

# 적응적 시차 추정기법을 이용한 웹 기반의 원격 다시점 3D 화상 통신 시스템의 구현

정회원 고정환\*, 김은수\*

## Implementation of Web-based Remote Multi-View 3D Imaging Communication System Using Adaptive Disparity Estimation Scheme

Jung-Hwan Ko\*, Eun-Soo Kim\* *Regular Members*

### 요 약

본 논문에서는 실시간으로 입력되는 스테레오 영상에 효과적인 시차 추정 기법을 적용한 새로운 웹 기반의 원격 다시점 3D 화상 통신 시스템을 구현하였다. 제시된 방법에서는 먼저, IEEE 1394 방식의 스테레오 카메라를 이용하여 자체 설계된 Intel Xeon 서버 시스템을 통해 실시간으로 입력되며, 입력된 스테레오 영상을 중심으로 특징점의 크기에 따라 세밀하면서도 정확한 시차 지도를 검출한 뒤, 검출된 시차지도와 좌 영상을 Directshow SDK를 이용하여 네트워크 채널로 연결된 클라이언트 영역으로 전송하게 된다. 전송된 클라이언트 영역에서는 수신된 시차지도와 좌 영상을 중심으로 정확한 우 영상을 복원한 뒤, 중간시점 합성 기법을 이용하여 16시점의 다시점 입체 화상 통신을 이루게 된다.

각기 다른 상황에서 촬영된 'Joo' 영상과 'Hoon' 영상을 웹 기반의 실시간 전송 및 중간영상 합성 실험결과, 본 논문에서 제안된 전송기법으로 우 영상을 복원할 경우 각각 30dB, 27dB의 높은 PSNR이 측정되었으며, 4시점의 8비트 영상을 전송할 경우, 각각 67.2ms의 속도로 디스플레이 함으로써 이를 통한 새로운 실시간 웹 기반의 원격 3D 화상 통신 시스템의 실질적인 구현 가능성을 제시하였다.

Key Words : 3D Communication, Disparity estimation, stereo camera

### ABSTRACT

In this paper, a new web-based remote 3D imaging communication system employing an adaptive matching algorithm is suggested. In the proposed method, feature values are extracted from the stereo image pair through estimation of the disparity and similarities between each pixel of the stereo image. And then, the matching window size for disparity estimation is adaptively selected depending on the magnitude of this feature value. Finally, the detected disparity map and the left image is transmitted into the client region through the network channel. And then, in the client region, right image is reconstructed and intermediate views be synthesized by a linear combination of the left and right images using interpolation in real-time. From some experiments on web based-transmission in real-time and synthesis of the intermediate views by using two kinds of stereo images of 'Joo' & 'Hoon' captured by real camera, it is analyzed that PSNRs of the intermediate views reconstructed by using the proposed transmission scheme are highly measured by 30dB for 'Joo', 27dB for 'Hoon' and the delay time required to obtain the intermediate image of 4 view is also kept to be very fast value of 67.2ms on average, respectively.

※ 본 연구는 정보통신부 및 정보통신연구진흥원의 대학 IT연구센터 지원사업의 연구결과로 수행되었습니다.(IITA-2005-C1090-0502-0038)

\* 광운대학교 차세대 3D 디스플레이 연구센터 (misagie@kw.ac.kr, eskim@daisy.kw.ac.kr)

논문번호 : KICS2005-11-460, 접수일자 : 2005년 11월 14일

## I. 서 론

최근, 보다 자연스럽고 현실감 있는 비전 시스템의 요구가 증가함에 따라 3차원 영상 디스플레이 및 전송 기술에 대한 많은 연구들이 진행되고 있다<sup>[1-3]</sup>. 특히, 심해저, 우주 공간 및 고 방사선 구역 등의 극한 작업 환경이나 원격 화상회의 시스템, 원격 의료서비스, 무인감시 시스템, 공장 자동화, 지능형 교통 시스템 등 다양한 산업 분야에서 기존의 2차원 영상시스템의 한계를 넘어설 수 있는 3차원 영상 통신시스템에 대한 필요성이 날로 증가되고 있으며, 일부는 이미 실용화가 추진되고 있는 실정이다. 특히 원격지 상호간의 동일 시간, 동일 장소에서 회의하는 것과 같은 효과를 주는 원격 입체 화상회의 시스템의 경우, 보다 현실감 있고 입장감 있는 회의 환경을 조성할 수 있다는 측면에서 그 관심이 더욱 고조되고 있으며, 이에 대한 활발한 연구가 진행되고 있다. 이러한 원격 3차원 화상 통신시스템을 구축하기 위해서는 다시점의 3차원 영상 획득 및 정확한 실시간 전송, 그리고 이를 현장감 있게 표현해 줄 수 있는 디스플레이에 이르기까지 고난이도의 방법과 기술들이 요구된다. 그러나 현재 보급 중인 원격 화상 통신 시스템의 경우, 대형 디스플레이, 영상/음성 처리 장치, 미팅룸 구성, 멀티캐스트 지원 및 네트워크와 같은 확장성 있는 하드웨어 시스템의 요구와 이를 뒷받침하기 위해 상호호환될 소프트웨어 개발의 한계가 있으며, 영상 전송에 있어서도 Vic에 근거한 구형 압축방식에 의존하여 300kbps 수준의 2차원 영상만을 전송할 수 있는 문제점을 안고 있다. 또한, 기존의 스테레오 영상의 화상신호에 기반을 두어 개발된 대부분의 3차원 화상 전송기술은 화질과 속도, 전용장비 및 회선 관리에 따른 네트워크 운영에 대한 추가적인 비용 등에 능동적으로 대처하는데 많은 어려움이 있다. 더욱이, 현재의 전송기술로서 필요로 하는 대역폭을 3차원 화상전송의 가능성에 대해 고려해 볼 때, 전송매체 및 설비 등의 문제점이 나타나게 되며, 지상에 있어서 무선전송 설비에 대한 주파수 할당이 진행되고 있고 3차원 화상전송을 위한 전용 통신로의 확보가 어려운 문제점이 나타나게 된다<sup>[2]</sup>. 또한, 기존의 3D 입체 화상 시스템은 다안식 입체 카메라를 통해 다시점의 영상획득과 디스플레이가 이루어지기 때문에 시점의 개수가 증가함에 따라 시역이 확대되고 보다 자연스러운 3D 디스플레이가 가능하지만, 시점의 개수가 증가됨에 따라 영상 데이터도 동

