

## 깊이영상의 스테레오 매칭을 이용한 고속 가상시점 영상 생성

\*구자명 \*\*최현준 \*서영호 \*김동욱

\*광운대학교 \*\*목포해양대학교

\*dwkim@kw.ac.kr

## Fast Virtual Image Generation using stereo matching of depth images

\*Koo, Ja-Myung \*\*Choi, Hyun-Jun \*Seo, Young-Ho \*Kim, Dong-Wook

\*Kwangwoon Univ. \*\*Mokpo National Maritime Univ.

## 요약

본 논문에서는 자유시점 또는 오토스테레오스코픽 비디오나 홀로그래픽 비디오 서비스 들을 목표로 가상시점 영상을 빠르게 생성하는 방법을 제안한다. 이 방법에서 좌/우 깊이영상이 주어졌다고 가정하고, 두 깊이 영상을 이용하여 깊이 값과 변위를 계산 한 후 깊이 당 변위증분을 계산하여 사용한다. 계산된 깊이 값 당 변위증분과 해당 픽셀의 깊이정보를 이용하여 텍스처 이미지를 원하는 가상 시점으로 이동시킨다. 좌영상과 우영상을 원하는 가상시점으로 각각 생성한 후, 잡음과 disocclusion 을 효과적으로 제거하기 위하여 두 영상을 결합한다. 그 다음 남은 홀을 효과적인 hole-filling으로 제거하여 가상 시점 영상을 생성한다.

## 1. 서론

최근 영상 매체에 대한 대중의 관심은 HD(High Definition)을 넘어 UHD(Ultra-High Definition) 그리고, 3D에 쏠리고 있다. 특히 3D를 통해 입체감을 느끼고 실제 사물이 앞에 있는 듯한 착각에 들며 좀더 현실감 있는 영상을 볼 수 있는 것에 매력을 느끼고 있다. 하지만, 현재의 3D 영상에서는 입체감을 느낄 수 있지만 이를 위해서는 어디까지나 정면에서 바라보아야만 입체감을 느낄 수 있으며, 그에 따라 3D를 보는데 있어서 사용자는 불편함을 느끼며 3D에서 중요한 입체감을 제대로 느낄 수가 없는 것과 같은 문제점이 발생하게 된다. 시점에 제약이 없는 3D 영상에 대한 욕구로 발전하면서 시점에 제약이 없는 자유시점(free view-point) 영상과 오토스테레오스코프(auto-stereoscope)가 나타났다. 하지만 이를 위해서는 다시점 영상이 필요하지만 현재의 HD 해상도에서 다양한 시점의 화면을 전송할 경우 현재보다 몇 배 이상 많은 데이터를 보내야하는 점에서 실현에 어려움을 느끼고 있다. 이를 해결하기 위해서 사용자에게 필요한 데이터를 보낸 후 사용자의 단말에서 데이터를 사용하여 다시점을 생성하는 방법을 연구하게 되었다.

지금까지 IVR과 같은 중간 시점 생성 방법에 관한 연구가 많이 진행되었다.[1][2]. 또한, 깊이 정보를 이용한 중간 시점 생성도 다양한 환경과 목적에 의해 제안되었다.[3][4][5][6].

이 논문에서는 두 시점의 깊이 영상을 이용하여 원하는 가상 시점을 생성하는 방법에 대하여 기술하고 있다. 이 방법에서 좌/우 깊이영상이 주어졌다고 가정하고, 이 깊이영상들은 카메라 보정과 영상보정이 수행되었다고 가정하였다. 이미 주어진 왼쪽 깊이 영상과 오른쪽 깊이 영상을 기반으로 특정 깊이 값에 대한 변위정보를 얻고 깊이 당 변위증분을 계산하여 원하는 가상시점 영상을 생성하는 방법이다.

