

Gaussian 가중 최소 자승법을 이용한 깊이 영상의 시간적 상관도 향상 방법

Temporally Consistent Depth Video Filtering Using Gaussian-weighted Least Squares

이상범 · 호요성
Sang-Beom Lee · Yo-Sung Ho

다시점 영상을 이용해서 깊이 영상을 획득하기 위한 방법으로는 스테레오 정합 방법이 일반적으로 가장 많이 사용되고 있다. 이 방법은 비교적 높은 정확도의 깊이 영상을 제공할 수 있지만 매 화면마다 독립적으로 깊이 영상을 탐색하기 때문에 깊이 영상의 시간적 상관도가 낮은 단점이 있다. 시간적 상관도 저하 문제는 깊이 영상을 이용하여 가상 시점의 영상을 합성했을 때 장면이 떨리게 되어 시청자에게 시각적인 불편함을 유발한다. 또한 깊이 영상 자체의 부호화 과정에서 시간적 예측 효과를 저하시키기 때문에 깊이 영상의 부호화 효율을 저하시킨다. 본 논문에서는 깊이 영상의 시간적 상관도를 향상시키는 방법을 제안한다. 제안하는 방법은 깊이 영상에 변형된 결합형 다방향 필터(modified joint multilateral filter)를 적용하는데 이 필터는 하나의 공간 필터와 밝기 및 깊이값의 유사도에 대한 두 개의 범위 필터로 구성된다. 또한 제안한 방법은 시간축에서 움직이는 객체에 대한 필터링을 위해 Gaussian 가중 최소 자승법을 사용해서 외곽오차를 제거한다. 실험 결과를 통해 제안하는 방법이 깊이 영상의 시간적 상관도를 향상시켰을 뿐만 아니라 화질도 약 1.22dB 향상시켰음을 확인할 수 있었다.

주제어: Depth video filtering, Modified joint multilateral filter, Gaussian-weighted least squares, Temporal consistency, Three-dimensional video

In general, a stereo matching algorithm is widely used to estimate the depth map by the aid of the multi-view video. Although this method provides us with reliable quality of depth maps, it has a problem that the temporal consistency of the depth video is low since the estimation process is independently performed on a frame-basis. The temporal inconsistency problem affects the rendering quality of the synthesized views, as the flickering artifact causes viewer discomfort. In addition, performance degradation of temporal prediction transpires in depth coding. In this paper, we propose a depth filtering algorithm for temporally consistent depth video. The proposed method applies a Modified Joint Multilateral Filter (MJMF) to depth video. Fundamentally, MJMF is based on three Gaussian distributions: one spatial filter and two range filters for color and depth similarities. In addition, we perform outlier reduction in the temporal domain using Gaussian-weighted least squares to deal with temporal motion. Experimental results show that the proposed method improves the depth quality by 1.22 dB on average compared to previous algorithms due to enhanced temporal consistency.

Keywords: Depth video filtering, Modified joint multilateral filter, Gaussian-weighted least squares, Temporal consistency, Three-dimensional video

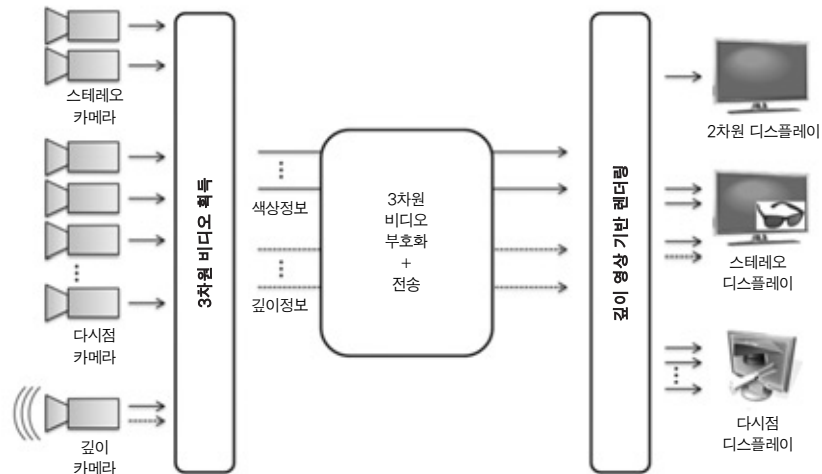


그림 1. 3차원 비디오 시스템

I. 서론

색상 영상과 깊이 영상으로 표현되는 3차원 비디오는 차세대 영상 미디어로 각광받고 있다[1]. 스테레오 디스플레이 혹은 다시점 디스플레이와 같은 3차원 디스플레이의 발전으로 인해 3차원 TV는 사용자에게 보다 실감나는 콘텐츠를 제공할 수 있게 되었다. 그 예로 영화 "아바타"의 흥행으로 인해 3차원 영상에 대한 관심이 폭발적으로 증가했고 이에 따른 공급 효과로 다양한 장르의 3차원 영화가 극장에서 상영되기 시작했으며 심지어 3차원 TV가 가정까지 보급되기 시작했다[2]. 멀지 않은 미래에 3차원 디지털 미디어가 영상 산업의 또다른 혁신을 주도할 것으로 기대를 모으고 있다.

그림 1은 3차원 비디오 시스템의 개념도를 나타낸 것이다. 3차원 비디오 시스템은 다시점 색상 영상과 이에 상응하는 깊이 영상을 포함하는 3차원 비디오의 획득/처리/전송/재생에 대한 모든 과정을 포함하고 있다. 3차원 비디오는 스테레오 카메라, 깊이 카메라, 다시점 카메라 등 다양한 저작도구들을 이용하여 직접 얻을 수 있으며 특히 깊이 카메라를 이용할 경우 카메라에 부착된 레이저, 구조광 등의 능동센서를 통해 깊이 영상을 직접 획득할 수 있다. 하지만 스테레오 카메라 혹은 다시점 카메라의 경우나 기존에 제작된 2차원 영상의 경우 깊이 영상을 직접적으로 획득할 수 없기 때문에 촬영된 장면으로부터 3차원 정보를 계산해서 깊이 영상을 획득해야 한다. 스테레오 정합(stereo matching) 기술이 그 대표적인 예이다.

