

FPGA/GPU 기반 다시점 영상 생성 시스템

신흥창, 엄기문, 김찬, 정원식, 허남호
한국전자통신연구원

{hcshin, gmum, ckim, wscheong, namho}@etri.re.kr

FPGA/GPU-based Autostereoscopic 3D Video Generation System

Hong-Chang Shin, Gi-Mun Um, Chan Kim, Won-Sik Cheong, Namho Hur
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문에서는 스테레오 영상으로부터 무안경 3D 디스플레이를 위한 다시점 영상을 생성하는 시스템을 제안한다. 제안한 시스템에서는 먼저 비디오 캡처 카드를 통해 입력되는 스테레오 영상으로부터 FPGA 상에서 구현된 Trellis 동적 프로그래밍 기법에 의해 좌우 변이 영상을 실시간으로 추출한다. 이 변이 영상을 기반으로 좌우 영상 사이에서 중간 시점 영상을 생성한다. 이렇게 추출된 좌우 변이 영상과 좌우 스테레오 영상은 각각 USB 3.0 과 PCI-express 인터페이스를 통해 GPU 로 전송되고, GPU 에서는 이들 데이터를 사용하여 변이 기반 영상 합성 방법을 통해 다시점 영상을 생성한다. 생성된 다시점 영상은 다시점 3 차원 디스플레이 규격에 맞게 재배치되어 재생된다.

1. 서론

최근 시장에 출시되는 TV 의 경우, 3DTV 기능이 기본적으로 탑재되어 있는 경우가 대부분을 차지하고 있을 정도로 3DTV 방송이 이제는 당연한 것이 되고 있다. 이에 따라 3D 영화를 비롯하여 많은 종류의 3D 콘텐츠가 제작되고 있으며, 3D 콘텐츠 시청에 따른 시각피로에 대한 연구도 많이 진행되고 있어, 보다 편안하게 3D 콘텐츠를 감상할 수 있게 되었다. 하지만, 3DTV 의 주류를 이루고 있는 안경식 스테레오스코픽(Stereoscopic) 3D 디스플레이 의 경우, 시청자들이 안경을 쓰는 것에 대한 불편함이 여전히 존재하고 있어, 이를 해결해야 한다는 지속적인 요구가 있다. 이에 대한 대안으로서 무안경 (Autostereoscopic) 3D 디스플레이가 개발되었지만, 다시점 콘텐츠 제작 및 전송 문제는 여전히 해결해야 할 숙제로 남아있다.

이를 위해 다시점 디스플레이 개발과 이에 적절한 다시점 콘텐츠를 생성하는 기술에 대한 많은 연구가 진행되고 있다. 다시점 콘텐츠 생성하고 재생하는 디스플레이로서 한 시점 영상과 그에 맞는 깊이 영상(image+depth)를 입력하면 다시점 영상을 생성하고 재생하는 디스플레이가 출시되었으며 [1], 일정 입력 형식으로 구조화된(Tiled) 다시점 영상 혹은 스테레오스코픽 영상을 입력하면 다시점 영상을 생성하여 재생하는 디스플레이가 출시되기도 하였다[2]. 하지만, 다시점 영상 생성 기술은 2D/3D 변환된 영상의 경우처럼 3 차원 깊이감 및 실재감을 재현하기엔 아직까지는 부족함이 있으며, 아직까지 해결되어야 할 문제점이 많이 남아 있다.

현재의 기술개발 흐름을 보면, 다시점 3DTV 서비스를 제공하기 위한 방법으로는 스테레오/다시점 변환 기법이 가장 현실적인 방안으로 꼽히고 있다. 왜냐하면 현재 개발되어

출시되고 있는 3DTV 및 방송중인 3D 방송시스템과 호환이 되면서 디스플레이 제조사마다 서로 다른 규격으로 개발되고 있는 다시점 3 차원 디스플레이에 적합한 콘텐츠를 생성하여 재생할 수 있어야 하기 때문이다.

관련 기술로서 여러 연구가 진행되고 있다. Berrety 는 image-plus-depth 인터페이스를 가진 다시점 3 차원 디스플레이를 고려하여 실시간 깊이 영상 생성 시스템을 개발하였다[3]. 이 시스템은 720x480 해상도를 가진 깊이 영상을 30fps 로 생성한다. G. Yi 는 FPGA 기반으로 실시간 변이를 추출하고 가상 시점 영상을 생성하는 시스템을 개발하였다[4]. 1024x768 해상도의 시점 영상을 60fps 로 생성할 수 있으며, 기본적으로 우영상의 위치를 조절하여 스테레오스코픽 영상의 3 차원 깊이 범위를 조절하는 방법을 사용하였다. 최근에 Fraunhofer HHI 에서 두 개의 Hexa-Core CPU 와 Nvidia GPU 기반으로 실시간 스테레오/다시점 변환 시스템을 개발하기도 하였다[5].

