

양안식 비디오에서 빠른 깊이 정보 탐색 방법

문지훈, 호요성
광주과학기술원 정보통신공학부
{jhm,hoyo}@gist.ac.kr

Fast Depth Estimation Method in Stereoscopic Video

Ji-Hun Mun, Yo-Sung Ho
School of Information and Communications
Gwangju Institute of Science and Technology (GIST)

요약

양안식 카메라로 촬영된 영상으로부터 깊이 정보를 획득하기 위해 전역 혹은 지역 기반 스테레오 매칭을 수행한다. 지역 기반 스테레오 매칭 방법은 윈도우의 크기와 변위 탐색 범위에 따라 깊이 정보를 얻는 시간이 많이 다르다. 본 논문에서는 지역 기반 스테레오 매칭 방법을 사용하는 과정에서 시간 축 정보를 이용하여 스테레오 매칭의 시간 복잡도를 감소시키는 방법을 제안한다. 시간 축 정보로 사용되는 초기 변위 정보로는 일반적인 스테레오 매칭 결과, 프레임 간 움직임 차이 그리고 주어진 깊이 정보를 사용한다. 본 논문에서 제안한 방법의 실험 결과를 통해 기존의 스테레오 매칭 방법보다 약 1/11 배 정도 시간 복잡도를 감소시킬 수 있었다. 또한 주어진 깊이 정보를 변위 정보로 변환하여 초기 정보로 사용할 경우, PSNR은 기존 매칭 방법에 비해 약 1dB 증가한 것을 확인할 수 있었다.

1. 서론

다양한 3D 영화 흥행과 3D 기능을 지원하는 스마트 TV의 발달로 인해 많은 사람들이 3D 기술에 관심을 갖고 다양한 콘텐츠를 제작하고 있다.

3D 영상을 제작하기 위해서는 기본적으로 동일 장면에 대한 색상 영상과 깊이 영상을 동시에 가지고 있어야 한다. 최근에는 디스플레이 기기들의 해상도 증가와 하드웨어의 연산속도 개선으로 인해 다시점 카메라를 통해 영상 콘텐츠를 제작하고 있다.

동일한 시점에서 색상 영상과 일치하는 깊이 영상을 획득하기 위해 일반적으로 두 가지 방법을 사용한다. 먼저 Kinect 카메라와 ToF(Time of Flight) 카메라와 같이 깊이 정보를 측정할 수 있는 카메라를 색상 영상과 동일 위치에서 촬영하여 깊이 정보를 획득할 수 있다 [1].

깊이 측정 카메라를 사용하지 않고 깊이 정보를 획득할 수 있는 다른 수단으로는 스테레오 매칭을 이용하는 방법이 있다. 스테레오 카메라는 사람의 눈과 유사한 원리를 이용한다. 사람이 입체감을 느낄 수 있는 원리는 두 눈에서 받아들이는 시각 정보들 간의 시간차이 때문에 입체감을 느낄 수 있게 된다.

스테레오 매칭 방법은 일반적으로 전역과 지역 방법으로 나눌 수 있다. 전역 기반 매칭 방법은 깊이 정보를 획득하기 위해 하나의 화소에 대해 대응되는 영상의 모든 화소와 비교를 통해 깊이 정보를 획득하는 방법이다 [2].

지역 기반 매칭은 사용자에 의해 지정된 변위 탐색 범위와 윈도우의 크기에 따라 매칭을 수행하게 된다. 이 과정에서 변위 탐색 범위나 윈도우의 크기에 따라 지역 기

반 스테레오 매칭 결과를 획득하기 까지 시간이 크게 차이날 수 있다.

본 논문에서는 이러한 복잡도 문제를 해결하기 위해 시간 축 정보를 이용한 지역 기반 스테레오 매칭 방법에 대해 설명한다. 초기 정보로 사용되는 3가지 정보에 따라 시간 복잡도와 획득한 깊이 영상의 정확성도 달라짐을 확인할 수 있었다. 다양한 초기 정보를 이용해 스테레오 매칭을 수행한 결과 기존 방법대비 복잡도 감소와 정보의 정확성을 개선할 수 있었다.