## 2. 본론

깊이 정보를 얻기 위해서는 현재 두 가지 방법이 존재하는데 하나는 stereo matching 기법과 다른 하나는 depth camera를 이용하여 깊이정보를 얻는 방법이다. stereo matching은 깊이정보를 얻는 방법 중에서 가장 많이 알려지고 오랫동안 연구가 되어온 방법이다. 두 영상에 대한 stereo matching을 통하여 두 영상에서 대응되는 두 점사이의 위치, 즉 변위(disparity)를 얻게 된다. 이렇게 stereo matching을 통하여 변위지도를 얻게 되고 이를 이용하여 깊이영상을 얻을 수 있다. 변위정보와 깊이정보의 관계는 다음 식 (1)과 같다.

$$z = \frac{Bfs}{d'} \quad (1)$$

여기서  $z$ 는 깊이 값,  $B$ 는 두 카메라간 거리,  $f$ 는 카메라의 초점거리,  $d'$ 는 스케일링된 변위 값,  $s$ 는 스케일링 요소를 각각 나타낸다. 따라서  $d' = sd$ 이며,  $d$ 가 실제 변위 값이다. 즉, 식 (1)에서  $B$ ,  $f$ ,  $s$ 를 알면 변위정보를 깊이장버러 쉽게 변환할 수 있다.

최근 출시된 다양한 depth camera는 대부분 TOF(time-of-flight) 방식을 사용하고 있다. 이 방법은 카메라의 적외선 수신부 주변에 있는 적외선 발생기를 물체에 조사하여 그것이 물체에서 반사 되어 돌아오는 시간을 계상하여 해당 물체의 깊이를 측정하는 방식이다. 일반적인 시간차 측정 방식은 phase difference를 이용한 방식인데, 식 (2)에 나타내고 있다.

$$z = \frac{1}{2} \cdot \frac{c\Delta\phi}{\omega} \quad (2)$$

c는 빛의 속도,  $\Delta\phi$ 는 phase difference,  $\omega$ 는 파장의 각속도이다. 자세한 내용은 [7][8]를 참조한다.

그림 1.은 제안하는 방법의 전체 흐름도를 나타내고 있다. 그림에서 보는 바와 같이, 제안한 방법은 전처리(pre-processing) 과정, 초기 생성(initial generation 과정, 후처리(post-processing) 과정의 세 과정으로 이루어진다. 본 논문에서는 좌/우의 깊이영상이 주어지고, 이 두 깊이 영상은 보정과정을 거쳤다고 가정한다.

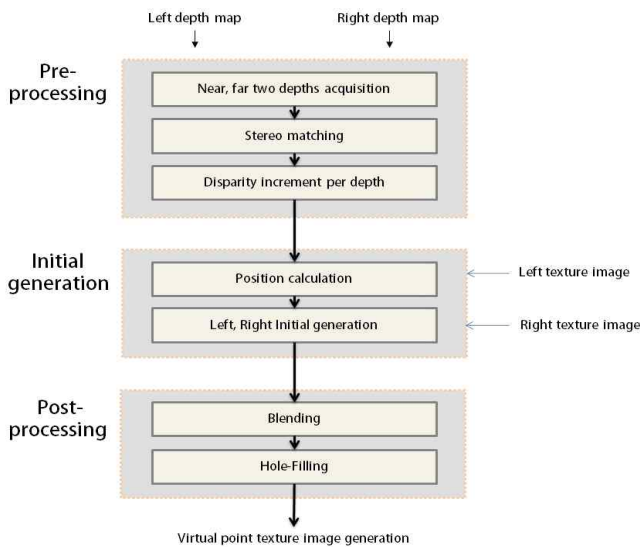


그림 1. 제안한 전체 알고리즘

전처리 과정은 깊이 값 1당 변화하는 변위증분값을 계산하는 것이다. 변위 값을 계산하기 앞서 정확한 변위 값을 얻기 위해서 먼저 에지 추출을 한다. 그 다음 변화가 심한, 즉 에지가 강한 픽셀 순서대로 샘플을 추출한다. 추출한 샘플들을 선별하여 미리 정의가 된 일정 값에 따라  $z_{T, near}$ ,  $z_{T, far}$ , 즉 두 개의 depth를 결정한다. 그 후, 두 개의 depth를 위주로 stereo matching을 수행하여 왼쪽 영상에서의 한 점과 오른쪽 영상에서의 한 점의 거리, 즉, 변위를 얻게 된다. 그렇게 얻은 변위값들 중에서도 선별을 하여 신뢰할 수 있는 두 개의 변위 값  $d_{near}$ ,  $d_{far}$ 을 얻게 된다. 이렇게 얻은 4개의 정보를 가지고 깊이 값 1당 변화하는 변위증분값을 계산하게 되는데 그 식은 다음 식 (3)에 서술하였다.