이상의 최근 연구 사례를 보면 다시점 디스플레이의 콘텐츠 생성을 위해 필요한 변이 영상을 실시간으로 생성하거나 혹은 3 차원 정보를 이용하여 가상 시점 영상을 생성하는 영상 합성을 실시간으로 처리하기 위해 FPGA 혹은 GPU 하나만을 이용하고 있다. 이와 다르게 본 논문에서는 FPGA 와 GPU 를 동시에 사용하고 역할을 분담하는 시스템을 제안한다. 제안한 시스템에서는 크게 FPGA 에서 좌우 변이 영상을 실시간으로 생성하고, GPU 에서 입력 영상 변환과 생성된 변이영상을 이용하여 다시점 가상 시점 영상을 생성한다

본 논문의 구성은 다음과 같다. 1 장 서론에 이어 2 장에서는 본 논문에서 제안한 시스템의 전체 구조에 대해 설명하고, 3 장에서는 FPGA 기반 스테레오 정합 모듈과 GPU

기반 다시점 영상 생성 기법에 대해 설명한다. 다음으로, 4 장에서는 시스템 구현 및 시험 결과를 보이며, 5 장에서 결론을 맺는다.

2. 제안 시스템 구조

그림 1 은 본 논문이 제안하는 전체 시스템 구조를 도식화한 것이다.

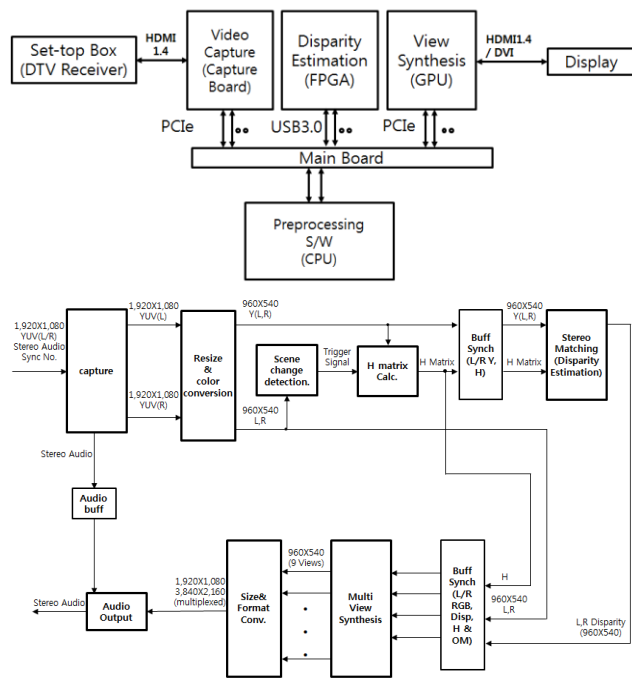


그림 1. 제안 시스템 구조

제안한 시스템의 동작은 다음과 같다. 먼저, 3DTV 수신기로부터 수신된 HD 스테레오 영상 비디오 스트림을 HDMI 1.4 인터페이스를 통해 영상 획득(capture)보드에 입력 받는다. 영상 획득 보드로는 Black Magic Design 사의 Decklink HD Extreme 3D+ card 를 사용하였다. 이때 입력영상신호는 HDMI 인터페이스에 맞게 UYVY 규격으로 들어오게 된다. UYVY 규격에 맞게 입력된 영상은 영상변환 모듈로 전달되는데, 영상변환 과정은 GPU 에서 처리되기 때문에 PCI-Express 인터페이스를 통해 GPU 메모리 상으로 복사된다. 영상변환 모듈에서는 좌우 UYVY 영상을 qHD(960x540)로 크기를 줄임과 동시에 RGB 와 Y 영상으로 변환한다. 이 중 Y 성분 영상은 다시 변이영상을 추출하는 스테레오 정합 모듈과 영상정렬을 위한 H 행렬 계산 모듈에 전달되며, RGB 영상은 장면전환검출 모듈과 다시점 영상합성 모듈에 전해진다. 이 때, 영상정렬 행렬 H 계산에 필요한 특징점 추출은 계산복잡도가 비교적 낮은 Harris corner 검출 방법을 이용하였다[6] [7]. 또한 행렬 H 계산은 매 프레임마다 계산하지 않고, 장면전환이 일어날 때에만 계산되어 갱신된다. 이러한 장면 전환 검출 모듈과 H 행렬 계산 모듈은 CPU 상에서 수행된다.