시에 증가하기 때문에 실시간적 영상처리 프로세서 및 초고속, 광대역의 전송채널이 또한 요구되게 된다<sup>[2]</sup>.

따라서 본 논문에서는 IEEE 1394 방식의 스테레오 카메라를 이용하여 자체 설계된 Intel Xeon 서버 시스템을 통해 실시간으로 입력되며, 입력된 스테레오 영상을 중심으로 특징점의 크기에 따라 세밀하면서도 정확한 시차 지도를 검출한 뒤<sup>[4]</sup>, 좌영상과 상대적으로 데이터양이 적은 시차정보만을 마이크로소프트 Directshow SDK<sup>[5]</sup>를 기반으로 개발된 제어 프로그램을 이용하여 네트워크를 통해 전송함으로써 전체적인 데이터양을 줄일 수 있고, 전송된 클라이언트 영역에서는 수신된 시차지도와 좌영상을 중심으로 정확한 우 영상을 복원을 한 뒤, 최근 3D 입체 디스플레이 시스템의 새로운 접근방법으로 제안된 중간시점 합성 기법<sup>[3]</sup>을 이용하여 다시점 입체 화상 통신을 이루게 된다. 그리고, 전후배경이 틀린 각기 다른 상황하의 2사람을 대상으로 촬영된 테스트용 스테레오 입력영상에 대해 카메라 촬영에서 전송, 디스플레이에 필요한 지연시간을 측정, 분석하고 서버측으로부터 전송받은 클라이언트 영역에서 복원된 우 영상과 원본 우 영상간의 PSNR 비교를 통해 제안된 시스템의 성능을 분석함으로써 본 논문에서 제안한 실시간적 원격 다시점 3D 화상 통신 시스템의 실질적인 구현 가능성을 제시하고자 한다. 각기 다른 상황에서 촬영된 'Joo' 영상과 'Hoon'영상을 웹 기반의 실시간 전송 및 중간영상 합성 실험결과, 본 논문에서 제안된 전송기법으로 우 영상을 복원할 경우 각각 30dB, 27dB의 높은 PSNR이 측정되었으며, 4시점과 16시점의 8비트 영상을 전송할 경우, 각각 67.2ms, 278.8ms의 속도로 디스플레이 함으로써 이를 통한 새로운 실시간 웹 기반의 원격 3D 화상 통신 시스템의 실질적인 구현 가능성을 제시하였다.

## II. 제안된 웹 기반의 다시점 3D 화상 시스템

실감형 원격 다시점 3D 화상 시스템을 구현 위해서는 우선적으로 영상을 획득하기 위한 카메라 시스템과 스테레오 영상간의 시차추출 및 웹기반의 안정적이고 효율적인 전송을 위한 서버시스템, 그리고 원격지에서 서버에 접속하여 3차원 디스플레이를 하게 될 클라이언트 시스템이 필요하다. 그림 1은 본 논문에서 새로이 제안한 웹기반 실시간 3D 화상 시스템의 흐름도를 나타낸 것으로 크게 2부분으로 이루어진다.

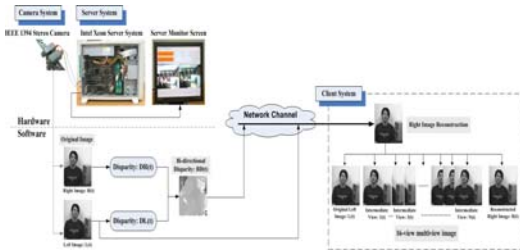


그림 1. 제안된 웹기반 실시간 3D 화상 시스템의 흐름도  
Fig. 1. Flowchart of the proposed web-based real-time 3D Imaging Communication System

즉, 스테레오 카메라로부터 영상 획득 및 전송을 담당하는 서버 시스템에서는 실시간으로 영상을 입력받아 특징값을 추출하고, 특징값의 크기에 따라 보다 효과적으로 좌, 우 영상간의 시차정보를 추출한 뒤<sup>[3]</sup>, 원본 좌 영상과 시차지도(disparity map)를 웹 기반의 네트워크 채널(network channel)을 통해 전송하게 된다. 원격지에서 네트워크를 통해 접속된 서버 시스템으로부터 원본 좌 영상과 그에 따른 시차지도를 전송받은 클라이언트 시스템은 이로부터 시점에 따른 우 영상을 복원한 후, 시점에 따라 연속적으로 중간 시점의 영상이 생성되며, 복원된 다시점의 3D 입체영상은 3D 디스플레이 장치에 전송되어 자연스러운 입체화상을 구현하게 된다.

### 2.1 스테레오 영상 처리용 서버 시스템(server system)

일반적으로 스테레오 카메라 시스템에서 실시간 영상획득과 알고리즘 프로세싱을 위해서는 방대한 데이터양이 발생한다. 이러한 데이터양을 실시간 구동하기 위해서는 컴퓨터 시스템의 성능이 매우 중요하며 여기에서 CPU와 메모리, 그리고 데이터 버스의 전송 대역폭이 성능 결정의 주된 요인으로 작용하게 된다.