2. 지역 기반 매칭을 위한 시간 축 정보

2.1 일반적인 지역 기반 매칭 결과 기반

지역 기반 스테레오 매칭의 복잡도 감소를 위해 첫 번째 프레임의 좌우 영상으로부터 변위 정보를 획득한다. 첫 프레임에서 연속되는 프레임들은 이전 프레임과 객체의 움직임 정도가 매우 유사하거나 약간 차이가 있다.

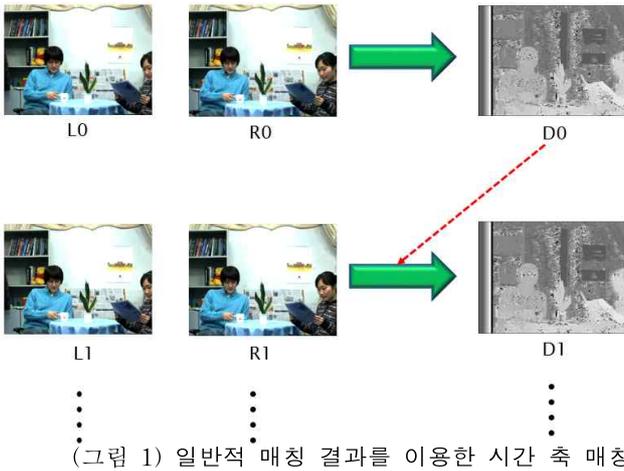
이어지는 프레임에서도 마찬가지로 움직임 차이가 거의 나타나지 않기 때문에 이전 변위 지도를 기반으로 변위 탐색 범위를 제한하여 매칭을 수행한다. 이때 획득한 변위 정보는 연속해서 이어지는 프레임에서도 이전의 프레임과 움직임 차이가 크지 않기 때문에 획득한 변위 지도를 기반으로 지역 기반 스테레오 매칭을 수행한다.

그림 1은 초기 변위 정보를 이용하여 지역 기반 스테레오 매칭을 수행하는 과정을 나타낸다. 좌영상 L0와 우영상 R0에 대해 지역 기반 스테레오 매칭을 수행하면 변위 지도 D0이 생성된다. 일반적으로 L0와 R0로부터 D0을 얻

기 위해서 지역 기반 스테레오 매칭을 수행한다.

획득한 초기 변위 정보는 이어지는 프레임인 L1과 R1의 매칭 과정에서 변위 탐색 범위를 제한하는데 사용된다. 초기 변위 정보를 기반으로 L1과 R1매칭을 수행하여 획득한 변위 정보도 초기 변위 정보와 유사한 변위 값을 갖게 된다.

L0와 R0로부터 변위 정보를 획득하기 위한 과정은 초기 변위 정보를 계산하기 위한 과정이므로 시간 측정 과정에서 제외한다. L1과 R1의 변위 정보를 획득하는 과정부터 매칭의 수행시간을 측정하도록 한다.



2.2 프레임 간 움직임 차이 기반

양안식 비디오 영상에서 색상 영상의 n-1번째 프레임과 n번째 프레임 사이에는 객체의 움직임이 존재한다. 스테레오 매칭을 수행할 때 객체의 움직임이 있는 영역에서 정확한 변위 정보를 얻게 된다면 전체적인 매칭 결과 영상의 정확성도 높아지게 된다.

본 절에서는 양안식 비디오 영상에서 객체의 움직임 정보를 얻기 위해 동일 시점의 두 장의 연속된 프레임으로부터 움직임 영역만을 획득하는 방법을 소개한다.

연속된 두 장의 프레임인 L0와 L1의 차이를 계산함으로써 움직임 정보를 얻게 된다. 동일 시점 영상에서의 차이를 연산하기 때문에 영상 내에 위치한 객체의 움직임이 있는 영역만 나타나게 된다. 동일 시점인 경우 대부분의 배경 정보는 바뀌지 않으므로 객체의 움직임 정보를 정확히 얻을 수 있다.

