

## 2차원 동영상의 3차원 입체 변환 방법

□ 김 만 배 / 강원대학교 공과대학 전기전자정보통신공학부 조교수

### I. 머리 말

일반적으로 입체 영상 기술은 2 대의 카메라로 구성된 입체 카메라로 촬영한 입체 영상을 3차원 디스플레이 기기를 이용하여 입체 영상을 보여주는 것이다. 이에 반해 2차원 동영상의 3차원 입체 변환(2D-3D 동영상 입체 변환) 기술은 1 대의 카메라로 촬영한 동영상을 본 논문에서 소개하는 입체 변환 기술을 이용하여 자동적으로 입체 영상을 생성하는 기술이다. 입체 변환 기술은 값비싼 입체 영상의 제작 과정을 거치지 않고, 기존의 단안 영상, 텔레비전, VCR, CD(Compact Disk), DVD(Digital Versatile Disk) 등의 실시간으로 전송되거나 또는 저장되어 있는 2차원 동영상을 입체 영상으로 변환할 수 있는 장점이 있다. 그러나, 입체 변환은 복잡한 영상 처리 및 해석 기술을 필요로 한다. 본 논문에서는 입체

변환 과정에서 필요한 영상 처리에 대해 소개하고자 한다.

2D-3D 동영상 입체 변환은 지난 90년 초부터 관심을 가져왔던 기술로서, 영상 처리 기술의 발달로 점진적으로 발전되어온 분야이다. 그러나, 타 기술처럼 상업용 응용 제품을 보기 힘든 것은 소비자의 요구를 만족할 수 있는 기술 개발을 위해서는 극복해야할 기술적 문제와 하드웨어로 구현했을 때의 복잡도 때문이라고 판단한다. 실제, 2D-3D 입체 변환 기술은 텔레비전, 케이블 TV, VCR 등의 아날로그 기기 및 CD, DVD, 디지털 TV 등의 디지털 기기 등에 응용되어 사용할 수 있는 기술로 응용 범위는 매우 넓다고 할 수 있다.

본 논문의 구성은 다음과 같다. II 장에서는 입체 변환 기술의 기술적 측면에 대해 자세히 설명한다. III 장에서는 최근까지 연구 개발되어 온 2D-3D 변환 기술의 응용 제품을 소개한다.

마지막으로 IV 장에서는 결론 및 향후 연구 방향을 정리한다.

## II. 변환 기술 원리 및 기술

본 장에서는 2차원 동영상의 3차원 입체영상 변환을 가능하게 하는 원리 및 필요 기술에 대해 서술한다. 머리말에서 소개했듯이 2 대의 카메라로 구성된 입체 카메라를 이용하여 얻어진 입체 영상을 직접 디스플레이 기기에 보내주는 입체 영상 처리와는 달리 2D-3D 동영상 입체 변환은 상대적으로 복잡한 원리 및 영상 처리 기술을 필요로 한다.

그림 1은 2D-3D 동영상 입체 변환의 기본 개념을 보여준다. TV, VCR, HDTV, CD, DVD에서 얻어지는 영상은 2차원 영상이다. 각 영상마다 좌영상과 우영상으로 구성되어 있는 입체 영상을 생성하는 것이 2D-3D 동영상 입체 변환의 기본 개념이다. 구현 방법은 크게 세 가지로 분류된다. 첫째, 현재 영상이 좌, 우영상 중 하나가 되면, 다른 하나는 이전 영상(delayed image)들 중에서 1개를 선택한

다[1, 2]. 둘째, 현재 영상은 좌, 우영상 중에 하나가 되고, 다른 영상은 합성영상으로 구현한다[3]. 셋째, 좌영상과 우영상을 모두 합성 영상으로 구현한다[4]. 입체 영상은 양안으로 장면을 볼 때 두 눈에 투시되는 영상이 미묘한 차이가 있듯이, 동일점이 좌영상과 우영상에 시될 때 두 투시점 간의 거리인 시차(parallax)는 깊이감을 결정하는 중요한 요소이다.

