoyo@gist.ac.kr
 Tel: +82-62-715-2211
 Fax: +82-62-715-3164