$$d_{incr} = \frac{d_{near} - d_{far}}{z_{T, near} - z_{T, far}} \quad (3)$$

그 다음, 초기 가상 시점 생성 과정에서는 모든 픽셀에 대하여 얻은 변위증분값을 가지고 원하는 가상 시점으로 이동하게 된다. 식 (4)(5)와 식 (6)(7)은 각각 왼쪽영상과 오른쪽 영상에서의 가상 시점 이동식을 나타낸다.

$$d_{i, left} = d_{near, L} + (z_{T, left} - z_{T, near}) \times d_{incr} \frac{h_{left}}{h} \quad (4)$$

$$x_{i, IVPD} = x_{i, left} - d_{i, left} \quad (5)$$

$$d_{i, right} = d_{near, L} + (z_{T, right} - z_{T, near}) \times d_{incr} \frac{h_{right}}{h} \quad (6)$$

$$x_{i, IVPD} = x_{i, right} + d_{i, right} \quad (7)$$

다음 그림 2는, 초기 가상 시점 생성을 한 결과 영상이다. 가상 시점은 왼쪽과 오른쪽 영상의 중간 시점으로 잡았다.

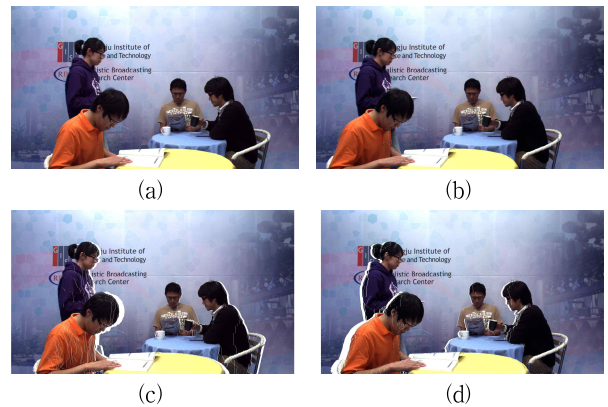


그림 2. 초기 가상시점 생성 과정: (a)왼쪽 원본 영상 (b)오른쪽 원본 영상 (c)왼쪽 중간시점 생성 (d)오른쪽 중간시점 생성

그림 2에서 (c),(d)에서 볼 수 있듯이 하얀색의 부분들을 확인할 수 있는데 시점이동을 하면서 원본영상에서 보이지 않던 영역이 가상 시점으로 이동하면서 새로 생기게 되는데, 이러한 영역을 비폐색영역(disocclusion)이라고 한다. 이러한 영역들은 원본영상에서는 없는 데이터이기 때문에 이 영역을 채우는 것이 문제이다. 이런 부분을 해결하기 위해서 원본 영상이 최소 2개 이상 필요한 것이다. 비폐색영역에 해당하는 정보를 다른 시점에서 찾아서 채우는 방법이 가능하기 때문이다. 그 이유는 오른쪽시점에서 왼쪽으로 이동한 영상은 비폐색영역이 객체의 왼쪽에 생기는데, 이 영역의 데이터를 왼쪽에서 오른쪽으로 이동한 영상에서 존재하는 데이터이므로, 비폐색영역은 다른 원본 영상에서의 데이터를 이용하여 채울 수가 있다[9].

하지만 그림 3처럼 blending후에도 hole들이 남게 되는데 이것은 원본 2개의 영상에 존재하지 않는 데이터가 있기 때문이다.



그림 3. 확대한 Hole 영역

이러한 hole들은 depth정보를 이용하여 없앨 수가 있다. 먼저 hole 부분을 추출한 후, left, right 방향으로 깊이 존재하는 부분까지 search를 하여 찾은 두 개의 픽셀 값 중에서 depth가 가장 큰 값, 즉 가장 먼 영역에 속하는 부분의 texture pixel로 채우는 것이다. 이러한 방법을 이용하여 나머지 hole들을 없앨 수가 있다.