단일 시점의 색상 영상과 깊이 영상을 이용하면 가상의 시점을 합성할 수 있다. 이와 같은 기술을 깊이 영상 기반 렌더링이라고 한다[3],[4]. 깊이 영상 기반 렌더링 기법(DIBR: Depth Image-Based Rendering)은 3차원 장면을 효과적으로 재현할 수 있기 때문에 3

차원 영상 산업의 주요 기술 가운데 하나로 사용된다. 깊이 영상은 3차원 장면에서 객체와 카메라와의 거리를 나타내는 정보로 활용되고 가상의 시점 영상은 다음과 같은 단계로 진행된다. 색상 영상의 모든 화소들은 카메라 정보와 깊이 영상을 통해 세계 좌표계(world coordinate)로 역 사상(back-projection)되고 다시 가상 시점에서의 영상 평면으로 사상된다. 이러한 과정은 컴퓨터 그래픽스 분야에서 3차원 워핑으로 불린다[5].

깊이 영상의 정확도는 DIBR 과정에서 가상시점의 합성화질에 직접적인 영향을 미치기 때문에 정확한 깊이 영상을 획득하는 것은 매우 중요하다. 컴퓨터 비전, 영상처리 분야에서 진행된 다양한 연구들이 정확한 깊이 영상을 획득하는 것에 초점을 맞추고 있지만 여전히 많은 문제점을 가지고 있다.

부정확한 깊이 영상의 문제점 가운데 시간적 상관도 저하 문제는 깊이 영상의 시간상에서의 상관도를 고려하지 않은 화면간 독립적인 획득 과정 때문에 발생한다. 그 결과로 객체 경계 주변에서의 깊이값이 심하게 변하게 되고(depth fluctuation) 이는 합성된 영상에도 그대로 전파되어 시청자의 눈에 피로감을 주게 된다. 또한 깊이 영상 자체의 부호화 과정에서 시간적 예측 효과를 저하시키기 때문에 깊이 영상의 부호화 효율을 저하시킨다.

본 논문에서는 단일 시점에서 색상 영상의 정보를 활용해서 깊이 영상의 화질을 개선하는 깊이 영상 필터링 방법을 제안한다. 제안하는 방법은 크게 두가지 단계로 이루어지는데 하나는 결합형 다방향 필터(joint multilateral filter)이고 다른 하나는 Gaussian 가중 최소 자승법을 이용한 시간적 외곽오차 제거 방법이다. 외곽오차 제거 방법은 움직이는 객체 혹은 움직이는 카메라 환경에서 필터링 방법을 적용하기 위한 기술이다.

본 논문의 구성은 다음과 같다. II장에서는 깊이 영

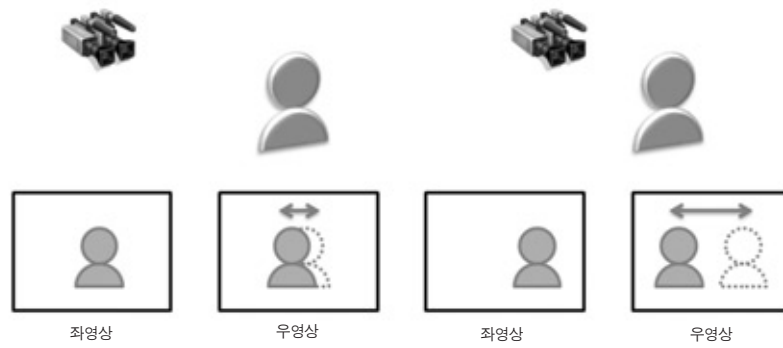


그림 2. 변위 계산의 기본 원리

상의 시간적 상관도 향상을 위한 기존의 방법들을 소개하고 III장에서는 제안하는 깊이 영상 필터링 방법을 설명한다. IV장에서는 제안하는 방법의 성능을 평가하기 위한 실험 결과를 분석하고 V장에서 결론을 맺는다.

II. 깊이 영상 탐색 및 필터링 기술

1. 스테레오 정합 기술

스테레오 카메라를 이용한 변위 계산 방법에서 변위와 깊이의 상관관계는 다음과 같은 방법으로 유도할 수 있다. 스테레오 카메라는 정확하게 수평이 맞도록 정렬이 되어 있고 또한 카메라의 광축이 정확하게 평행을 이룬다고 가정한다. 이러한 카메라 구성에서 멀리 있는 물체는 좌영상과 우영상에서 시차가 작게 발생하는 반면 가까이 있는 물체는 시차가 크게 나타난다. 이러한 양안시차 원리를 이용해서 기준 시점의 모든 화소가 참조시점의 어느 위치에 존재하는지 탐색하면 변위(disparity)를 얻을 수 있고 변위를 이용하면 실제 깊이 정보를 계산할 수 있다. 그림 2는 변위 계산의 기본 원리를 도식화한 것이다.

인간 시각 체계(human visual system)는 양쪽 눈으로부터 얻은 영상의 시차로 물체의 깊이 정보를 인지하는 것으로 알려져 있다. 스테레오 정합(stereo matching) 기술은 이러한 인간 시각체계를 영상처리에 차용한 예로써 현재까지도 컴퓨터 비전 분야에서 활발히 연구 중이다[6]. 스테레오 영상에서 객체의 변위값을 구하기 위해서는 두 영상에 모두 존재하는 동일한 객체의 화소 위치를 찾아야 한다. 이러한 과정을 대응점 탐색이라고 하며 좌영상에 존재하는 한 물체의 대응점을 찾기 위해 우영상에서 수평의 화소들에 대하여 실제 대응되는 점을 찾게 된다.