다수의 시스템에서 사용하고 있는 기본적인 입체 변환 기술의 기본 원리는 현재 영상과 이전 영상을 이용하여 입체 영상을 만드는 것이다. 그림 2를 이용하여 이 원리를 설명하고자 한다. 영상에서 물체(새)가 우측으로 운동하고 있고, 카메라는 정지된 상태이다. 원영상 #4 프레임은 좌영상으로 하고, 지연영상인 #1-3중 #2 영상을 우영상으로 구성된 입체 영상을 만든 후 양안에 디스플레이하면 새는 모니터 앞으로 튀어나오듯이 보여지고, 배경은 모니터 화면에 디스플레이되어 3차원 입체감을 느끼게

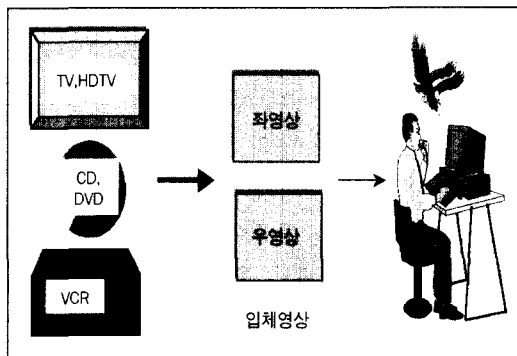


그림 1. 2D-3D 동영상 입체 변환

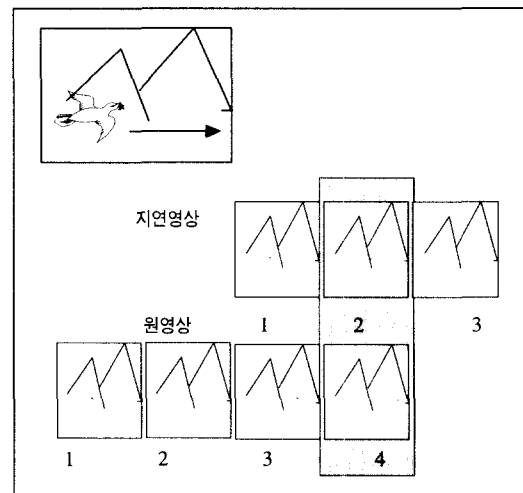


그림 2. 현재 영상과 지연 영상의 구조도

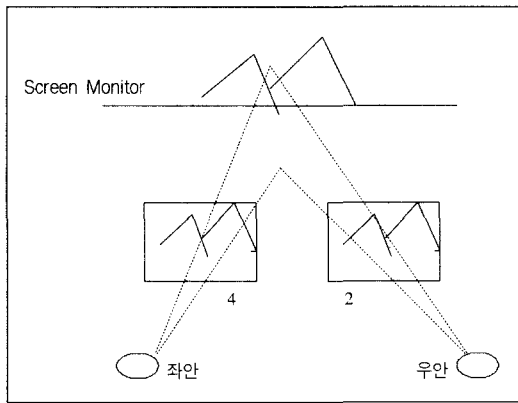


그림 3. 그림 2를 이용한 입체 영상

해준다. 그림 3은 이 과정을 보여준다. 만일 좌, 우영상이 바뀌게 되면 물체는 배경 뒤에 있는 듯이 인식된다. 따라서 인간의 3차원 인식과 상반되는 현상이 발생하므로 눈의 피로감이 발생하게 된다. 따라서 지연영상을 이용하는 입체영상의 생성 과정은 영상내의 물체 및 카메라 운동의 해석이 필요함을 알 수 있다. 전술한 방법은 Ross 현상에 기반을 두고 있기 때문에, 먼저 기본 원리인 Ross 현상에 대해 설명한 후, 영상 운동 유형에 따른 현재 영상과 지연 영상의 선택 방법을 소개하고자 한다.

Ross는 양눈의 시간적 지연이 입체 비전의 단서(cue)라는 사실을 발견하였다[5,6]. 이 현상은 운동하는 물체의 시각 추적을 가능하게 하는 단기 기억(short-term memory)과 밀접한 관계가 있다. 예를 들어 좌측에서 우측으로 날아가는 물체를 추적할 때, 배경 물체들은 양눈에 다른 시간대에 보여진다. 다시 말하면, 움직임을 추적 과정에서 각 눈은 같은 시각영역(visual field)을 다른 시간대에 보고 있는 것이다. 이 단계에서 우리는 물체의 깊이감을 얻을 수 있는데, 이 깊이 인식의 안정성과 연속

성은 단기 기억으로 이해될 수 있다. 배경을 보고 있을 때, 좌안은 리드(lead)하고, 우안은 래그(lag)된다. 반대로 움직임을 보고 있을 때에는 이 현상은 바뀐다. 따라서 좌측에서 우측으로 움직이는 물체의 경우에는 좌안에 현재 영상을 보고, 우안은 래그 또는 지연 영상을 보면, 적절한 시각 시차를 얻을 수 있는데 이 시각 시차는 물체의 운동에 의해 얻어지는 운동시차(motion parallax)에서 변환된 양안시차(binocular parallax)로 불리운다. 양안시차는 깊이감을 결정한다.