그림 2는 방대한 양의 데이터를 실시간으로 전송하기 위해 IEEEa 1394 스테레오 카메라로부터 입력받은 그래픽 어댑터와 이를 통해 입력된 스테레오 영상을 처리하기 위한 인텔 Xeon 프로세서로 구성된 서버 시스템을 나타낸 것이다. 그림 2에서, CPU는 인텔사에서 개발된 Xeon 프로세서 2.4GHz를 듀얼(dual) 모드로 설치하였으며, 두개의 512M DDR 메모리와 SE7501HGZ의 서버보드 및 SCSI Ultra320 HDD의 주 저장 장치를 이용하여 고속의 스테레오 영상 처리를 수행하였다. 특히, 서버보드로 사용된 SE7501HGZ의 경우, 32 bit/33 MHz PCI, 2개의 64 bit/100 MHz PCI-X, 1개의 64 bit/

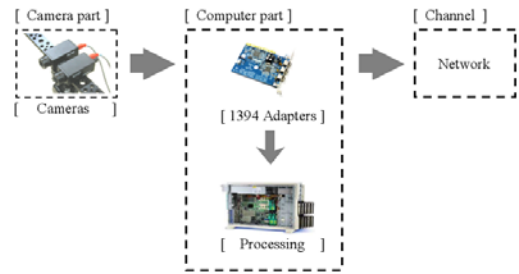


그림 2. 고속 스테레오 영상처리를 위해 구현된 서버시스템  
Fig. 2. The implemented server system for the fast stereo image processing

133 MHz PCI-X와 같이 3개의 PCI 버스를 통해, 각각 독립적으로 동작하며 상호간의 데이터 전송에는 영향을 주지 않는다. 또한, 제안된 시스템에서는 카메라 시스템을 통해 총 26.3671875MB(640(W)×480(H)×24(bits)×15(frame/sec)×2(cameras))의 데이터가 발생하지만, 본 논문에서는 256×256의 영상을 실험에 이용하였으므로, 1.875 MB(256(W)×256(H)×8(bits)×15(frame/sec)×2(cameras))의 데이터가 실시간으로 알고리즘에 적용되며 네트워크를 통해 수신측으로 전송될 수 있다.

### 2.2 16시점 3D 디스플레이를 위한 효율적인 스테레오 영상 전송

본 논문에서는 데이터양의 축소와 함께 보다 효율적인 영상 전송을 위해 좌측 영상과 평행식 스테레오 카메라를 통해 검출된 시차지도만을 전송하게 된다. 스테레오 입력영상으로부터 시차를 검출하기 위해 사용된 비용함수로는 계산량은 늘어나지만 보다 정확한 시차 추출을 위해 MSE (mean square error)를 사용하였다. 식 (1)은 스테레오 영상의 시차를 검출하기 위해 사용된 MSE 함수를 나타낸 것으로, 식 (1)에서  $N_x \times N_y$ 는 블록의 크기,  $I_L$ 과  $I_R$ 은 좌측영상 블록과 이에 대응하는 우측영상 블록을 각각 나타낸 것이다.

$$MSE = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} |I_L(i, j) - I_R(i+d, j+l)|^2 \quad (1)$$

일반적으로, 인간시각 시스템에서 수직시차는 수평시차에 비해 그 영향이 적기 때문에<sup>[2]</sup>, 스테레오 카메라를 평행식 카메라로 구성하면 그림 3와 같이 수평 성분만으로 시차 벡터를 찾을 수 있다. 그림 3은 입력된 스테레오 영상의 에피폴라(epipolar) 선상에서 좌영상과 우영상간에 서로 대응되는 블록을

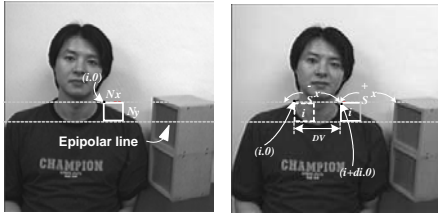


그림 3. 에피폴라 선상의 대응점 탐색  
Fig. 3. Correspondence point search on the epipolar line

찾는 과정을 나타낸 것으로, 좌영상을 일정한 크기의 블록으로 나눈 후 임의의 블록에 대응하는 같은 에피폴라 선상에 존재하는 우영상에서의 일치점을 일정한 탐색범위 ( $\pm S$ )내에서 찾는 과정을 보여준 것이다.

식 (2)는 그림 3에서 표현된 탐색 범위를 나타낸 것으로,  $\pm S$ 는 우영상에서 대응되는 블록 또는 화소를 찾는 탐색범위를 나타내고,  $x_0$ 와  $y_0$ 는 적용적시차 추정시 좌영상에서의 블록의 시작점을 각각 나타낸 것이다.

$$S = [R(x, y) \mid (x_0 - S_x \leq x \leq x_0 + S_x + N_x), (y_0 \leq y \leq y_0 + N_x)] \quad (2)$$

따라서 탐색범위  $\pm S$  내에서 MSE를 최소로 하는 블록을 좌영상과 정합된 우영상의 블록으로 선택하게 되면, 우영상과 좌영상간의 좌표값 차이로 그 블록을 대표하는 시차값을 얻게 된다. 즉, 좌측영상을 기준으로 한 우측영상과의 시차는 식 (3)과 같이 주어지게 된다.

$$\hat{a}_{ij} = \arg \left[ \min_{(i, j) \in \beta_i} \frac{1}{N_x N_y} |I_L(i, j) - I_R(i + d, j)|^2 \right] \quad (3)$$

여기서,  $I_R$ 은 우영상을,  $I_L$ 는 좌영상을,  $\hat{a}_{ij}$ 는 탐색범위 내의 시차값을 각각 나타내며,  $\beta_i$ 는  $I_R$ 의  $i$ 번째  $N_x \times N_y$  크기의 블록을 나타낸다. 식 (3)에서 얻어지는 시차벡터는 블록을 대표하는 값으로 시차 정보를 추출하는데 사용된다.