변위 계산을 위해서는 정합 함수(matching function)를 이용하여 정합비용을 계산하며 정합비용이 가장 작은 값을 가질 때의 변위를 취하도록 한다. 정합

함수는 두가지 가정을 포함하고 있는데 첫번째는 두 시점에서 변위만큼 떨어진 화소는 서로 유사해야 한다는 것이며 두번째는 주변의 화소들은 비슷한 변위를 가져야 한다는 것이다. 이 두가지 가정은 정합 함수에서 데이터항(data term)과 평활화항(smoothness term)으로 표현된다. 정합 함수는 다음과 같이 정의된다.

$$D(x,y) = \underset{d}{\operatorname{argmin}} \{ E(x,y,d) \} \quad (1)$$

$$E(x,y,d) = E_{data}(x,y,d) + E_{smooth}(x,y,d) \quad (2)$$

여기서 $D(x,y)$ 는 화소 (x,y) 에서의 변위를 말하고 정합 함수 $E(x,y,d)$ 는 데이터항 $E_{data}(x,y,d)$ 와 평활화 항 $E_{smooth}(x,y,d)$ 로 구성된다.

스테레오 정합 방법은 변위 계산 방법에 따라 크게 지역적인 것과 전역적인 방법으로 나눌 수 있다. 일반적으로 지역적 방법은 계산 속도가 빠르지만 화소별로 각각 대응점을 찾기 때문에 정보가 모호한 영역에서는 성능이 저하되는 단점이 있다.

전역적 방법은 주변의 정보를 같이 활용하여 영상 전체의 오차를 최소화하기 때문에 보다 정확한 결과를 얻을 수 있지만 계산시간은 더 많이 소요된다. 대표적인 전역적 방법으로는 마르코프 랜덤 필드(MRF: Markov Random Field) 모형을 들 수 있는데[7] 이 방법은 입력 영상을 이용하여 필드를 구성하고 최소 에너지를 구하여 변위를 계산한다. 따라서 필드를 구성하는 방법과 최소 에너지 함수를 정의하는 것이 알고리즘의 성능을 좌우한다.

최근에는 전역적 방법이 지역적 방법에 비해 비교적 높은 성능을 보이면서 많은 연구들이 이 방법을 따르고 있다[8]~[11]. 그림 3은 스테레오 정합 기술을 통해 획득한 변위 영상을 나타내고 있다. 그림 3에서 얻은



그림 3. 스테레오 정합 기술로 얻은 변위 영상

변위 영상은 전역적 방법 중 하나인 신뢰 전파 (belief propagation) 기술을 이용해서 얻은 결과이다[10]. 그림 3(b)에서 알 수 있듯이 스테레오 정합 기술을 이용하면 객체의 깊이정보를 효과적으로 얻을 수 있다.

2. 깊이 영상 필터링 기술

스테레오 정합 기술은 특별한 센서 없이도 다수의 카메라만으로 깊이 영상을 획득할 수 있다는 장점을 가지지만 폐색(occlusion) 영역, 반복되는 텍스처 혹은 텍스처가 없는 영역, 빛의 반사가 서로 다른 시점에서 동일하지 않은 평면, 시간적 상관도 저하 문제 등 다양한 문제점을 가지고 있다. 이러한 문제들을 해결하기 위해 지금까지도 다양한 방법들이 제안되고 있다. 이 가운데 깊이 영상의 시간적 상관도 향상을 위한 기존의 기술들을 자세히 살펴보면 동적 객체를 고려한 깊이 탐색 기술, 깊이 영상 필터링 기술로 분류할 수 있다.

동적 객체를 고려한 깊이 탐색 기술은 깊이 탐색 과정에서 사용하는 정합 함수를 시간상으로 확장한다. *Tao et al.*은 3차원 장면이 다양한 조각의 평면들로 구성되어 있다는 가정 하에 영상을 분할한 다음 분할된 영역 각각에 대해서 깊이값을 구하는 방법을 제안했다[12]. 이 방법에서 정합 함수를 이전 화면과 이후 화면을 고려하도록 설계했기 때문에 깊이 영상의 변동 현상이 줄어들었다. *Tao et al.*은 시간상으로 확장된 신뢰 전파 기술을 사용함으로써 이 문제를 해결했다[13].

깊이 영상 필터링 기술은 색상값, 깊이값을 사용해서 저역 통과 필터를 설계함으로써 이 문제를 해결한다. 가장 대표적인 방법으로는 결합형 양방향 필터가 있다 (JBF: Joint Bilateral Filter)[14]. JBF는 표적 화소와 그 주변의 화소들 사이의 좌표상 거리로부터 유도되는 공간 필터(spatial filter)와 밝기 유사도로부터 유도되는 범위 필터(range filter)를 사용한다. 각각의 필터는 표적 화소와 주변 화소들 사이의 거리가 가까울수록 밝기 유사도가 높을수록 가중치를 증가시킨다.

깊이 영상에서 표적 화소와 주변 화소 중 하나가 각각 좌표 p 와 q 에 위치해 있고 S_p 와 S_q 는 각 화소에 대응

하는 깊이값, I_p 와 I_q 는 각 화소에 대응하는 밝기값이라고 가정하자. 그렇다면 JBF를 통해 계산된 새로운 깊이값 \tilde{S}_p 는 다음과 같이 나타낼 수 있다.

$$\tilde{S}_p = \frac{\sum_{q \in \Omega} S_q \cdot f(\|p-q\|)g(\|I_p-I_q\|)}{\sum_{q \in \Omega} f(\|p-q\|)g(\|I_p-I_q\|)} \quad (3)$$

여기서 f 와 g 는 각각 공간 필터, 범위 필터를 나타낸다. Ω 는 필터의 커널 크기를 나타내며 화소 좌표 p 와 q 는 모두 Ω 의 범위 내에 존재한다. f 와 g 가 Gaussian 분포를 사용한다면 다음과 같이 나타낼 수 있다.

$$f(x) = \exp\left(-\frac{x^2}{2\sigma_f^2}\right), g(x) = \exp\left(-\frac{x^2}{2\sigma_g^2}\right) \quad (4)$$

여기서 σ_f^2 과 σ_g^2 은 각각 공간 필터, 범위 필터의 표준편차를 나타낸다.