다음으로 스테레오 정합 모듈에서는 USB 3.0

인터페이스를 통해 qHD(960x540) 좌우 Y 영상과 영상정렬행렬 H 를 입력받아 좌우 영상정렬을 수행하고 정렬된 스테레오 영상을 기준으로 좌우 변이영상을 추출한다. 스테레오 정합 모듈은 Xlink's XC6SLX150 기반의 FPGA 상에서 수행된다. 이렇게 추출된 좌우 변이영상은 호스트인 PC 로 USB 3.0 을 통해 전송되고, 다시점 영상합성을 위해 GPU 메모리로 복사된다.

마지막으로 다시점 영상합성 모듈에서는 변이 기반의 영상합성 기법을 통해 입력된 좌우 컬러 영상과 변이 영상을 사용하여 변이 기반의 영상합성 기법을 통해 9 개의 가상시점 영상을 생성하고, 재생을 위해 이를 다시 화소 재배치 (multiplexing) 하여 다시점 디스플레이로 출력한다. 본 논문에서는 다시점 디스플레이의 시점수를 9 시점으로 가정하였다.

3. FPGA 기반 스테레오 정합 및 GPU 기반 다시점 영상 생성 기법

본 장에서는 제안한 시스템의 핵심 모듈인 FPGA 기반 스테레오 정합 및 GPU 기반 다시점 영상 생성 모듈에 사용된 기법에 대해 좀더 상세하게 살펴보기로 한다.

3.1 FPGA 기반 스테레오 정합 기법

본 논문에서는 FPGA 기반 스테레오 정합 모듈에 사용된 변이 추출 기법으로 Trellis 동적 프로그래밍(Dynamic Programming) 기법[8]을 이용하였다. 그림 2 는 이 기법의 기본 구조 및 개념을 나타내고 있다. 본 논문에서는 그림 2 의 구조를 이용하되 좌우 영상 각각에 대하여 변이영상을 구하도록 구현 하였다.

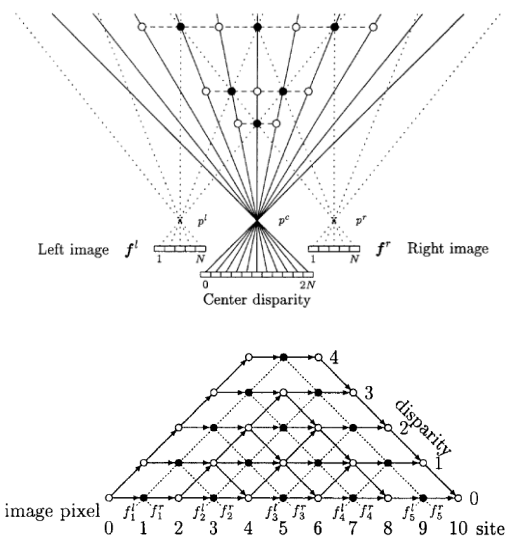


그림 2. Trellis 동적 프로그래밍 기반 스테레오 정합 기법 구조 및 개념 [8]

3.2 GPU 기반 다시점 영상 생성 기법

본 논문에서는 GPU 기반으로 좌우 영상 중간의 다시점 영상생성을 위해 변이 기반 영상합성 기법을 사용한다. 이

기법은 스테레오 영상 (I_L, I_R)과 해당 변이 영상 (D_L, D_R)이 주어지면, 먼저 좌우 영상의 각 화소에 대해 변이를 이용하여 원하는 시점 위치로 사상(mapping)하여 각 입력 영상에 대한 가상 시점 영상을 만들어 낸다. 이 때, 좌우 시점 영상 사이의 거리는 0에서 1로 정규화 되어 있으며, 좌 영상은 0, 우 영상은 1로 정한다. 임의의 중간 시점 위치 α 는 $0 < \alpha < 1$ 의 값을 가진다.