### 3. 실험 및 결과

우리가 제안한 방법을 적용한 컴퓨터의 프로세서는 Intel Core 2 Quad Q6600 2.40GHz이며 OS는 windows 7 ultimate, 사용한 프로그램은 Microsoft Visual Studio 2010이다. 표 1에 우리가 사용한 테스트 영상들의 정보를 나타냈다.

표1. Test Sequences

Name	Resolution	Sequence length
Cafe	1,920 × 1,080	300 frames
Lovebird1	1,024 × 768	200 frames

그림 4, 5는 테스트 영상들을 이용한 최종 결과영상을 나타내고 있다. 비교 대상으로 MPEG의 VSRS 3.5로 선택하였다. VSRS 3.5는 1D parallel warping을 한 후, 경계선 노이즈 제거를 실시하고, hole들을 제거하는 hole-filling을 한다.



(a)



(b)



(c)

그림 4. 가상시점 생성: (a)중간 시점 원본 영상 (b)제안한 중간 시점 생성 영상 (c)VSRS 3.5 중간 시점 생성 영상



(a)



(b)



(c)

그림 5. 가상시점 생성: (a)중간 시점 원본 영상 (b)제안한 중간 시점 생성 영상 (c)VSRS 3.5 중간 시점 생성 영상

좀 더 정확한 객관적인 평가를 위해서 PSNR를 이용하여 원본 영상과 제안한 방법이 얼마나 일치하는지를 평가하였다. 그림 6, 7이 바

로 그 결과이다.

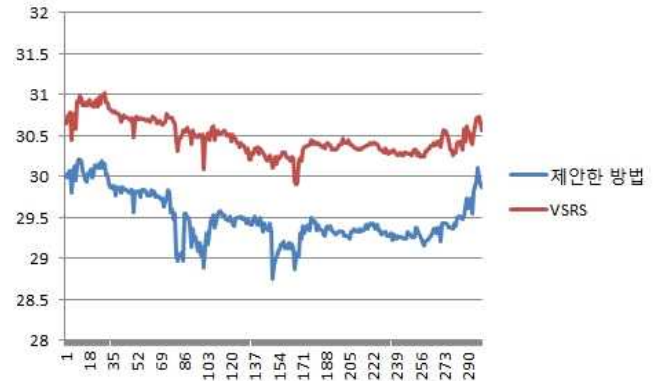


그림 6. Cafe PSNR

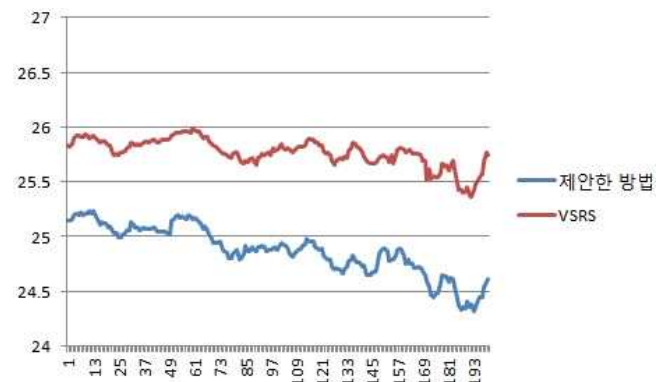


그림 7. Lovebird1 PSNR

표 2는 전체 frame에 대한 평균 PSNR과 제안한 방법이 얼마나 빠르게 가상 시점 영상을 생성하는 지를 1 frame 생성 평균 시간을 나타내고 있다. 우리가 제안한 방법은 첫 frame에서만 깊이 당 변위증분을 구하여 그것을 다음 frame에 적용하기 때문에 생성 속도가 많이 떨어졌음을 알 수 있다.