최근 JBF를 사용한 다양한 방법이 제안됐다. *Lai et al.*은 기존의 JBF에 깊이값 임계 필터를 추가한 반복적 결합형 다방향 필터(JMF: iterative Joint Multilateral Filter)를 제안했다[15]. *Yang et al.*은 변형된 JBF(MJBF: Modified JBF)를 소개했는데[16] 이 방법은 표적 화소 상하좌우의 네 화소를 깊이값 후보로 두고 필터 함수를 최소 인수(argument of minimum)로 수정해서 깊이값 후보 중 필터 함수값이 최소가 될 때의 후보를 선택한다.

이러한 기존의 필터링 기술들은 깊이 영상의 화질을 향상시키는데에는 성공했지만 시간적 상관도 문제는 여전히 해결되지 않았다. 최근에 3D JBF 기반의 깊이 영상 필터링 방법이 제안됐는데[17] 이 방법은 깊이 영상의 변동 현상을 막기 위해 필터를 시간축으로 확장했다. 또한 색상 및 깊이를 고려한 범위 필터가 필터 커널

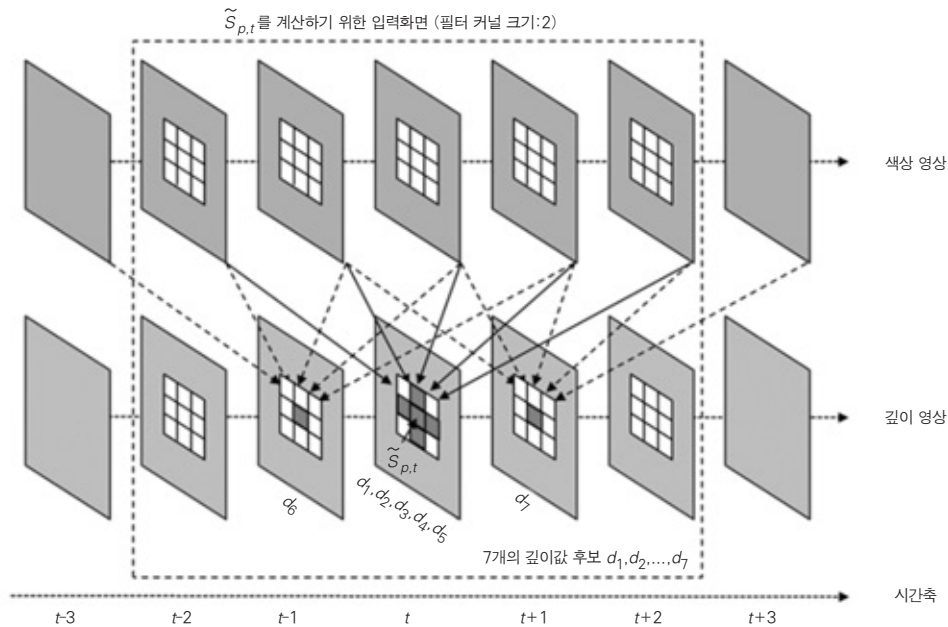


그림 4. 제안한 방법의 필터링 과정

내부의 깊이값 분포에 따라 적응적으로 적용됐다. 하지만 객체 및 카메라의 움직임을 고려하지 않은 이 기술은 움직이는 객체가 존재할 때 동작 흐림(motion blur) 현상을 유발한다.

III. 제안하는 방법

본 장에서는 깊이 영상의 시간적 상관도를 위해 MJBF 기반의 깊이 영상 필터링 기술을 제안한다. 기존의 방법들과 다르게 제안하는 방법은 객체 및 카메라의 움직임을 고려한 필터를 설계한다.

1. 변형된 결합형 다방향 필터

다양한 3차원 비디오 응용에서의 활용을 위해 깊이 영상의 정확도는 매우 중요하다. 앞서 언급한 MJBF는 객체의 경계 주변에서 깊이 오차를 효과적으로 줄이지만 화면간 깊이값이 여전히 변동한다. 따라서 제안한 방법은 기존의 MJBF를 기반으로 필터를 시간상으로 확장한다. 색상 영상에서의 불연속점은 깊이 영상에서의 불연속점을 바로잡을 수 있다는 가정 하에 제안하는 변형된 결합형 다방향 필터(MJMF)는 하나의 공간 필터, 화소 및 깊이값의 유사도로부터 유도되는 두 범위 필터를 사용하며 모두 Gaussian 분포로부터 모델링된다. 깊이 영상에서 표적 화소가 공간 좌표 p , 시간 좌표 t 에 존재하고 이 화소에 대해 필터링이 적용됐을 때 필

터링된 깊이값 $\tilde{S}_{p,t}$ 는 다음과 같이 나타낼 수 있다.

$$\tilde{S}_{p,t} = \underset{d \in N_d}{\operatorname{argmin}} \frac{\sum_{q \in \Omega} \sum_{n \in \Pi} W_{q,n} \cdot C_{q,n,d}}{\sum_{q \in \Omega} \sum_{n \in \Pi} W_{q,n}} \quad (5)$$

여기서 t 와 n 은 표적 화소와 주변 화소의 시간 좌표를 나타내며 Π 는 시간축에서의 필터 커널 크기를 나타낸다. N_d 는 깊이값의 후보인 d 의 집합을 의미하며 표적 화소와 동일한 화면에서의 상하좌우 네개의 주변 화소, 즉 $S_{q,t}$ 와 동일한 공간 좌표에서 이전 화면과 이후 화면의 두개의 주변 화소, 즉 $S_{p,t-1}$, $S_{p,t+1}$ 을 포함한다. 동작 흐림 현상없이 깨끗한 객체 경계를 얻기 위해서 필터링된 깊이값은 식 (5)에서와 같이 주변 6개의 화소가운데 하나로 선택된다. $W_{q,n}$ 은 다음과 같이 정의된다.

$$W_{q,n} = f(\|p_t - q_n\|) g_I(\|I_{p,t} - I_{q,n}\|) g_S(\|S_{p,t} - S_{q,n}\|) \quad (6)$$