한편, 네트워크를 통한 영상 전송을 위해 선행될 영상 압축에서는 ISO/IEC MPEG-4, ITU-T H.263 등 고성능 영상 부호화 기법들이 소프트웨어로 지원되어야 하지만, 이들의 광대역 스테레오 영상의 실시간 소프트웨어 부호화 기법은 현재의 일반적인

PC 프로세싱으로는 한계가 있다. 따라서 본 논문에서는 데이터양의 축소와 함께 보다 효율적인 영상 전송을 위해 선도망의 대역폭 한계인 155Mbps를 맞추어 전처리 필터링을 포함한 영상 샘플링과 이의 부수적인 처리 과정들을 이용한 간단한 압축과정을 이용하였다. 먼저 스테레오 카메라로부터 획득된 최대 15Mbps의 처리 속도를 가진 좌영상과 시차지도를 네트워크와 시스템의 대역폭 및 네트워크 제약에 맞도록 처리하고 압축된 영상 스트림은 네트워크 모듈로 전달되어 프레임 단위로 패킷화 된다. 또한, 영상 획득 및 처리와 전송 과정의 독립을 위해서 더블버퍼(double buffer)를 사용하였으며, RTP/UDP 기반의 표준화된 전송으로 정리될 패킷화 부분은 MTU(maximum transfer unit)를 참조하여 적당한 패킷 크기로 분할되며, 자체적으로 정의된 순서 번호는 전송순서에 따라 순차적으로 증가하게 된다.

### 2.3 16시점 3D 디스플레이를 위한 클라이언트 시스템 (client system)

서버 시스템으로 부터 전송되어온 좌영상과 시차정보는 접속된 클라이언트 시스템의 중간영상합성 알고리즘에 따라 우영상으로 복원된 뒤, 연쇄적인 중간시점의 영상을 생성함으로써 보다 자연스러운 다시점 3D 영상 통신이 이루어지게 된다. 또한, 송신측으로부터 전송 받은 패킷은 그 손실 여부를 검사한 후 해당 신뢰성 기법에 따라 처리한 다음, 프레임의 단위로 병합하여 스테레오 영상 재현을 위한 모듈로 전달하게 된다. 따라서, 전송 받은 영상 스트림은 오류 은닉 기법을 포함하는 복원단계를 거쳐서 클라이언트 시스템의 디스플레이 환경에 맞는 포맷으로 변환된 후 디스플레이 하게 된다.

그림 4는 본 논문에서 제안한 16시점 3D 디스플레이 화상 통신 시스템을 구현하기 위해 이용된 중

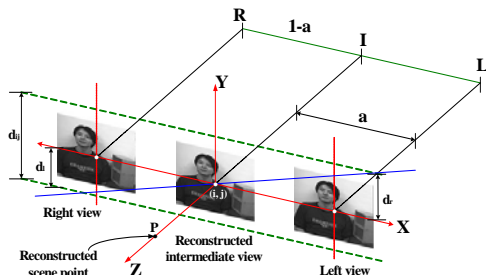


그림 4. 좌우 영상으로부터 획득한 중간영상의 대응점  
Fig. 4. Corresponding point of intermediate view image from left and right image



간시점 영상합성의 개념도를 나타낸 것이다. 그림 4에서 I축은 중간영상 합성 개념에서의 선택된 축의 위치를 보여주고 있으며, 중간영상 'I'의 면을 통해서 좌영상면 'L'에서 우영상면 'R'까지의 대응 위치 관계를 보여준 것이다. 또한, 스테레오 입력영상으로부터 중간영상을 합성할 때 시점의 위치는  $\alpha$ 로 나타내며,  $\alpha$ 는 0과 1사이의 값으로 설정되게 된다.

여기서, 좌 영상을 기준으로 할 때,  $\alpha=0$ 인 영상은 좌 영상을,  $\alpha=1$ 인 영상은 우 영상을 각각 나타내게 되며, 중간값에 해당하는 영역이 중간영상에 해당된다.

식 (4)는 시점의 위치  $\alpha$ 에 따라 가중 평균값으로 내삽하는 경우를 나타낸 식이다.

$$I_p(i, j) = (1 - \alpha) \cdot I_L(i - \hat{d}_{ij}(i, j), j) + \alpha \cdot I_R(i - \hat{d}_{ij}(i, j), j) \quad (4)$$

식 (4)에서,  $\hat{d}_{ij}$ 는 탐색범위 내의 시차값을,  $I_p$ 는 좌, 우 영상의 시차가 교차되는 지점에 합성된 중간시점 영상의 위치를 각각 나타낸 것이다.

일반적으로, 중간시점 영상을 합성할 때 스테레오 영상쌍의 한쪽 영상에서는 서로 가려진 영역(occluded region)이 존재하게 된다. 이 때, 이러한 가려진 영역에는 시차가 할당되지 못한 영역이 존재할 수 있기 때문에 이러한 영역에 대해서는 시차 평활화 과정을 통하여 주변 시차 값들의 평균값을 구하여 그 시차 값으로 대체하게 된다. 식 (5)는 수평방향의 시차와 좌우 영상간의 관계를 나타내는 식이다.

$$I_R = \begin{bmatrix} i_R \\ j_R \end{bmatrix} = \begin{bmatrix} i_L + \hat{d}(i_L, j_L) \\ j_L \end{bmatrix} = I_L + \begin{bmatrix} \hat{d}(i_L, j_L) \\ 0 \end{bmatrix} \quad (5)$$

따라서 본 논문에서는 스테레오 입력영상의 특성에 따라 적응적 정합 알고리즘을 이용하여 영상을 복원하고 중간시점의 영상을 합성하는 방식을 제안함으로써 기존 방식들의 장점들을 상호보완적으로 사용하여 성능을 개선하였으며, 이를 이용하여 3차원 통신 시스템에서의 데이터 압축과 다시점 3D 디스플레이 시스템을 구현을 통해 보다 자연스러운 디스플레이가 가능한 스테레오 비전 시스템을 제시하였다.

