

양안 입체 TV 를 위한 가상 광축 조절

한상효, 박종일, 엄기문*, 안충현*
한양대학교 전자통신전파공학과
*한국전자통신연구원 방송시스템연구부

Virtual Control of Optical Axis of Stereo TV

Sang Hyo Han, Jong-Il Park, Gi Mun Um*, Chung Hyun Ahn*
Div. ECE, Hanyang University

*Broadcasting System Research Dept., Electronics and Telecommunications Research Institute
E-mail : {guy_cody, jipark}@mr.hanyang.ac.kr, {gmum, hyun}@etri.re.kr

Abstract

양안 입체 TV 에서는 3 차원 카메라의 시점거리에 따라 눈의 편안함과 3 차원 효과 사이에 대차 관계가 존재한다. 일반적으로 카메라 사이의 거리가 인간의 동공 사이의 거리인 65mm 일 때 최적의 효과를 낸다고 한다[1]. 그러나 일반적인 방송용 화질의 상용 카메라는 렌즈 크기가 크기 때문에 이 거리를 맞춘 3 차원 카메라를 제작하기는 매우 어렵다. 그래서 보기 편한 양안 입체 영상을 제공하기 위해서는 스테레오 카메라에서 얻은 영상을 조작하여 원하는 가상 시점에서의 영상으로 만들어 줄 필요가 있다. 본 논문에서는 먼저 적응적 다중 창틀 정합을 이용한 계층적 변이 추정을 사용하여 변이 지도를 추출하고, 이것을 이용하여 가상 시점에서의 영상을 합성했다. 다양한 스테레오 영상을 이용한 실험을 통해 제안된 기법의 타당성을 확인하였다.

I. 서론

최근 보다 실감나는 영상에 대한 수요가 증대하여 3 차원 입체 TV 에 관한 연구가 활발히 이루어지고 있다[2]. 본 논문에서는 양안 입체 TV 의 광축을 조절하여 가상 시점에서의 영상을 합성하는 시스템을 제안하였다. 이 시스템을 이용해 원하는 거리-인간의 동공 사이의 거리인 65mm-로 양안 입체 TV 의 광축을 가상으로 조

절함으로써 사람의 눈에 보다 편안한 영상을 만들어 내 고자 하였다.

3 차원 카메라의 렌즈 크기가 크기 때문에 광축 사이의 거리를 65mm 로 맞출 수 없다. 이러한 카메라로 얻은 영상들은 puppet theater effect 와 같은 현상을 나타 낸다[3]. 그러므로 보기 편안한 양안 입체 영상을 만들기 위해 스테레오 카메라에서 얻은 영상을 조작하여 원하는 가상 시점에서의 영상으로 만들어 줄 필요가 있다. 이것을 양안 입체 TV 광축 조절이라 하고, 그림 1 은 제안된 시스템의 개념을 나타낸다.

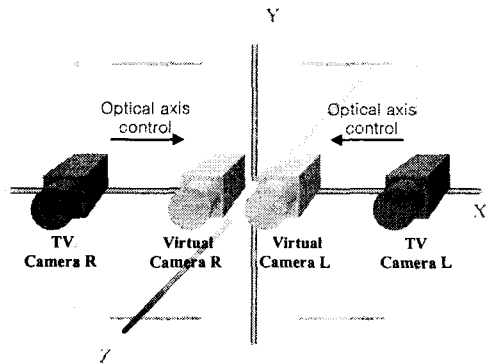


그림 1. 3 차원 TV 카메라의 가상 광축 조절

II. 광축 조절 시스템

제안된 광축 조절 시스템은 적응적 다중 창틀을

이용한 계층적 변이 추정과, 가상 시점 영상 합성의 두 가지 단계로 구성되어 있다.

2.1 적응적 다중 창틀을 이용한 계층적 변이 추정

스테레오 카메라로 얻은 영상을 가상 시점으로 조절하면 원 영상에서는 보이지만 가상 시점 영상에서는 보이지 않는 부분이 생길 수 있으며, 반대로 없던 부분이 새로이 생겨나는 부분들이 생긴다. 따라서 광축 조절된 임의시점 영상을 생성하기 위해 단순히 화소를 좌우로 움직이는 것만으로는 3 차원적 입체감을 줄 수 없다.