표2. Measurement

Name	AVR PSNR (dB)		1 frame Time (s)	
	제안한 방법	VSRS	제안한 방법	VSRS
Cafe	29.49	30.48	0.589	4.242
Lovebird1	24.88	25.77	0.17	1.64

### 4. 결론

본 논문에서는 좌/우 깊이영상의 깊이 1당 변위증분을 계산하여 원하는 가상시점의 영상을 빠르게 생성하는 방법을 제안하였다. 현재 가상시점의 영상을 생성하는데 warping 방법을 사용하여 생성하고 있다. 이 방법은 정확하지만 warping에 필요한 데이터인 카메라 내 외부 파라미터들을 사용하여야 한다. 그러나 우리가 제안하는 방법은 이런

데이터가 주어지지 않는다고 가정하여, 좌/우 깊이영상에 대하여 stereo matching을 수행하여 필요한 데이터를 얻고 만드는 것이다. 그리고 얻은 깊이를 1당 변위증분을 첫 번째 프레임만을 대상으로 수행하여 얻은 후 그것을 다음 프레임에도 적용할 수 있기 때문에 추가적인 변위증분값을 얻을 필요가 없다. 이렇게 변위증분값을 계산하여 얻은 후, 이것으로 왼쪽, 오른쪽 영상을 원하는 가상 시점으로 이동시킨 후, blending을 하여 효과적으로 disocclusion을 없앤 후, 남은 hole들은 hole-filling 방식을 사용하여 원하는 가상시점 영상을 생성하게 된다.

제안한 방식의 성능을 평가하기 위해서 기존의 warping 방식을 사용하는 MPEG의 VSRS3.5와 비교하였다. test sequences를 사용하여 객관적인 평가를 위한 PSNR를 구한 결과 우리가 제안한 방법이 평균적으로 05-1dB정도 떨어진 것을 알 수 있다. depth와 texture image와의 불일치로 인하여 생기는 노이즈에 대한 대비 미흡과 효과적으로 hole-filling이 이루어지지 않았기 때문이다.

그러나 1 프레임당의 가상 시점 생성 속도를 비교한 결과 우리가 제안한 방법이 좀 더 빠르고 효과적으로 가상 시점을 생성하는 것을 알 수가 있다. 좀 더 현 제안한 방법을 연구하고 해결책을 마련한다면 좀 더 질적으로 뛰어나고 빠르게 생성하는 방법이 되지 않을까 생각한다.

## 감사의 글

이 논문은 2012년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(NRF-2012-0004985).

## 참고문헌

- [1] A. Smolic, et al., "Intermediate View Interpolation based on Multiview Video plus Depth for Advanced 3D Video Systems," IEEE ICIP, pp.2448-2451, 2008.
- [2] Y. Mori, et al., "View Generation with 3D Warping using Depth Information for FTV," Elsevier J.of Image Communications, Vol. 24, pp. 65-72, Oct. 2008
- [3]. M. Zararim, et al., "A Joint Multiview plus Depth image Coding Scheme Based on 3D Warping," ACM Conf. on 3DVP, 2010.
- [4] E. K. Lee, et al., "High-Resolution Depth Map Generation by Applying Stereo Matching based on Initial Depth Information," IEEE 3DTV-CON, pp. 201-204, 2008
- [5] S.Y. Kim, et al., "Generation of ROI Enhanced Depth Maps using Stereoscopic Cameras and a Depth Camera," IEEE Trans. on Broadcasting, Vol. 54, No. 4, pp. 732-739, Dec. 2008.
- [6] X.Huang, et al., "A Depth Extraction Method based on Motion and Geometry for 2D to 3D Conversion," IEEE Intl. Symp. on IITA, pp. 294-298, 2009
- [7] SR4000 Data Sheet, <http://www.mesa-imaging.ch/prodview4k.php>, MESA Imaging, Oct. 2010.
- [8] Sonam Jamtsho, "Geometric Modeling of 3D Range Cameras and their Application for Structural Deformation Measurement," Dissertation for master degree, Dept of Geomatics Eng., Galgary Univeristy, Dec. 2010
- [9] Cheon Lee et al., "Boundary Noise Removal in Synthesized Intermediate Viewpoint Images for 3D Video." pp. 109-112, Nov. 2008