여기서 p_t 와 q_n 은 각각 표적 화소와 주변 화소의 공간 및 시간 좌표를 나타내고, $I_{p,t}$ 와 $I_{q,n}$ 은 각각 표적 화소와 주변 화소의 밝기값, $S_{p,t}$ 와 $S_{q,n}$ 은 각각 표적 화소와 주변 화소의 깊이값을 나타낸다. 그리고 g_I 와 g_S 는 각각 밝기 및 깊이 유사도에 대한 범위 필터를 나타낸다. 깊이 유사



그림 5. 단순 확장한 MJMF의 필터링 오차

도에 대한 범위 필터는 다음 절에서 자세하게 설명한다. 절삭 선형 모델(truncated linear model)을 사용한 깊이값의 비용 함수 $C_{q,n,d}$ 는 다음과 같다.

$$C_{q,n,d} = \min(\lambda L, \|S_{q,n} - d\|) \quad (7)$$

여기서 λ 는 외곽오차를 제거하기 위한 상수이다. L 은 깊이값의 범위를 나타내는데 통상적으로 깊이 영상은 8비트의 단일 채널 영상으로 표현하기 때문에 L 값은 255으로 고정된다. 그림 4는 시간상 필터 커널 기가 2일 때 제안한 방법의 필터링 과정을 도식화한 것이다.

제안한 방법은 이미 획득된 깊이 영상에 대한 필터링을 목적으로 하고 있기 때문에 미래의 화면을 사용할 수 있다. 또한 현재 화면을 필터링 하는 과정에서 필터링된 이전 화면을 사용하지 않고 입력 영상을 그대로 사용하는데 그 이유는 필터링 과정에서 발생 가능한 오차가 이후 화면으로 전파될 가능성이 있기 때문이다.

2. 시간상 외곽오차 제거 방법

MJMF를 단순히 시간상으로 확장만 하는 것은 여전히 동작 흐림 현상이나 시간상 외곽 오차와 같은 문제를 해결하지 못한다. 시간상 외곽오차란 시간상 필터 커널 내부의 모든 화면에 대한 동일 위치에서의 화소값 가운데 다른 화소들과의 차이가 큰 화소를 말한다. 시간상 외곽오차를 포함한 채로 필터링을 수행하면 깊이값이 왜곡된다. 그림 5는 시간상으로 확장한 MJMF의 필터링 오차를 나타낸다. 그림 5(b)는 그래프 컷 기반의 스테레오 정합 방법을 사용하는 Depth Estimation Reference Software(DERS)를 통해 획득한 깊이 영상

이며[18] 그림 5(c)는 그림 5(b)에 필터링을 적용한 결과를 보여준다. 그림에서 알 수 있듯이 시간상 외곽오차로 인해 객체 내부의 깊이값이 필터링 과정에서 망가진 것을 확인할 수 있다.

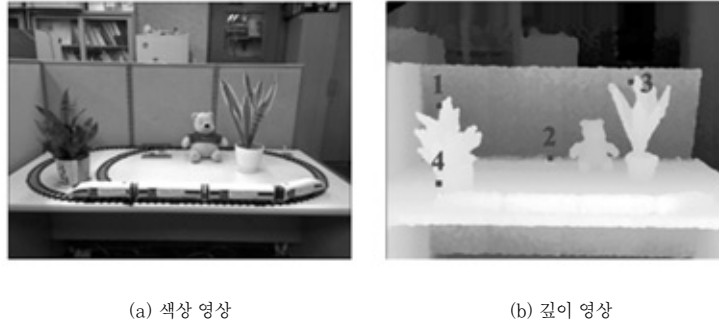
시간상 외곽오차를 제거하기 위해 제안한 방법은 첫 번째로 깊이 유사도로부터 유도된 범위 필터를 추가한다. 이 과정을 통해 큰 차이가 나는 외곽오차를 제거할 수 있다. 단순한 임계 필터를 사용한 기존의 JMF 방법과 다르게 제안한 방법은 밝기 유사도에서와 같이 Gaussian 분포를 사용하며 다음과 같이 정의된다.

$$g_S(\|S_{p,t} - S_{q,n}\|) = \exp\left(-\frac{\|S_{p,t} - S_{q,n}\|^2}{2\sigma_S^2}\right) \quad (8)$$

여기서 σ_S 는 깊이값의 표준편차를 나타낸다.

외곽오차 제거를 위한 추가적인 방법으로 제안한 방법은 Gaussian 가중 최소 자승법(GWLS: Gaussian-Weighted Least Squares)을 사용한다. 이를 뒷받침할 근거로 두가지의 가정을 세워보면 첫번째는 밝기(혹은 깊이)값은 동일한 좌표에서 시간상으로 움직임이 없는 영역에 대해 선형적으로 변화한다는 것이다. 그렇기 때문에 움직임이 없는 영역에서 동일한 좌표의 모든 화면에서의 밝기(혹은 깊이)값은 선형 모델로 예측할 수 있다. 두번째는 선형 모델 예측은 밝기(혹은 깊이) 차이와 시간상 거리에 의존적이라는 점이다. 따라서 제안한 방법에서 GWLS를 사용하면 움직임의 유무에 따라 외곽오차를 판별할 수 있게 된다.

GWLS 방법은 다음과 같이 계산된다. 다음 식(9)와 같이 1차 다항식으로 모델링될 수 있는 밝기(혹은 깊이)값이 있다고 가정한다.