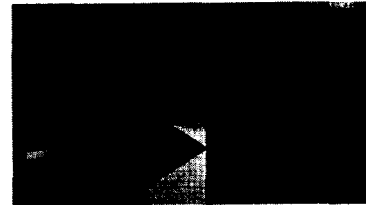
그러므로 이러한 문제를 해결하면서 변이 지도를 추출하는 과정이 필요하게 된다. 변이 지도 추출 과정은 많은 계산량을 필요로 하는 과정으로, 빠르면서도 정확한 변이 지도를 구하기 위해 계층적 변이 추정 방법을 사용하였다.

원 영상을 Gaussian 저역 통과 필터와 Decimation을 통해 여러 계층으로 만든 후, 에피폴라 선 상에서의 전역 탐색 블록 정합을 통해 얻은 최상위 계층의 변이 지도 정보를 이용하여 블록 정합 창틀 크기 조절과 후보 탐색 창틀 조절을 통한 하위 계층에서 적소의 블록 정합을 통하여 빠르면서도 정확한 변이 지도를 구한다.

블록 정합 수행 시에는 그림 2 와 같이 적응적 다중 창틀 정합을 적용하여 여러 창틀들 중 최적의 정합쌍을 선택하여 에지(edge) 부분에서 훨씬 정확한 변이 지도를 구할 수 있다. 그런데 모든 영역에 대해 이러한 방법을 적용하면 계산량이 많아지므로 에지가 존재하는 영역에 대해서만 적응적으로 다중 창틀 정합을 적용한다. 에지검출의 계산량은 정합의 계산량에 비해 훨씬 작으므로 적응적 방법을 도입함으로써 계산량을 많이 줄일 수 있다. 이 과정에서 양 카메라에서 서로 보이지 않는 차폐 영역이라고 불리는 영역을 검출하여 차폐 영역 주위 화소의 변이 정보를 이용한 보간을 수행한다. 왼쪽 영상의 변이값과 이를 이용해 구한 오른쪽 영상의 정합되는 위치에서의 변이값을 비교하여 그 차이가 일정한 문턱값을 넘으면 차폐영역이라고 판단한다. 마찬가지로 오른쪽 영상에 대해서도 같은 과정을 수행한다. 이러한 영역은 앞에 있는 물체에 의해 가려진 배경 영역이라고 볼 수 있으므로 에피폴라 선을 따라 배경의 깊이값을 보간한다. 이와 같은 방법을 통하여 빠르고 정확하게 변이 지도를 얻어낸다.



(a) Left-Shifted Window



(b) Centered Window



(c) Right-Shifted Window

그림 2. 세 종류의 Shifted Windows

2.2 가상 시점 영상 합성

2.1 절에서 얻은 변이 지도 정보를 이용하여 정방향 사상하면 임의시점의 영상을 만들 수 있다. 그런데 임의의 시점 카메라에서 새로이 생겨나는 영역에서는 변이 미 계산 지역이 발생하므로, 그림 3에서의 과정으로 해결한다. 먼저 변이 미 계산 지역 주위 화소의 변이값을 이용하여 에피폴라 선을 따라 선행보간을 행한다. 변이 미 계산 지역의 점 P에 대해 이렇게 얻은 변이값을 이용하여 반대편 카메라의 정합되는 점 P'으로부터 화소값을 역 방향 사상함으로써 적절한 화소값을 구할 수 있다. 단, 이 때 정합되는 점 P'이 차폐영역에 속하지 않는다면 그 카메라에서만 보이는 고유영역이라고 볼 수 없으므로 주변 화소로부터 공간 보간하여 화소값을 구한다. 이와 같이 정방향 사상과 변이 미계산 영역에 대한 역방향 사상과 보간에 의해 임의의 시점 영상 생성을 완성한다.