Ross 현상은 기본적인 운동 시차의 양안시차의 변환을 실험을 통해 증명하였는데, 실제 응용 과정은 보다 정확한 영상 운동 해석 이론을 필요로 한다. 영상에서는 다양한 물체 및 카메라 운동이 존재하는데, 이러한 운동 유형에 따라 현재 영상과 지연 영상을 좌영상과 우영상 중 어느 쪽으로 디스플레이해 주어야 하는 것은 표 1에서 자세히 설명하고 있다. 카메라가 우측으로 이동하고 있고, 물체도 우측으로 운동하고 있는 운동 유형 ⑤의 경우에는 좌영상은 현재 영상으로, 우영상은 지연 영상이 되어야 한다. 이 과정에서 중요한 것은 운동 유형을 결정하는 알고리즘의 필요성이다. 시차 종류는 음의 시차(negative parallax), 영의 시차(zero parallax), 양의 시차(positive parallax)가 있는데, 입체가 디스플레이 기기의 모니터 바깥에서 느껴지면 음의 시차, 모니터 화면에서 느껴지면 영의 시차, 모니터 안쪽에서 느껴지면 양의 시차라고 한다. 예를 들어 운동유형 ④에서는 움직이는 물체는 음의 시차로 모니터 바깥에서 움직이는 것처럼 느끼고, 배경은 영의 시차로 모니터 화면에서 보여지게 된다.

표 1. 운동 유형과 좌,우영상의 관계 및 시차 현상

운동유형	카메라 운동	물체 운동	좌영상	우영상	시 차 종류
①	우측		지연영상	현재영상	배경은 음의 시차
②	좌측		현재영상	지연영상	배경은 음의 시차
③		우측	현재영상	지연영상	물체는 음의 시차, 배경은 영의 시차
④		좌측	지연영상	현재영상	물체는 음의 시차, 배경은 영의 시차
⑤	우측	우측	현재영상	지연영상	물체는 영의 시차 배경은 양의 시차
⑥	우측	좌측	지연영상	현재영상	물체 및 배경은 음의 시차. 물체는 전면에 보임.
⑦	좌측	좌측	지연영상	현재영상	물체는 영의 시차, 배경은 양의 시차
⑧	좌측	우측	현재영상	지연영상	물체 및 배경 모두 음의 시차. 물체는 전면에 보임.
⑨	우측	좌우	지연영상	현재영상	카메라와 같은 방향으로 진행하는 물체는 영의 시차임, 배경은 후면에, 반대방향으로 진행하는 물체는 배경보다 뒤에 있음.
⑩	좌측	좌우	지연영상	현재영상	카메라와 같은 방향으로 진행하는 물체는 영의 시차임, 배경은 후면에, 반대방향으로 진행하는 물체는 배경보다 뒤에 있음.

그림 4는 이전 영상을 사용하여 얻은 실험 결과이다. Translating Tree 영상은 물체 운동없이 카메라만 좌측으로 이동한다. 따라서 이 영상은 운동 유형 ②이므로 현재영상은 좌영상으로, 이전 영상은 우영상으로 하여야한다. FUN

영상은 카메라 운동 없이 물체만 우측으로 이동하는 영상이다. 따라서 이 영상은 운동유형 ③이므로 현재 영상은 좌영상으로, 이전 영상은 우영상으로 하여야한다. 각 좌, 우영상의 비월영상(interlaced image)은 우측에 보여준다.