객체의 움직임 정보와 일반적인 지역 기반 스테레오 매칭 결과 정보를 동시에 사용하여 움직임 영역에서의 변위 정보를 정확히 획득할 수 있다. 하지만 움직임 차이 정보를 얻기 위해 연속되는 프레임간의 차이를 연산하게 되면 객체의 움직임 영역과 배경 영역의 불필요한 잡음 영역까지 결과를 얻게 된다.

객체의 움직임 영역만을 정확히 획득하기 위해서는 불필요한 잡음 영역을 제거해야 할 필요가 있다. 잡음 영역을 제거하기 위해 문턱 값을 변화시켜가며 최대한 객체의

움직임 정보는 남겨두면서 불필요한 영역의 정보는 제거한다. 그림 2에서 문턱 값의 변화에 따른 불필요한 잡음 영역의 제거 변화 정도를 나타내고 있다.

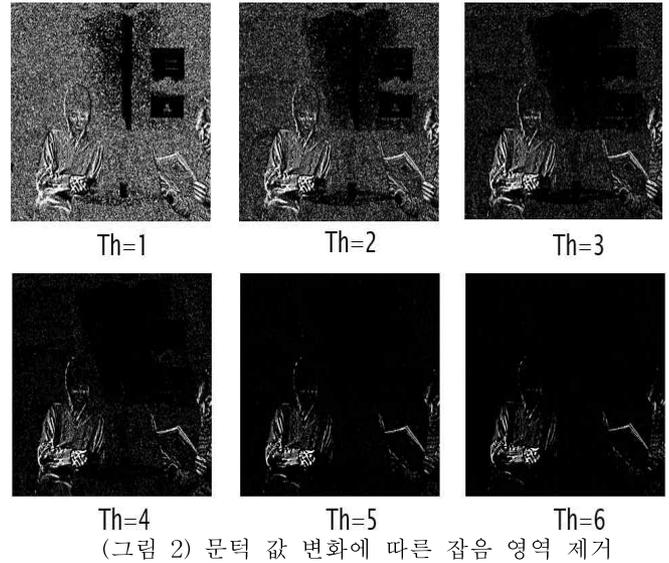


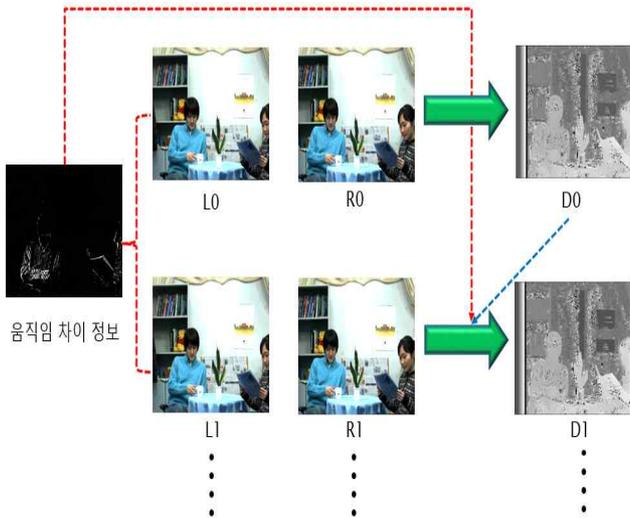
그림 2에서 문턱 값이 1인 경우에는 객체를 포함한 배경 영역에서 잡음이 많이 검출되는 것을 확인할 수 있다. 문턱 값을 증가시켜 가면서 배경영역의 정보를 제한하며 객체 정보를 유지할 수 있는 문턱 값을 찾는다. 본 논문에서는 문턱 값으로 5를 사용했는데, 그 이유는 문턱 값이 5를 초과할 경우 객체의 움직임 영역 정보도 제거되어 정확한 스테레오 매칭 정보로 사용할 수 없기 때문이다.

2.1절에서는 초기 변위 정보를 획득한 뒤 이어지는 모든 프레임들의 스테레오 매칭 정보를 획득하기 위해 동일하게 변위 탐색 범위를 제한했다. 하지만 본 절에서 제안하는 방법은 움직임이 탐지된 영역에서만 변위 탐색 범위를 증가시키고, 나머지 영역에서는 기존의 초기 변위정보 기반 스테레오 매칭 방법을 따른다.