### III. 실험 및 결과분석

본 논문에서 새로이 제안한 실험형 원격 다시점

3D 화상 통신 시스템의 실험 시나리오는 화상회의를 주관하는 각기 다른 조건하의 배경으로 획득된 2사람의 스테레오 영상을 서버와 클라이언트로 구분하여 실시간 전송 및 자연스러운 다시점의 중간영상을 합성하는 것이다. 따라서 이를 위해서는 상호 네트워크 채널을 이용한 전송과정에서 상대적으로 전송량이 큰 영상 데이터의 효율적인 처리 기능과 수신단에서의 보다 자연스러운 3D 시역 생성을 가능하게 해주어야 한다.

즉, 그림 5와 같이 순차적으로 입력되는 좌 영상과 우 영상으로부터 단방향성의 시차지도도를 생성한 뒤, 네트워크 모듈로 전달되어 MTU를 참조하여 적당한 패킷 크기로 분할하고, 영상 획득 및 처리와 전송 과정의 독립을 위해 더블버퍼를 사용하여, 자체적으로 정의된 순서 번호에 따라 전송한 후, 수신단에서는 원본 좌 영상과 시차지도도로부터 다시점의 중간영상을 재구성하게 되며, 이를 통한 자연스러운 3D 화상 통신을 이루게 된다.

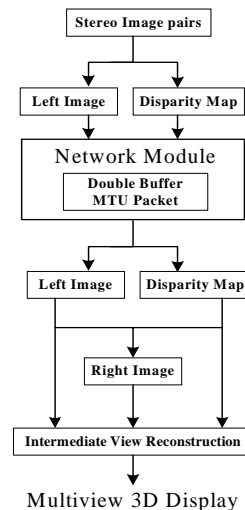


그림 5. 다시점 3D 화상 통신 시스템의 전송과정  
Fig. 5. Flowchart of multiview 3D communication system

#### 3.1 네트워크 채널에서의 전송

현재 기업체나 연구소등에 범용적으로 구축되어 있는 네트워크 표준인 Fast-Ethernet (100Base-T)은 100Mbps의 전송용량을 지원한다. 이것은 범용 네트워크의 수준이며 실제 네트워크상의 노드에 해당하는 중심들은 이보다 훨씬 더 방대한 대역폭을 지원한다. 따라서, 본 논문에서 제안하는 시스템의 초당 데이터 전송량은 영상 압축을 하지 않았을 때 15Mbps(가로256×세로 256×흑백 8비트×좌우영상 2

시점×초당 15프레임)로 근거리 네트워크에서 실시간 전송하기에 충분하며 JPEG이나 ZIP 등의 압축을 사용할 경우 이보다 더 낮은 데이터양이 발생하게 된다. 또한, 본 논문에서는 디지털 카메라 제어와 획득된 영상의 프로세싱, 다시점 영상의 디스플레이를 위해 Microsoft Visual C++.Net and Microsoft DirectX 9.1 software development kit를 이용하였다. Microsoft DirectX SDK는 윈도우 시스템에서 멀티미디어를 지원하기 위한 개발 키트이며 멀티미디어 장치의 데이터획득, 디스플레이를 지원하고, 압축, 해제 등을 지원하게 된다.

3.2 영상 획득 및 전송을 위한 서버 시뮬레이션

본 실험에서 사용한 스테레오 카메라는 그림 2와 같이 2대의 IEEE 1394 포트용 웹 카메라를 평행식으로 하여 기하학적으로 에피폴라 선(Epipolar line)상에 위치하도록 구성하였으며, 두 카메라의 광축간의 거리는 인간의 양안시차 거리인 6.5cm로 설정하였다. 실험에 사용된 Aplus C102T 카메라는 640×480에서 Y400 모드의 8비트 흑백영상에서 초당 15프레임으로 영상을 획득하게 된다.

서버 시스템으로는 2.4GHz 두개의 CPU를 가지며 1GB의 메모리로 구성된 인텔 제온 서버시스템(Intel Xeon Server)를 사용하였으며, 이는 고속처리와 방대한 데이터양을 처리하기에 적합한 모델이다.

그림 6은 본 논문에서 구현된 실시간 원격 3D 화상 시스템의 서버용 구동 시뮬레이터로써, 카메라로부터 영상의 획득과 제어를 위해 Microsoft Directshow 9.0 SDK를 기반으로 Visual C++.Net을 이용하여 다이얼로그 기반에서 구현하였다<sup>[5]</sup>. Microsoft Directshow는 마이크로소프트가 멀티미디어와 게임의 성능향상을 지원하기 위하여 개발한 도구인 DirectX에 포함되어 멀티미디어의 재생과 제어, 그리고 카메라 영상 획득과 관련된 다양한 기능을 지원하고 있는 멀티미디어 도구이다.