그림 4. Translating Tree (위) 영상 Fun (아래영상)

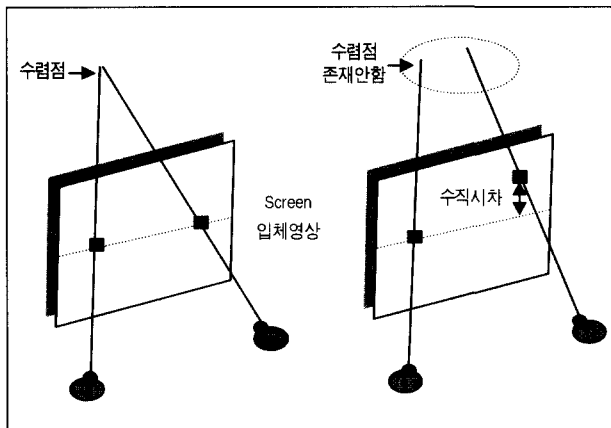


그림 5. 수직시차 및 영상 퓨전

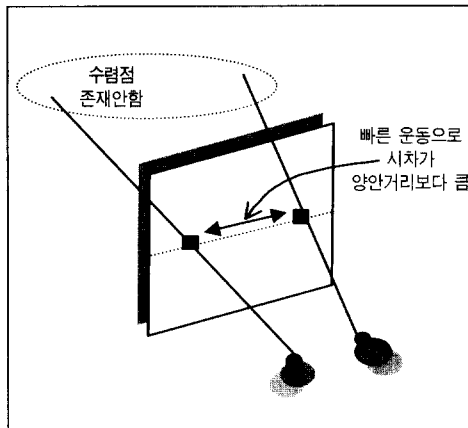


그림 6. 고속 운동의 수렴점

현재 영상을 기준으로 이전 영상 중에서 어떤 영상을 선택하는 문제가 있다. 기본 개념은 운동이 저속인 경우일 때에는 시간적으로 멀리 있는 이전 영상을 선택하면 시차를 크게 할 수 있고, 운동이 고속인 경우에는 가까운 지연영상을 선택하면, 시차를 줄일 수 있다. 수평 운동인 경우에는 표 1을 이용한 입체 변환이 가능하다. 비수평 운동 및 고속 운동 영상인 경우는 다음과 같은 이유로 Ross 현상을 기반으로 하는 입체 변환을 사용할 수 없다.

그림 5, 6은 카메라 및 물체의 운동이 수직 성분을 가지고 있을 때와 고속 운동으로 인한 시차의 불일치 현상을 보여준다. 그림 5는 양눈이 각각 우영상 및 좌영상을 보고 있을 때 영상 퓨전(image fusion) 현상을 보여준다. 왼쪽 그림은 두 매칭점(■로 표시됨)이 수직시차를 가지고 있지 않는 경우로 수평으로 매칭점이 놓여 있다. 두 직선간의 수렴점이 스크린 뒤에 발생하여 정확한 깊이감을 줄 수 있다. 오른쪽 그림에서는 두 매칭점이 수직 시차(↔로 표시됨)를 가지는 경우를 보여준다. 두 매칭점에

수직시차가 존재하므로 수렴점이 존재하지 않는다. 따라서 영상 퓨전이 일어나지 않는다. 이 현상은 입체로 시청할 때에 깊이감의 상실로 눈의 피로감의 원인이 된다. 그림 6은 빠른 카메라 또는 물체 운동으로 매칭점의 시차가 양안 거리보다 크면 수렴점이 없다. 따라서 입체감의 상실로 인한 눈의 피로의 원인이 된다. 따라서 2D-3D 동영상 입체 변환에서는 위에서 언급한 수직 시차 및 고속 운동 처리가 중요한 문제이다.

이 문제를 해결하는 방법으로 다수의 기술이 소개되었다. [4]에서 제안한 3DR(3-D Reconstruction) 방법은 그림 7에서 설명한다. 3차원 데이터는 카메라 투영면으로 원근 투영된다. 카메라가  $(u,v,w)$ 만큼 이동하였을 때에 이전 영상의 한 픽셀  $(X,Y)$ 는  $P(X,Y,Z)$ 의 원근 투영된 투영면의 한 점으로, 이는 현재 영상의  $(X',Y')$ 로 이동한다.  $(X,Y)$ ,  $(X',Y')$  및  $(u,v,w)$ 을 이용하여 영상에서의 각 픽셀의 상대 깊이 값을 얻을 수 있다. 이 깊이 값을 원근투영 기법을 이용하여 입체 영상의 좌, 우영상을 복원