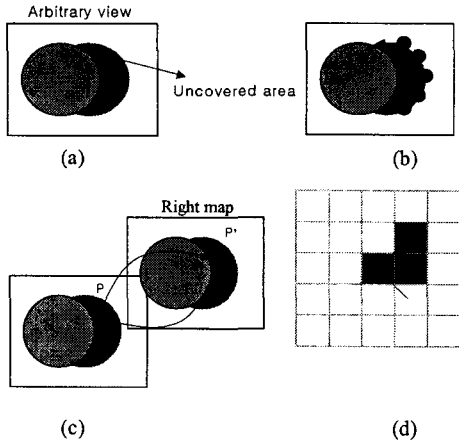


그림 3. 시점 합성 과정

- (a) 정방향 사상, (b) 변이지도의 선형보간, (c) 역방향 사상, (d) 공간적 보간

III. 실험결과

본 논문에서 제안한 기법의 성능을 입증하기 위해 기존 데이터 베이스 영상에서부터 방송용과 같은 화질을 가진 3DTV 카메라로 찍은 실내, 실외 영상에 이르기까지 다양한 양안 입체 영상으로 실험하였다. 시점 거리는 150mm 보다 큰 스테레오 카메라를 사용하였고, Microsoft Visual C++을 사용하여 본 논문에서 제안한 기법을 구현하였다.

본 논문에서 제안한 기법을 정량적으로 평가하기 위해 계산된 변이 지도의 품질을 측정하였다. 평가를 위해 ground truth data(true disparity map : Univ. of Tsukuba)에 대한 error statistics 를 계산하였고, 다음과 같은 quality measure 를 사용하였다.

Percentage of bad matching pixels[4]은 다음과 같이 정의한다.

$$B = \frac{1}{N} \sum_{(x,y)} (|d_{est}(x,y) - d_{true}(x,y)| > \delta_d)$$

여기서 δ_d 는 허용 오차이며, 본 논문에서는 $\delta_d = 1.0$ 을 사용하였다.

그림 4 은 true disparity map 과 제안된 알고리즘을 이용한 estimated disparity map 사이의 오차를 나타내고, 표 1 은 제안된 알고리즘을 이용한 estimated disparity map 의 error percentage 를 나타낸다. 원 영상의 사이즈는 384x288 pixels 이고, true disparity map 과의 비교해 사용된 영역은 상하좌우 18 픽셀이 제거된 348x252 pixels 이다.

그림 4 (a), (b)는 각각 원 영상의 왼쪽, 오른쪽 영상이다. 계층적 블록 정합을 위해 5-layer pyramid 가 구성되었고, 각 레벨에서 정합 창들의 사이즈는 5x5 으로 설정되었다. 그림 4 (c)는 단일 창들을 이용한 계층적 블록 정합에 의해 얻은 변이 지도의 오차를 나타낸다. 검은 픽셀들이 bad matching pixels 이다. 그림 4 (d) 는 적응적 다중 창들을 이용한 계층적 블록 정합에 의해 얻은 변이 지도의 오차를 나타내고, 그림 4 (e)는 edge-adaptive scheme 이 없는 다중 창들을 이용한 계층적 블록 정합에 의해 얻은 변이 지도의 오차를 나타낸다. 그림 4 (c)와 (d) 를 비교해보면 12.4%의 성능향상이 있음을 알 수 있다. 하지만 그림 4 (d)와 (e)를 비교해 보면 눈에 뵈는 차이점을 알 수 없다. 이러한 결과들로부터 논문에서 제안한 적응적 다중 창 정합이 성능향상과 계산량 감소에 타당한 것을 알 수 있다.



(a) Left

(b) Right

(Original Image)



(c) Single Window



(d) Edge-Adaptive Shifted Window



(e) Without Edge-Adaptive Scheme

(True Disparity Map 과 비교한 오차)

그림 4. True Disparity Map 과 비교한 오차

표 1: Percentage of bad matching pixels.

	Error Percentage
Single Window	30.5%
Edge-Adaptive Shifted Window	18.1%
Shifted Window Without Edge-Adaptive Scheme	17.9%

그리고 실험 결과 영상으로 연구실 내에서 주관 평가를 실시하였다. 원 영상(시점 거리=L)과 합성된 영상들(시점 거리=0.8L, 0.6L, 0.4L, 0.2L, 0)은 stereoscopic display 에 표시하였으며, 눈의 편안함, 3D 효과, 영상의 품질, 그리고 원 영상과 합성된 영상의 선호도, 총 4 개 항목에 대해서 5 단계의 평가를 실시하였다.