(a) 색상 영상

(b) 깊이 영상

그림 6. 깊이 카메라로부터 획득한 영상

$$y = b_1 x + b_2 \quad (9)$$

여기서 x 와 y 는 각각 필터 커널 Π 안에서 시간 좌표와 밝기(혹은 깊이)값을 나타낸다. 위 식에 대해서 미지의 변수 b_1 과 b_2 를 풀기 위해서는 SSE (Sum of Square Error)로 정의되는 시스템을 식 (10)과 같이 정의한다.

$$SSE = \sum_{n \in \Pi} w_n \{y_n - (b_1 x_n + b_2)\}^2 \quad (10)$$

식 (10)으로부터 b_1 은 식 (11)과 같이 얻을 수 있다.

$$b_1 = \frac{\sum_{n \in \Pi} w_n \sum_{n \in \Pi} w_n x_n y_n - \sum_{n \in \Pi} w_n x_n \sum_{n \in \Pi} w_n y_n}{\sum_{n \in \Pi} w_n \sum_{n \in \Pi} w_n x_n^2 - \left\{ \sum_{n \in \Pi} w_n x_n \right\}^2} \quad (11)$$

b_1 을 이용해서 b_2 를 풀면 식 (12)와 같다.

$$b_2 = \frac{\sum_{n \in \Pi} w_n y_n - b_1 \sum_{n \in \Pi} w_n x_n}{\sum_{n \in \Pi} w_n} \quad (12)$$

여기서 Gaussian 가중값은 식 (13)과 같이 계산된다.

$$w_n = \exp \left\{ -\frac{\|x_n - x_t\|^2}{2\sigma_x^2} \right\} \exp \left\{ -\frac{\|y_n - y_t\|^2}{2\sigma_y^2} \right\} \quad (13)$$

여기서 x_t 와 y_t 는 시간 t 에서의 시간 좌표 및 밝기(혹은 깊이)값을 나타낸다.

위 과정을 통해 우리는 선형 모델의 계수를 얻을 수 있다. 외곽오차를 판별하기 위해서 제안한 방법은 예측된 선형 모델과 각 화면에서의 화소와의 거리를 계산한다. 거리가 임계값보다 클 경우에는 해당 화소가 외곽오차로 간주된다. 여기서 사용한 임계값 T_{GWLS} 는 식 (14)와 같이 정의된다.

$$T_{GWLS} = \frac{|b_1 x_t - y_t + b_2|}{\sqrt{b_1^2 + 1}} \quad (14)$$

GWLS를 사용한 외곽오차 제거 방법은 색상 영상과 깊이 영상 각각에 대해 동일하게 수행된다. 색상 영상과 깊이 영상에서의 외곽오차를 포함하는 화면은 필터링 과정에서 제외되고 나머지 화면들만 필터링 과정에서 사용된다.

IV. 실험 결과 및 분석

제안한 방법의 성능을 평가하기 위해 깊이 영상 획득을 위한 능동센서 방식, 수동센서 방식의 두가지 접근 방식에 대해 실험을 진행했다. 비교 분석을 위해서 기존의 방법 중 MJBF, JMF, 3D JBF를 사용했다. 제안한 필터링 방법에서 표준편차는 밝기에 대해서는 10, 깊이에 대해서는 255값을 사용했으며 식 (7)에서 외곽오차 제거를 위한 상수 λ 는 0.1로 설정했다.

1. 깊이 카메라를 이용한 실험 결과

능동센서 방식의 경우 제안한 방법은 구조광 깊이 카메라를 사용해서 획득한 해상도 640×480 의 깊이 영

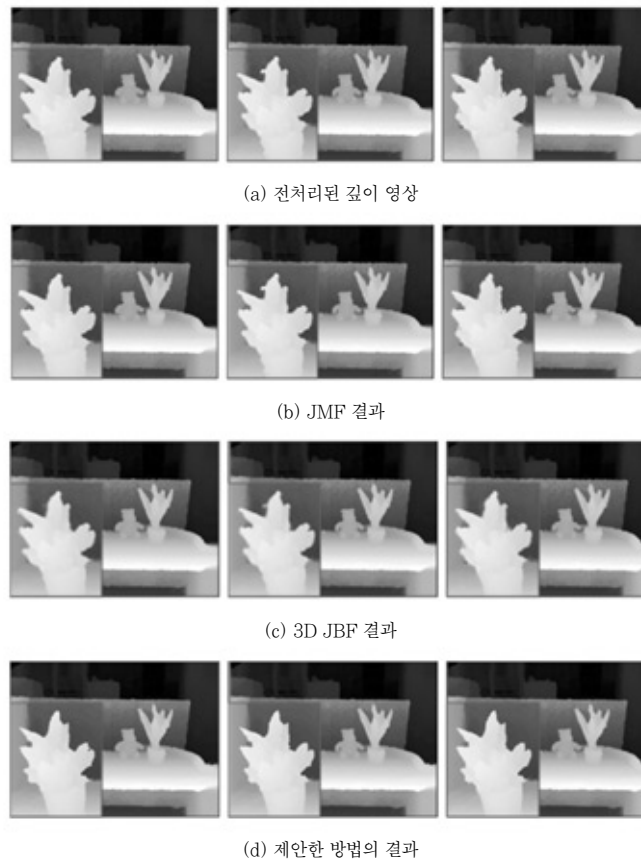


그림 7. 깊이 카메라로부터 획득한 깊이 영상의 필터링 결과

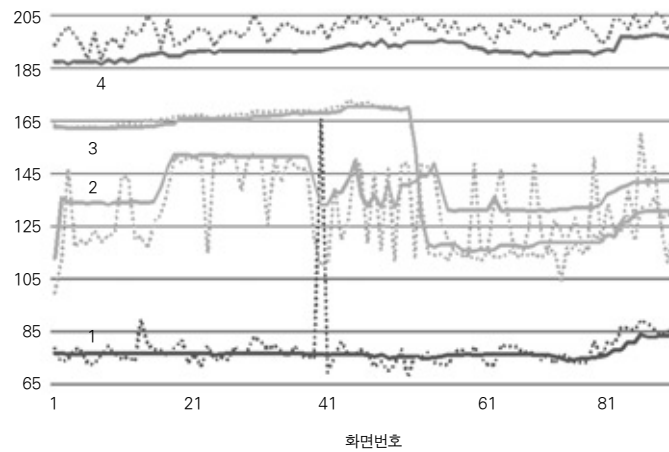


그림 8. 네 영역의 평균 깊이값

상에 대해 깊이 영상의 화질을 비교했다. 가공되지 않은 깊이 영상은 적외선 구조광 센서가 추적할 수 없는 영역과 센서의 송신부와 수신부의 시점 차이로 인한 폐

색 영역(occlusion)으로 인해 빈 영역이 존재하기 때문에 전처리 과정으로써 인페인팅(inpainting) 기술을 사용해서 빈 영역을 채웠다[19]. 그림 6은 깊이 카메라