그림 6에서 보는바와 같이, 본 논문에서 구현한



그림 6. 네트워크 서버용 구동 시뮬레이터  
Fig. 6. Performance simulator for network server

서버용 시뮬레이터를 구동하게 되면 카메라 2대로부터 초당 15프레임의 속도로 640×480 해상도의 Y400 모드 흑백영상을 획득하게 된다. 획득된 640×480 해상도의 영상에서 초기 설정된 영역들이 매 프레임마다 적응적 시차추정 알고리즘에 적용되어 시차지도가 생성되며, 생성된 시차지도와 원본 좌 영상은 클라이언트가 접속 상태라면 접속된 클라이언트에 전송되고 접속되지 않은 경우에는 데이터를 소멸시킨다.

3.3 시차 정보 검출 실험

본 실험에서는 에피폴라 기하학의 구속조건에 따라 설정된 스테레오 카메라를 이용하여 그림 7과 같이 실제 촬영된 640\*480 픽셀크기의 각기 다른 조건의 배경을 가진 스테레오 영상을 사용하였다.

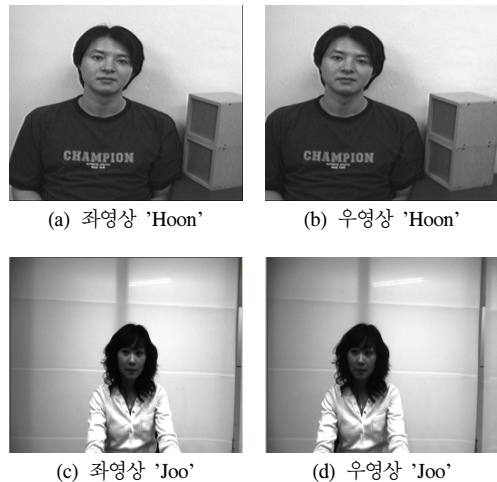
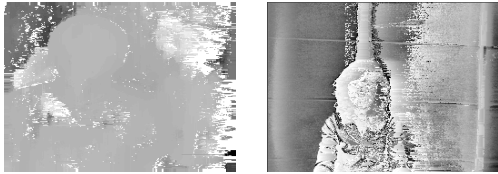


그림 7. 획득된 1st 프레임의 2시점 원본영상  
Fig. 7. Captured 2-view images of 1st frame

즉, 그림 7(a), (b)는 배경에 인물의 그림자(silhouette)와 함께 측면으로 다른 물체가 존재하는 다소 복잡한 영상이고 그림 7(c), (d)는 몇 개의 격자들을 가진 창을 배경이 있는 영상이다.

그림 8은 보다 정확한 결과 분석을 위해 시뮬레이터로부터 입력받은 좌, 우 각각 30 프레임의 순차영상을 획득한 후, 이 중 1st 프레임의 영상들에 대해 본 논문에서 이용한 적응적 시차 추정 기법을 이용하여 검출한 시차 지도를 나타낸 것이다.

그림 8에서 보면, 안정된 시차 벡터가 나타남을 볼 수 있으며, 이러한 안정된 시차 벡터를 통해 성능이 보다 개선된 중간영상의 합성결과를 예측할 수 있다.



(a) 'Hoon' 영상 (b) 'Joo' 영상

그림 8. 적응적 시차 추정 기법에 의한 시차지도  
Fig. 8. The detected disparity maps by using adaptive disparity estimation

### 3.4 다시점 3D 화상 통신을 위한 클라이언트 시뮬레이션

네트워크를 통해 서버로부터 원본 좌 영상과 시차지도를 전송받은 클라이언트 시스템에서는 2.3의 내삽법을 이용한 중간시점 영상합성 기법을 이용하여 다시점의 중간영상을 복원한 뒤, 자체 내장된 다시점 3D 모니터를 통해 실시간으로 화상을 디스플레이하게 된다.

본 실험에 사용된 클라이언트 시스템은 펜티엄 IV 2.6GHz의 512MB의 메모리를 가지는 개인용 컴퓨터이다. 그림 9와 같이 클라이언트용 시뮬레이터를 실행하게 되면 먼저 서버에 접속하게 되며, 이후 초당 15프레임의 속도로 서버측으로부터 전송된 원본 좌 영상과 시차지도를 수신하게 된다. 그러나 클라이언트 시스템이 좌 영상과 시차지도로부터 N 시점의 중간영상을 합성하기까지의 지연시간은 시점을 얼마나 생성하느냐에 따라 변하게 되므로, 생성될 시점의 개수에 따라 초당 처리할 프레임의 수를 설정하도록 하여 실험하였다. 따라서 그림 9에서는 16시점을 생성하기 위한 클라이언트 시스템의 설정으로 281ms의 지연시간이 소요되었으며, 이 시간에는 16시점의 중간영상 합성과 우 영상의 복원시간이 포함되어있으며, 시뮬레이션에서 좌 영상의 경우



그림 9. 3D 화상 통신을 위한 클라이언트 시뮬레이터  
Fig. 9. Client simulator for 3D image communication

$\alpha = 0$ , 우영상의 경우  $\alpha = 1$ 이 되며 중간영상의 1시점부터 16시점까지 얻을 수 있다.

표 1은 본 논문에서 제안한 3D 화상 전송 시스템의 시점별 중간영상 합성 소요시간을 나타낸 것이다.