그림 5는 주관 평가에 사용된 "Dog and Flower" 영상의 원 영상과 합성된 영상이며, 영상의 크기는 720x480 pixels 이고 양안 입체 카메라의 시점 거리 L 은 110mm 이다. 그림 6은 "Dog and Flower"영상에 대한 주관 평가의 평균 점수를 나타낸다. 실험 결과로부터 [1]에서 보고된 바와 같이 시점 거리가 줄어들수록 3D 효과는 줄어드는 반면 눈의 편안함은 증가함을 알 수 있다. 그리고 시점 거리가 줄어들수록 변이 추정 오차에 의한 영상의 품질 저하가 증가함을 알 수 있다. 하지만 영상의 선호도는 0.2L 에서 최고치를 보이는데, 이것은 artifact 가 없는 원 영상보다는 합성된 영상을 더 선호한다는 것을 보여준다. 이 결과로부터 제안된 기법의 효율성을 알 수 있다.

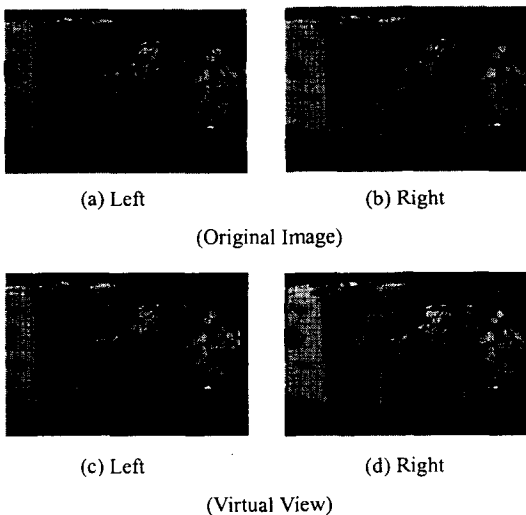


그림 5. "Dog and Flower" 실험 결과

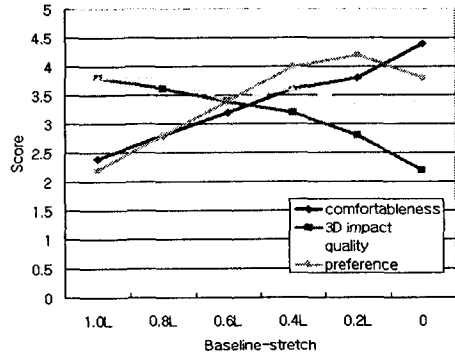


그림 6. "Dog and Flower"영상의 주관 평가 결과

IV. 결론

본 논문에서는 두 카메라 사이의 원하는 위치에서 가상 시점 영상을 합성하는 3DTV 카메라의 가상 광축 조절 기법을 소개하였다. 먼저 적응적 다중 창틀을 이용한 스테레오 정합의 계층적 구현으로 변이 지도를 얻고, 얻어진 변이 지도를 이용하여 가상 시점 영상을 효과적으로 합성하였다. 다양한 스테레오 영상을 이용한 컴퓨터 시뮬레이션을 통해 고화질의 임의 가상 시점 영상을 만들어 낼 수 있음을 확인하였다. 또, 주관평가를 통해 원 영상보다 주관적으로 선호되는 가상 양안 입체 영상을 만들어 낼 수 있음을 보여 제안된 기법의 타당성을 확인하였다.

V. 참고문헌

- [1] L. Stelmach, W. Tam, F. Speranza, A. Vincent, G. Um, "Visual Fatigue Reduction Techniques for Binocular 3D System," *Final Report of CRC/ETRI Collaborative Research*, June 2002.
- [2] ETRI, Test broadcasting of stereo HDTV for 2002 Worldcup Soccer Game, June 2002.
- [3] H.Yamanoue and I.Yuyama, "The relation between size distortion and pick-up conditions for stereoscopic images," *Journal of ITE*, vol. 48, no.10, pp.116-124, Oct. 1994 (in Japanese).
- [4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision* 47(1/2/3), 7-42, 2002.