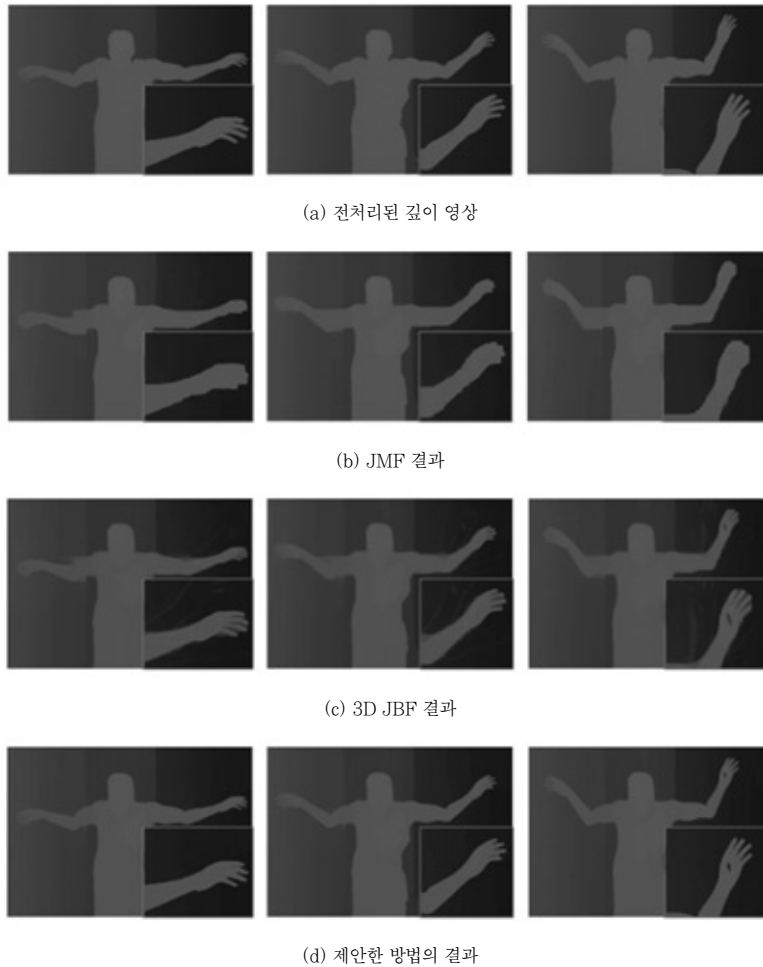


그림 9. “Undo_Dancer” 깊이 영상의 24, 25, 26번째 화면에서의 필터링 결과

표 1. 100 화면의 깊이값의 분산

영역	원본 영상	필터링 결과
1	101.3260	3.8489
2	232.0264	69.4549
3	571.6297	510.8932
4	13.3535	7.4868

로부터 획득한 영상을 보여준다.

시간적 상관도 향상을 정량적으로 평가하기 위해 그림 6(b)에서 표시한 네개의 4×4 블록에 대해 평균 깊이값을 매 화면 계산함으로써 깊이 변동 현상의 추이를 살펴봤다. 그림 7은 필터링된 깊이 영상의 연속된 세 화면을 보여준다.

그림 8은 네개의 블록에 대한 평균 깊이값을 나타낸다. 점선과 실선은 각각 전처리된 깊이값, 필터링된 깊이값을 나타낸다. 표 1은 100 화면에 대한 깊이값의

분산을 보여준다. 다양한 실험 결과를 통해 제안한 방법이 시간적 상관도를 향상시켜서 깊이값의 변동 현상을 효과적으로 줄인 것을 확인할 수 있었다.

2. 스테레오 정합 방법을 이용한 실험 결과

스테레오 정합 방법의 경우에는 DERS를 사용했고 DERS로 얻은 깊이 영상에 다양한 필터링 방법을 적용했다. 또한 원본 깊이 영상이 존재해서 깊이 영상의 화

표 2. 200 화면의 깊이값의 평균 PSNR

방법	필터 윈도우 크기	테스트 영상	
		Undo_Dancer	Mobile
DERS	해당 없음	38.6973	29.3727
MJBF	1	39.0275	29.4972
	2	40.7222	29.6723
	3	40.7386	29.7597
JMF	1	38.6618	29.4971
	2	38.6652	29.5862
	3	38.5531	29.6783
3D JBF	1	39.6497	29.4717
	2	39.9297	29.3825
	3	40.0300	29.7063
제안한 방법	1	40.1845	29.7258
	2	41.5713	29.8434
	3	41.7083	29.8589

질을 정량적으로 비교 가능한 “Undo_Dancer”, “Mobile” 테스트 영상을 사용했다. 여기서 “Undo_Dancer”, “Mobile” 테스트 영상은 각각 노키아, 필립스에서 배포한 영상으로 컴퓨터 그래픽으로 제작되었으며 움직이는 카메라 환경에서 획득한 영상이다 [20],[21]. “Undo_Dancer”는 1920×1080, “Mobile”은 720×540의 해상도를 가진다. 필터링된 영상은 PSNR 값으로 화질을 비교했다.

표 2는 각 테스트 영상의 200 화면에 대한 깊이값의 평균 PSNR을 보여준다. 이 결과를 통해 제안한 방법이 기존의 방법들에 비해 깊이 영상의 화질을 향상시켰음을 확인할 수 있었다. 필터 커널크기가 3인 경우 평균 PSNR은 DERS, MJBF, JMF, 3D JBF에 비해 각각 1.75dB, 0.53dB, 1.67dB, and 0.92dB만큼 증가한 것을 확인할 수 있었다.

그림 9는 “Undo_Dancer” 영상에 대해 필터링을 적용한 깊이 영상의 연속된 세 화면을 보여준다. 이런 실험 결과를 통해 제안한 방법이 시간적 상관도를 향상시켜서 깊이값의 변동 현상을 효과적으로 줄인 것을 확인할 수 있었다. 그림 9(d)에서 알 수 있듯이 제안한 방법이 기존의 방법들에 비해 객체의 경계를 깨끗하게 복원하면서 시간적 상관도를 향상시킨 것을 확인했다.