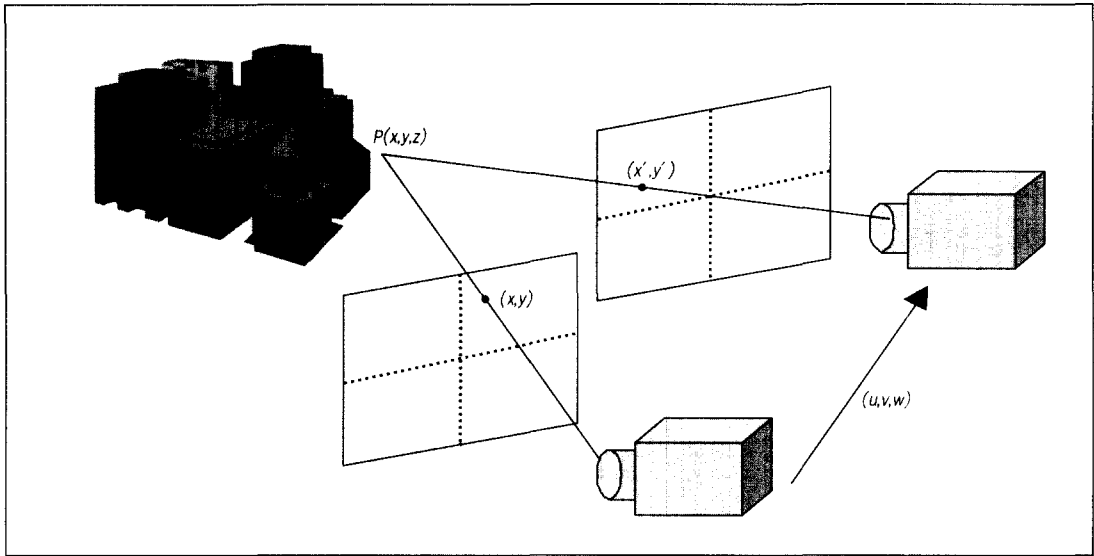


그림 7. 3DR의 깊이 추출

할 수 있다. 따라서 좌, 우영상에는 수직 시차가 존재하지 않기 때문에 비수평 운동 영상의 수직 시차를 극복할 수 있다. [3]에서는 각 블록의 운동 벡터를 계산한 후, 운동 벡터의 크기를 수평 시차로 변환한다. 이 시차 값에 따라 각 블록을 우측으로 수평 이동하여 우영상을 복원한다.

운동이 존재하지 않는 정지 영상의 입체 변환은 운동이 없기 때문에 단안 영상 정보를 이용하여 각 픽셀 또는 블록별 상대적 깊이 정보를 추출하여 입체 영상으로 변환한다. 이론적으로는 인간이 단안 영상에서 깊이를 인식하는 현상은 다음과 같다[7]. 먼데 있는 물체일수록 흐리거나 희미해져서 대조(contrast)가 약하게 보인다는 공기투기(aerial perspective)의 효과나 선과 선의 간격이 좁아지는 도형을 볼 때에 오행감이 생기는 선원근법(linear perspective), 가까울수록 망막에 투영되는 상의 텍스처

(texture)가 치밀하며, 또 많은 상이 조밀하게 투영되는 텍스처 구배(texture gradient)의 효과도 있다. 또 물체에 생기는 명암(light and shade)의 효과도 중요한 실마리가 된다. 뒤쪽의 것은 앞쪽의 것에 의해 그 일부가 가려지는 중복(overlap)의 효과나 적색은 가깝게, 청색은 멀게 느껴진다는 진출 색 및 후퇴 색과 같은 현상도 단안 정보가 될 수 있다. 실제로 일본 산요 전기에서는 평면 화면의 색의 상이, 피사체 분리, 대조에 의한 선명도 차, 원근법에 따른 깊이 등을 조사하여 원근을 판단하여 2차원 정지 영상을 입체 영상으로 변환하는 방법을 개발하였다.

### III. 변환시스템 소개

입체 변환 기술의 최초 등장은 1993년 일본

산요 전기의 2D-3D 변환 TV인데, 상업용 목적으로 개발한 이후로 종종 기술 소개 및 제품이 등장하고 있다. 본 장에서는 먼저 기존의 입체 변환 시스템을 소개한 후, 변환 기술 개발에 필요한 사항들을 소개하고자 한다.