이때, 움직임 정보가 있는 영역에서는 기존에 알고 있는 최대, 최소 변위 값을 이용하여 스테레오 매칭을 수행하며 움직임 정보가 없는 영역에서는 최소한의 변위 탐색 범위를 지정하여 스테레오 매칭을 수행한다.

이러한 이유 때문에 일반적인 지역 기반 스테레오 매칭 방법보다 약간의 시간 복잡도는 증가하지만 보다 정확한 변위 정보를 획득할 수 있게 된다. 즉 객체의 움직임이 탐지된 영역에서는 변위 탐색 영역을 증가 시켜 스테레오 매칭을 수행하고, 움직임이 없는 배경영역의 경우에는 아주 작은 변위 영역에서만 스테레오 매칭을 수행할 수 있도록 설정한다.

초기 정보로는 움직임 차이 정보와 첫 프레임의 스테레오 매칭 결과를 사용한다. 연속되는 프레임들 사이의 스테레오 매칭은 이전 단계에서 획득한 변위 정보를 사용한다. 그림 3에서 이 과정을 나타내고 있다.



(그림 3) 움직임 정보를 이용한 시간 축 매칭

그림 3에서 L0와 R0를 통해 획득한 D0은 다른 조건없이 기본적인 지역 기반 스테레오 매칭을 통해 획득하는 정보이기 때문에 시간 축적 과정에서 제외한다. L1과 R1의 스테레오 매칭 수행 시 움직임 차이 정보와 D0을 사용하며, 이어지는 프레임에서도 동일한 방식을 사용하기 때문에 이 과정의 연산 시간을 측정하도록 한다.

2.3 주어진 깊이 정보 기반

앞에서 설명했듯이, 깊이 정보는 깊이 카메라를 이용하는 방법과 스테레오 매칭을 통해 획득하는 방법이 있다. 이번 절에서는 초기 정보를 깊이 카메라로 촬영한 영상을 이용하여 변위 탐색 범위를 제한하는 스테레오 매칭 방법에 대해 설명한다.

깊이 카메라로 획득한 영상은 실제 촬영한 영상의 깊이 정보를 가지고 있으므로, 변위 정보로 변환을 해야 한다. 식 1은 Z_{near} , Z_{far} 와 변위 값들의 상관관계를 설명하고 있다.

$$\begin{aligned} Z_{near} &= \frac{f \cdot l}{d_{max} + \Delta d} \\ Z_{far} &= \frac{f \cdot l}{d_{min} + \Delta d} \end{aligned} \quad (1)$$

스테레오 매칭을 위한 초기 정보로 사용하기 위해서는 식 1에 있는 d_{max} 와 d_{min} 의 관계에 대한 식으로 변환해야 한다. 식 1에서 분모에 있는 Δd 는 카메라 offset값을 의미하는데 보통 변위 값을 계산할 때는 이 값을 고려하지 않아도 무관하다.

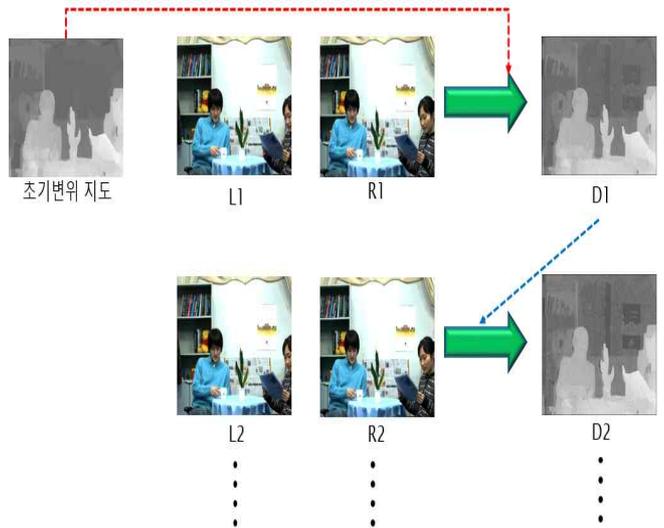
변위 값을 계산하기 위해 식 1에서 d_{max} 와 d_{min} 을 좌변으로 이항한 뒤 나머지 값들을 우변으로 이항한다. 이 결과 식은 깊이값과 카메라의 초점 거리와 카메라 거리의 값으로 이루어진 변위 정보를 얻을 수 있게 된다. 이 결과는 식 2에서 확인할 수 있다.