이렇게 중간 시점 각 위치로 사상된 두 영상 $C_L(x)$ 과 $C_R(x)$ 는 다음과 같은 식 (1)에 의해 컬러 값이 합쳐져서(bleending) α 위치에서의 중간 가상 시점영상 N_α 를 생성한다.

$$N_\alpha = C_L \times \frac{|\alpha|}{|1-\alpha|+|\alpha|} + C_R \times \frac{|1-\alpha|}{|1-\alpha|+|\alpha|} \quad (1)$$

이러한 과정을 거쳐 합성된 두 영상에서는 시점 이동에 따라 가려졌다 새롭게 나타나는 영역에 의해 어느 한 시점에서는 정보가 없는 영역이 다른 시점 영상의 사상에 의해 채워질 수 있다. 두 시점 영상의 결합에 의해서도 정보가 없는 영역(hole)이 나타날 수 있는데, 이는 합성에 쓰인 두 참조 시점 영상에서 모두 정보가 없는 경우이므로 이 경우에는 인접 화소들의 중앙값(median)을 이용하여 채워준다. 한편, 이 경우와 달리 변이영상의 부정확함으로 가려짐 영역이 발생하는 경우가 있을 수 있다. 이는 실제로 참조 영상에서의 대응점은 있지만 변이영상의 부정확함으로 가져오지 못하는 경우이다. 이런 경우는 변이정보가 크게 변하지 않는다는 가정 하에 가상시점에서 빈 영역의 인접 좌우 대응점을 찾아서 그 사이의 값들을 순차적으로 가져오으로써 왜곡을 최소화한다.

4. 시스템 구현 및 시험 결과

본 장에서는 제안한 시스템의 구현 및 시험결과에 대해 기술한다. 먼저 그림 3은 본 논문에서 구현한 FPGA 기반 스테레오 정합 장치의 형상을 나타내고 있다. 이 장치는 USB 3.0 인터페이스를 통해 호스트 PC와 연결되고, 입출력 영상을 전송한다. 한편, GPU 하드웨어로는 Nvidia Geforce GTX 580 그래픽카드를 이용하였다. 호스트 PC로는 Window 7 64bit 기반의 i7 프로세서가 탑재된 시스템이 사용되었다.

다음으로, 제안한 시스템의 시험을 위해 그림 4(a)와(b)에 주어진 바와 같은 HD 스테레오 영상을 입력하고 이로부터 9 시점 영상을 생성하여 화질을 살펴보고 속도를 측정해보았다. 그림 4(c)와 (d)는 그림 4(a)와 (b)에 주어진 좌우 영상으로부터 그림 3의 스테레오 정합 장치에 의해 변이를 추출한 결과를 나타내고 있다. 그림에서 보듯이 일부 객체 경계 주변에서 변이 오류에 의한 잡음 현상(streaking noise effect)이 나타남을 볼 수 있다. 이는 경계 보존 3자 필터링(Trilateral filtering) 등을 통해 개선이 가능할 것으로 판단된다. 그림 6은 적응적 교차-3자 변이 필터링(Adaptive Cross-Trilateral Depth Map Filtering)을 적용하여 잡음 현상(streaking noise effect)를 제거한 영상이다[9]. 전체적으로 완벽히 제거가 되진 않았지만, 예지 근처에서 경계를 보존하면서 잡음 제거할 수 있었다. 이 결과는 아직 실시간 처리가 되지 않은 영상이며, 추후 GPU를 통해

실시간으로 구현될 예정이다.

한편, 그림 5는 추출된 변이를 기반으로 각각 중간 시점 위치 0.3과 0.7에서 생성된 중간 시점 영상을 보여주고 있다. 합성된 결과영상을 다시점 디스플레이에서 확인해보면 눈에 많이 거슬리는 artifact는 잘 나타나지 않음을 알 수 있었다.

다음으로, 각 모듈별 수행속도를 측정한 결과, FPGA 기반의 스테레오 정합 모듈은 30fps의 속도로 qHD 해상도를 출력하고 있었으며, GPU 기반의 영상합성 모듈에서의 9 시점 qHD 영상 생성 속도는 0.03 초로 측정되었다. 또한 총 지연시간(latency)은 대략 4 프레임 정도로 나타났다.