표 1. 시점별 중간영상 합성시 소요시간

Table 1. Synthesis processing time of intermediate image as view

Experiment	Process time		
	1 view	4 view	16 view
1	30 ms	70 ms	281 ms
2	31 ms	67 ms	276 ms
3	30 ms	63 ms	280 ms
4	30 ms	66 ms	280 ms
5	30 ms	70 ms	277 ms
Average	30.2 ms	67.2 ms	278.8 ms

표 1에서, 중간영상 획득시 발생하는 지연시간이 4시점의 결과에서 초당 15프레임에 해당하는 66ms에 근접하기 때문에 실시간 15프레임의 전송에는 4시점 합성이 현재 제안된 시스템에서의 한계이다. 또한 1시점에서는 지연시간이 30ms로 초당 30프레임의 처리가 가능하며 16시점에서는 초당 3프레임으로 제한된다. 전송되는 영상은 좌 영상과 시차지도이기 때문에 1시점의 중간영상 합성에는  $\alpha = 0.5$ 의 중간영상과 우 영상 복원에 해당하는 2회의 영상합성이 발행하게 된다. 따라서 좌 영상, 중간영상, 우 영상의 3시점을 얻게 되며, 4시점에서는 총 6시점의 영상을 얻게 되고 16시점에서는 총 18시점의 영상을 얻을 수 있다는 분석을 내릴 수 있다.

### 3.5 다시점 영상 합성 및 우영상 복원 실험

그림 10은 전체 30 frame의 입력 영상 중 5th 프레임의 영상에 대해 원본 좌 영상과 적응적 시차 추정 방법에 의해 생성된 중간시점의 영상들 중 5번째 시점(N=5)과 10번째 시점(N=10)의 중간영상과 복원된 우 영상을 각각 나타낸 것이다. 그림 10에서 영상속의 사람이 보이는 시점이 각 영상마다 조금씩 다르며 시각적으로 보기에 별다른 오정합 없이 성공적으로 합성되었음을 알 수 있다.

그림 11은 그림 10과 같은 방법으로 순차적으로 입력되는 30frame의 영상 중 10th 프레임의 중간영상 합성 결과를 보인 것이다.

그림 12는 실험에 사용된 전체 30 frame의 영상 중 1st, 5th, 10th frame에 대해, 순차적으로 입력되는 원본 우영상과 전송 후 시차정보에 의해 복원된 우 영상을 각각 나타낸 것이다.



그림 10. 복원된 5th 프레임의 중간시점 영상  
 Fig. 10. Intermediate view image of the reconstructed 5th frame



그림 11. 복원된 10th 프레임의 중간시점 영상  
 Fig. 11. Intermediate view image of the reconstructed 10th frame

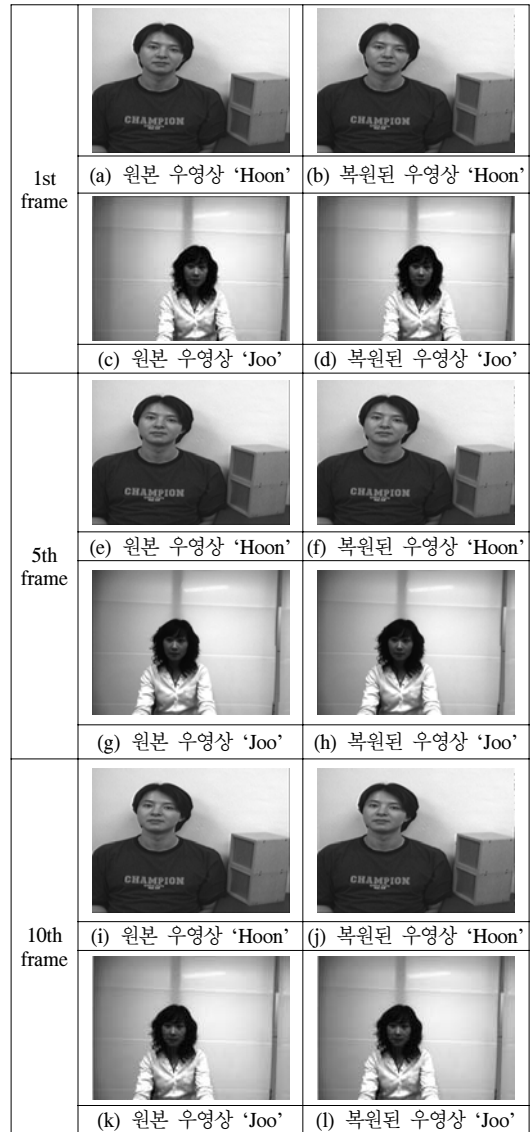


그림 12. 원본 우영상과 복원된 우영상의 비교  
 Fig. 12. Comparison between original right image and reconstructed right image

그림 12에서 보느바와 같이, 원 영상과 복원된 영상은 시각적으로 거의 차이를 없음을 알 수 있다.

표 2는 그림 12에서 나타난 시각적인 결과를 검증하기 위해 원 영상과 복원된 영상간의 유사도 측정을 위한 PSNR의 결과를 나타낸 것이다.

표 2에서 보면, 'Hoon' 영상과 'Joo' 영상의 경우, 전체 30frame의 평균 PSNR이 30.35dB과 27.35dB로 각각 측정되었으며, 더욱이 'Hoon' 영상의 경우, 'Joo' 영상에 비해 배경 물체가 존재하고 있음에도 불구하고 특별한 오류 없이 강건하게 동작하고 있음



표 2. 원본 영상과 복원된 영상간의 PSNR 결과 비교  
 Table 2. PSNR results between original right image and reconstructed right image

Frame	Test image	PSNR	
		'Joo'	'Hoon'
1st		30.47 dB	27.28 dB
5th		30.34 dB	27.33 dB
10th		30.25 dB	27.45 dB
Total average		30.35 dB	27.35 dB

을 알 수 있었다. 한편, 복원된 두 영상의 PSNR 평균값이 서로 다름을 알 수 있는데, 이는 비교적 배경이 단순한 'Joo' 영상에 비해 상대적으로 복잡한 배경이 존재한 'Hoon' 영상의 경우, 더 낮은 PSNR이 검출됨을 알 수 있다. 또한, 본 논문에서 제안된 시스템의 목적이 원격 화상 통신 시스템임을 감안할 때, 측정된 PSNR이 ∞에 근접하는 세밀한 결과가 아닌 시간적으로 큰 부담이 없는 27dB~30dB의 PSNR 이면 원격 화상시스템의 응용에 충분한 결과값으로 분석된다.