$$\begin{aligned} d_{max} &= \frac{f \cdot l}{Z_{near} + \Delta d} \\ d_{min} &= \frac{f \cdot l}{Z_{far} + \Delta d} \end{aligned} \quad (2)$$

주어진 깊이 정보에서 식 2를 이용하여 스테레오 매칭을 수행할 때 사용되는 최대, 최소 변위 값을 알 수 있다. 또한, 깊이 지도는 스테레오 매칭의 초기 정보로 사용되기 때문에 식 1을 이용하여 각 화소별로 깊이 값을 변위 값으로 변환한 변위 지도를 획득해야 한다. 이 관계식은 식 3과 같이 나타낼 수 있다.

$$Z = \frac{1.0}{\frac{v}{255.0} \cdot \left(\frac{1.0}{Z_{near}} - \frac{1.0}{Z_{far}} \right) + \frac{1.0}{Z_{far}}} \quad (3)$$

식 3에서 v 는 깊이 카메라로부터 촬영하여 획득한 깊이 지도에서 변위 정보로 변환하고자 하는 화소의 값을 의미한다. 즉, 모든 깊이 지도에서 화소의 위치를 바꿔가며 변위 정보를 획득한다. 그림 4에 예서는 깊이 정보를 변위 정보로 변환한 뒤, 변위 탐색 영역을 제한하여 지역 기반 스테레오 매칭을 수행하는 과정을 나타낸다.



(그림 4) 깊이 정보를 이용한 시간 축 매칭

이 방법 또한 마찬가지로 초기 변위 지도는 L0에 해당하는 정보이므로, L1과 R1의 스테레오 매칭 과정에서 사용된다.

정확한 시간 측정을 위해 초기 변위 지도 정보를 기반으로 스테레오 매칭을 수행하는 과정부터 연속되는 프레임들에 대한 스테레오 매칭 연산 시간을 측정한다. 즉 L2와 R2의 스테레오 매칭 수행 시간부터 측정을 해야 초기 정보를 이용한 스테레오 매칭의 시간 효율을 알 수 있다.

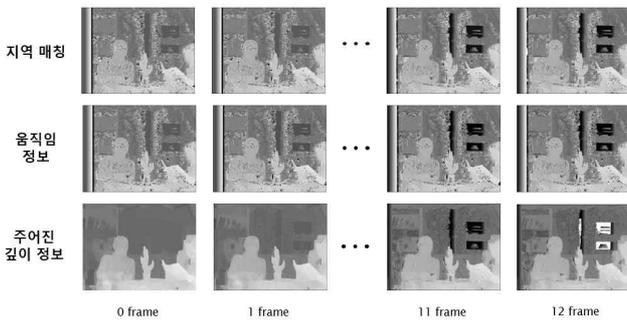
3. 실험 결과

본 논문에서 제안하는 다양한 지역 기반 스테레오 매칭 방법의 타당성을 증명하기 위해 실험한 환경은 표 1과 같다.

(표 1) 실험을 위한 PC환경

상세 사항	
CPU	Intel(R) Core(TM) i7-5960X 3.00GHz
RAM	32.0GB
Compiler	Visual Studio 2013
OS	Window 8.1

제안하는 다양한 초기 정보를 이용한 지역 기반 스테레오 매칭 결과의 성능을 입증하기 위해 각 실험 영상들을 동일한 프레임과 동일 시점을 갖도록 하였다. 실험 결과는 그림 5에서 확인할 수 있다. 초기 입력 정보에 따라 이어지는 영상에서 획득되는 변위 정보의 정확성이 달라지는 것을 알 수 있다.