그림 3. FPGA 기반의 스테레오 정합 장치 (100x160x68mm)

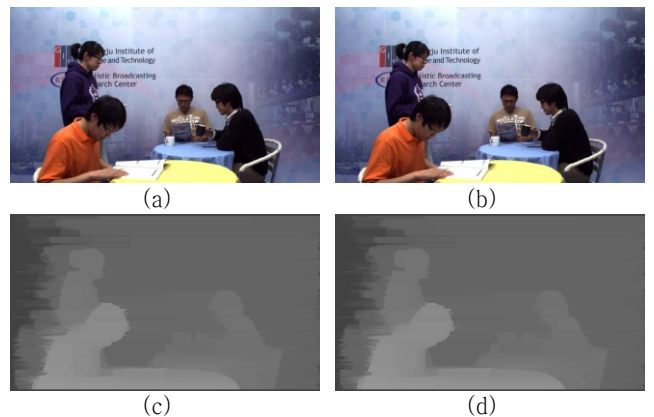


그림 4. 실험 영상(MPEG 3DV Café 영상) 및 변이 추출 결과

(a) 원본 좌영상, (b) 원본 우영상, (c) 좌 변이영상, (d) 우 변이영상

5. 결론

본 논문에서는 FPGA 및 GPU 기반으로 스테레오 영상으로부터 다시점 영상 실시간 생성 시스템을 제안하였고, 구현한 결과를 소개하였다. 제안한 시스템은 3DTV 수신기로부터 HD 스테레오 영상을 수신하고, GPU를 통해 영상크기조절 및 변환 등의 전처리를 수행한 후, FPGA를 통해 추출된 좌우 변이영상을 이용하여 GPU로 9 시점의 중간 시점 영상을 고속으로 생성할 수 있었다.

현재는 FPGA 기반 스테레오 정합 기법의 실시간 변이추출

성능의 한계로 qHD 해상도의 다시점 영상을 생성을 하고 있지만, 앞으로는 성능 개선을 통하여 HD 다시점 영상의 실시간 생성이 가능하도록 구현할 예정이다. 또한 Trellis 동적 프로그래밍 기반 스테레오 정합 기법의 문제점인 객체 경계에서의 잡음(streaking Noise)를 제거할 수 있는 변이 후처리 기법 및 변이결과에 강인한 실시간 영상합성 기법에 대한 연구도 이뤄질 예정이다.

감사의 글

본 연구는 한국방송통신위원회의 한국방송통신전파진흥원 지원 과제인 “다시점 3D 호환 UHD TV 방송 기술 개발(KCA-2012-11921-02001)”의 지원에 의해 수행되었습니다. 본 연구에 사용된 FPGA 기반 스테레오 영상 변이 추출 하드웨어 구현에 수고하신 (주)비전에스티 연구원 여러분께 감사 드립니다.



(a)



(b)

그림 5. 그림 4의 좌우 영상과 해당 변이영상으로부터의 중간시점 영상 합성 결과의 예 (중간시점 위치: (a) 0.3, (b) 0.7)



그림 6. 그림 4(c)변이 영상의 적응적 교차-3자 변이 필터링 적용 결과

참고문헌

- [1] Philips 3D Solutions, 3D Content Generation Guidelines, pages 1-10, Oct., 2005.
- [2] <http://www.techradar.com/reviews/audio-visual/televisions/plasma-and-lcd-tvs/toshiba-55zl2-1080745/review#articleContent>
- [3] R-P. M. Berretty, A. K. Riemens and P.F. Machado, "Real-time Embedded System for Stereo Video Processing for Multi-view Displays," in *Proceedings of the SPIE, - IS&T Electronic Imaging SPIE* Vol. 649014, 2007.
- [4] G. Yi, H-C. Yeh, G. Vanmeerbeeck, and Ke Zhang, "Demo: Real-Time Depth Extraction and Viewpoint Interpolation on FPGA," in *Proceedings of 2011 ICDSC*, pp. 1-4, Aug., 2011.
- [5] <http://www.hhi.fraunhofer.de/en/departments/image-processing/applications/real-time-stereo-to-multiview-conversion/>
- [6] Harris and M. Stephens, "A combined corner and edge detector," in *Proceedings of the 4th Alvey Vision Conference*. pp. 147- 151, 1988.
- [7] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision (2nd ed.)*. Cambridge University Press, 2004.
- [8] Y. Oh and Hong Jeong, "Trellis-based Parallel Stereo Matching," in *Proceedings of IEEE conference on Acoustics, Speech, and Signal Processing*, vol. 4, pp.2143-2146, 2000.
- [9] M. Mueller, F. Zilly and P. Kauff, "Adaptive Cross-Trilateral Depth Map Filtering," in *Proceedings of IEEE 3DTV conference: The True Vision - Capture, Transmission and Display of 3D video*, pp.1-4, 2010.