그림 13은 상용 8시점 모니터인 SiliconGraphics사의 1600SW를 통해 본 실험을 통해 합성한 8시점 영상을 나타낸 것이다.

SiliconGraphics 1600SW은 편광방식의 3차원 디스플레이 장치로 입체효과를 느끼기 위해서는 디스플레이 장치와 편광이 동일한 편광안경이 필요하다. 안경을 통하여 영상을 관측한 결과 매우 자연스러운 3차원 영상을 볼 수 있었다. 관측 실험에서 물체가 카메라와 너무 가까운 거리에 있을 경우에는 눈이 피로함을 느꼈으며 실험 결과 평균 1미터 이상의 거리를 유지할 때 자연스럽게 편안한 3차원 영상을 느낄 수 있었다.



그림 13. SiliconGraphics사의 1600SW를 통한 8시점 디스플레이  
 Fig. 13. 8-view image for the test image of 'Man' on the commercial 3D monitor of SiliconGraphics

## V. 결론

본 논문에서는 웹 기반의 다시점 3D 화상 통신 시스템이 제안되고 그 성능을 기존의 다른 시차추정 기법과 비교, 분석하였으며, 서버와 클라이언트 영역으로 각각 분리한 원격 시뮬레이션을 통한 실질적인 웹 기반의 3D 화상 통신 시스템을 구현하였다. 제안된 시스템에서는 보다 효율적인 중간영상합성을 위하여 세 가지의 시차추정 기법을 비교하여 그 중 성능이 좋은 것을 선택하였으며 이를 이용하여 웹 기반의 원격 3D 화상 통신 시스템의 구현하였다. 또한 스테레오 카메라를 사용하여 각기 다른 상황으로 촬영한 'Joo' 영상과 'Hoon'영상을 웹 기반의 실시간 전송 및 중간영상 합성 실험결과, 본 논문에서 제안된 전송기법으로 우 영상을 복원할 경우 각각 30dB, 27dB의 높은 PSNR이 측정되었고, 4시점과 16시점의 중간영상을 획득하는데 소요되는 지연시간 역시, 각각 67.2ms, 278.8ms로 나타남으로써 이를 통한 새로운 실시간 웹 기반의 원격 3D 화상 통신 시스템의 실질적인 응용 가능성을 제시하였다.

## 참고 문헌

- [1] V. S. Grinberg, G. Podner, and M. W. Siegel, "Geometry of binocular imaging", Proc. of SPIE, vol. 2177, pp. 56-65, February 1994.
- [2] 김은수/이승현 공역, 3차원 영상의 기초, 技多利, 1998.
- [3] K. H. Bae, J. J Kim, and E. S. Kim, "New disparity estimation scheme based on adaptive matching window for intermediate view reconstruction", Optical Engineering, vol.42, no.6, pp.1778-1786, 2003.
- [4] A. Redert, E. Hendriks, and Jan Biemond, "Correspondence Estimation in Image Pairs", IEEE Signal Processing Magazine, special issue on 3D and stereoscopic visual communication, vol. 16, no.3, pp.29-46, May 1999.
- [5] [http://www.microsoft.com/win\\_dows/directx/](http://www.microsoft.com/win_dows/directx/)
- [6] N. A. Dodgson, J. R. Moore, and S. R. Lang, "Multi-view autostereoscopic 3D display", IBC '99, 10th-14th, pp. 497-502, September 1999.

[7] M. E. Izquierdo, "Stereo matching for enhanced telepresence in three-dimensional video communications", IEEE Transactions on circuits and systems for video technology, vol. 7, no. 4, pp. 629~643, August 1997.

[8] C. H. Yang, "Geometric Models in Stereoscopic Video", Rapport technique de l'INRS-T'el'ecomunications, no. 95-12, 1995.

[9] Z. Zhang, "Determining the epipolar geometry and its uncertainty: A review", Tech. Rep. Technical Report 2927, 1996.

[10] W. E. L. Grimson, "Computational experiments with a feature based stereo algorithm," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.7, no.1, pp.17-34, January 1985.

[11] W. Ho and N. Ahuja, "Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol.11, no.2, pp.121-136, February 1989.

[12] A. Luo and H. Burkhardt, "An intensity-based cooperative bidirectional stereo matching with simultaneous detection of discontinuities and occlusions", SPIE Computer Vision, vol.15, pp.171-188, 1995.

[13] Jeffrey S. McVeigh, M. W. Siegel, and Angel G. Jordan "Intermediate view synthesis considering occluded and ambiguously referenced image regions", Signal Processing: Image Communication, vol. 9, pp. 21-28, February 1996.

[14] 박창주, 고정환, 김은수, "스테레오 영상 보정 알고리즘에 기반한 새로운 중간시점 영상합성 기법", 한국 통신학회 논문지, vol.29, no5C, pp.632~641, 2004.

고 정 환 (Jung-hwan Ko)

정회원



1999년 2월 광운대학교 제어계측공학과 학사

2001년 8월 광운대학교 전자공학과 석사

2005년 2월 광운대학교 전자공학과 박사

2005년 3월~현재 광운대학교

차세대 3D 디스플레이 연구센터 연구교수

<관심분야> 3D 로봇비전, 스테레오 카메라, 영상 통신

김 은 수 (Eun-Soo Kim)

정회원

광운대학교 차세대 3D 디스플레이 연구센터

(<http://3DRC.org>)

한국통신학회지 Vol.29, No. 4c 참조