(그림 5) 깊이 정보를 이용한 시간 축 매칭

표 2에서는 스테레오 매칭 과정에서 사용되는 윈도우 크기를 7로 지정한 뒤, 다양한 초기 정보를 이용한 매칭 결과를 나타내고 있다. 일반적으로 획득한 변위 정보를 평가하기 위해서는 원본 깊이 영상 값(ground truth)을 사용하여 BPR (bad pixel rate)을 측정한다.

하지만 본 실험에 사용한 비디오 영상은 정확한 원본 깊이 영상 값이 없으므로 우선 합성을 통해 중간 시점 영상을 생성 한다. 합성을 통해 생성된 영상과 기존에 촬영된 색상 영상과의 화질과 비교하여 획득한 변위 정보의 정확성을 판단한다. 합성된 영상은 생상 영상이므로 PSNR을 이용하여 비교하는데 PSNR을 이용해 결과를 비교하는 방법을 식 4에서 나타내고 있다.

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) \quad (4)$$

식 4에서 PSNR은 MAX_I^2 와 MSE (mean square error)의 비율에 의해 결정되는 것을 확인할 수 있다. MSE는 모든 동일 위치의 화소 값 차이와 비교하려는 영상의 해

상도에 따라 결정 된다. 스테레오 매칭을 통해 획득한 깊이 정보를 기반으로 합성한 영상을 이용해 PSNR을 비교하고, 제안한 각 스테레오 매칭 방법을 이용한 시간 복잡도를 비교한 결과를 표 2에서 나타내고 있다.

(표 2) 시간 축 정보를 이용한 매칭 결과

매칭 방법		소요시간(sec)	복잡도 감소	PSNR(dB)
시간 축 정보	기존 방법	50.67	-	23.81
	지역	4.744	1/10	23.70
	움직임	5.728	1/10	23.66
	깊이	4.752	1/11	24.80

표 2에서 움직임 기반 스테레오 매칭 방법은 다른 제안한 방법들과 다르게 움직임이 있는 영역에서는 정해진 범위 탐색 범위에 따라 스테레오 매칭을 수행하므로 다른 방법들에 비해 소요 시간이 증가하게 된다. 기존 방법 대비 깊이 정보를 초기 입력 정보로 사용하여 스테레오 매칭을 수행했을 때, 복잡도는 기존 방법에 비해 1/11배 감소하는 것을 확인할 수 있으며 영상의 화질도 약 1dB 증가하는 것을 알 수 있다.

4. 결론

본 논문에서는 지역 기반 스테레오 매칭을 수행할 때 3가지 서로 다른 초기 정보를 이용하여 시간 복잡도를 줄이는 방법을 제안했다. 실험을 통해 기존의 지역 기반 스테레오 매칭 방법에 비해 깊이 정보를 이용한 스테레오 매칭 방법을 사용해 시간 복잡도를 약 1/11배 감소시켰으며, PSNR은 약 1dB 증가하는 결과를 얻을 수 있었다.

감사의 글

본 연구는 미래창조과학부 ‘범부처 Giga KOREA 사업’의 일환으로 수행하였음. [GK13C0100, 기가급 대용량 양방향 실감 콘텐츠 기술 개발]

참고 문헌

- [1] S.B Lee and Y.S Ho, “Real-time Stereo View Generation using Kinect Depth Camera,” Asia-Pacific Signal and Information Processing Association (APSIPA), pp. 1-4, Oct. 2011.
- [2] Chen W, Zhang M-J and Xiong, Z-H, “Fast semi-global stereo matching via extracting disparity candidates from region boundaries,” Computer Vision, IET, vol. 5, no. 2, Mar. 2011.
- [3] Jiangbo Lu, Rogmans, Sammy, Lafrut. G, and Catthoor. F, “Stream-Centric Stereo Matching and View Synthesis,” IEEE Transaction on Circuits and Systems for Video Technology, vol. 19, no. 11, pp 1598-1611, Nov. 2009.