V. 결론

본 논문에서는 Gaussian 가중 최소 자승법을 이용한 깊이 영상의 시간적 상관도 향상 방법을 제안했다. 제안한 방법에서 사용한 변형된 결합형 다방향 필터는 공간 필터와 밝기 및 깊이에 대한 범위 필터로 구성됐다. 객체 및 카메라의 움직임을 고려하기 위해 제안한

방법은 Gaussian 가중 최소 자승법을 이용했다. 실험 결과를 통해 제안한 방법이 기존의 방법들에 비해 평균 약 1.22dB의 깊이 영상의 화질을 향상시킨 것을 확인했다. 시간적 상관도 측면에서 봤을 때에도 제안한 방법이 깊이 영상의 변동 현상을 효과적으로 줄인 것을 확인했다.

[참고문헌]

- [1] A. Smolic, K. Muller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, T. Wiegand, "3D Video and Free Viewpoint Video - Technologies, Applications and MPEG Standards," Proc. IEEE International Conference on Multimedia and Expo (ICME), Jul. 2006, pp. 2161-2164.
- [2] “제임스 캐머런 ‘상상력, 기술의 新 르네상스,’” <http://www.mt.co.kr/view/mtview.php?type=1&no=2010051308483641428&outlink=1>.
- [3] C. Fehn, "Depth-image-based Rendering (DIBR), Compression and Transmission for a New Approach on 3DTV," Proc. SPIE Stereoscopic Displays and Virtual Reality Systems, Vol. 5291, May 2004, pp. 93-104.
- [4] C. Fehn, K. Hopf and Q. Quante, "Key Technologies for an Advanced 3D-TV System," Proc. SPIE Three-Dimensional TV, Video and Display, Oct. 2004, pp. 66-80.
- [5] W. R. Mark, L. McMillan, G. Bishop, "Post-Rendering 3D Warping," Proc. Symposium on Interactive 3D Graphics, Apr. 1997, pp. 7-16.
- [6] D. Sharstein and R. Szeliski, "A Taxonomy and

- Evaluation of Dense Two-frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, Vol. 47, issue 1-3, Apr. 2002, pp 7-42.
- [7] P. Perez, "Markov Random Fields and Images," *CWI Quarterly*, Vol. 11, No. 4, Jul. 1998, pp. 413-437.
- [8] M. Bleyer and M. Gelautz, "Graph-based Surface Reconstruction from Stereo Pairs Using Image Segmentation," *Proc. SPIE Electronic Imaging*, Vol. 5665, Feb. 2005, pp. 288-299.
- [9] Y. Deng, Q. Yang, X. Lin, and X. Tang, "A Symmetric Patch-based Correspondence Model for Occlusion Handling," *Proc. IEEE International Conference on Computer Vision*, Oct. 2005, pp. 1316-1322.
- [10] P.F. Felzenszwalb and D.P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *International Journal of Computer Vision*, Vol. 70, No. 1, Jan. 2006, pp. 41-54.
- [11] J. Sun, Y. Li, S. Kang, and H. Shum, "Symmetric Stereo Matching for Occlusion Handling," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2005, pp. 399-406.
- [12] H. Tao, H. S. Sawhney, and R. Kumar, "Dynamic Depth Recovery from Multiple Synchronized Video Streams," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Dec. 2001, pp. 118-124.
- [13] E. S. Larsen, P. Mordohai, M. Pollefeys, and H. Fuchs, "Temporally Consistent Reconstruction from Multiple Video Streams Using Enhanced Belief Propagation," *Proc. IEEE International Conference on Computer Vision*, Oct. 2007, pp. 1-8.
- [14] J. Kopf, M. F. Cohen, D. Lischinski, and M. Uyttendaele, "Joint Bilateral Upsampling," *Proc. SIGGRAPH*, Aug. 2007, pp. 96-100.
- [15] P. Lai, D. Tian, and P. Lopez, "Depth Map Processing with Iterative Joint Multilateral Filtering," *Proc. Picture Coding Symposium*, Dec. 2010, pp. 9-12.
- [16] Q. Yang, L. Wang, and N. Ahuja, "A Constant-space Belief Propagation Algorithm for Stereo Matching," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 1458-1465.
- [17] J. Choi, D. Min, and K. Sohn, "2D-Plus-Depth Based Resolution and Frame-rate Up-conversion Technique for Depth Video," *IEEE Transactions on Consumer Electronics*, Nov. 2010, pp. 2489-2497.
- [18] M. Tanimoto, T. Fujii, and K. Suzuki, "Reference Software of Depth Estimation and View Synthesis for FTV/3DV," *ISO/IEC JTC1/SC29/WG11, M15836*, Oct. 2008.
- [19] A. Telea, "An Image Inpainting Technique Based on the Fast Marching Method," *Journal of Graphics Tools*, Vol. 9, No. 1, 2004, pp. 25-36.
- [20] M. Hannuksela and D. Rusanovskyy, "Extension of Existing 3DV Test Set toward Synthetic 3D Video Content," *ISO/IEC JTC1/SC29/WG11, M19221*, Jan. 2011.
- [21] F. Bruls, R. K. Gunnewiek, P. Walle, "Philips response to new Call for 3DV Test Material: Arrive book & Mobile," *ISO/IEC JTC1/SC29/WG11, M16419*, Apr. 2009.



이상범
(Sang-Beom Lee)

2004. 8: 경북대학교 전자전기공학부 학사
 2006. 8: 광주과학기술원 정보통신공학과 석사
 2006. 9~현재: 광주과학기술원 정보통신공학부 박사과정
 관심분야: 디지털 영상처리, 3차원 영상 획득,
 3차원 영상 부호화, 실감 방송
 E-mail: sblee@gist.ac.kr
 Tel: +82-62-715-2258



호요성
(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, 실감 방송
 E-mail: hoyo@gist.ac.kr
 Tel: +82-62-715-2211