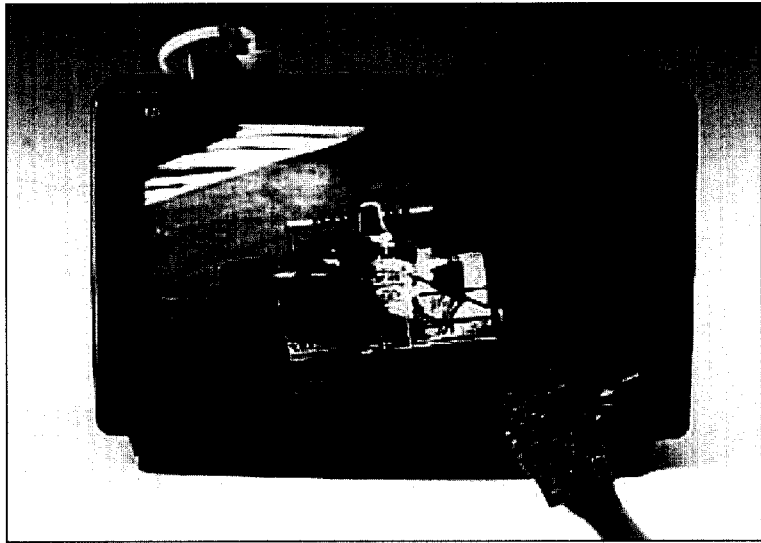


그림 8. 산요 전기의 2D/3D Conversion TV

### 1. 산요 전기

1993년에 일본 산요 전기는 2차원 TV 신호를 3차원 입체 신호로 변환하는 TV를 출시하였다(그림 8)[4]. 2D/3D conversion TV는 low-cost LSI ASIC으로 입체 성능은 우수하지 않았지만, 최초로 상업용 TV에 접목된 것으로 평가받고 있다. 변환 기술은 MTD(Modified Time Difference) 방식으로 불리우는데 현재 영상 프레임과 카메라 및 물체 운동이 존재할 때에 시간적으로 지연된 영상 프레임을 각각 다른 좌, 우눈에 보여 3D 입체 효과를 만드는 것이다. 그림 9는 MTD 방법의 개요도를 보여 준다. 전 장에서 소개한 지연 영상의 사용은 상당히 잘 알려진 기술로써 많은 입체 변환 기술에서는 빈번히 사용되는 기법이다.

1997년에 2차 제품을 출시하였는데, 지연 영상 기법과 함께 저속 운동 영상인 경우에는 영상 운동을 추출한 후 운동 기반 깊이 추출 알고리즘을 이용하여 현재 영상 블록(block)의 깊

이 값을 추출한 후 컴퓨터 그래픽스에 사용하는 원근 투영(perspective projection)을 거쳐 좌, 우영상을 생성하였다. 변환 기술에서 가장

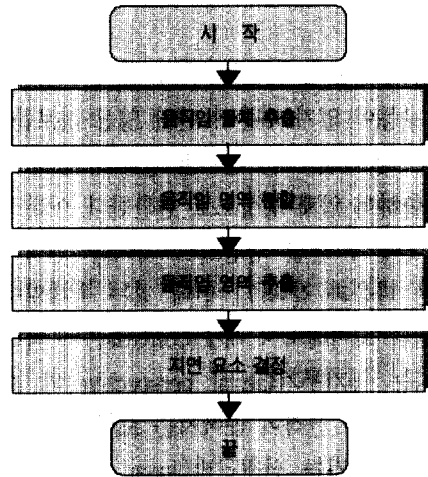


그림 9. MTD 방법의 개요도

큰 문제였던 수직 시차(veritcal parallax)를 해결할 수 있는 장점이 있다. 그러나, 단점은 원근 투영 때문에 영상 왜곡이 발생하여 화질이 약간 낮아지는 단점이 있다. 급격한 운동이 존재하는 경우보다는 카메라 및 물체의 운동이 크지 않을 때에 적용하면 좋은 효과를 얻을 수 있다.

## 2. TransVision

TransVision의 변환 기술은 두 영상에서 픽셀의 상대적 운동을 이용한다[1]. 영상간의 픽셀의 운동 변화를 이용하여 깊이 정보를 구하고 이를 이용하여 각각의 눈에 디스플레이할 영상과 오프셋(offset)으로 최대 시차값을 결정한 후 지연 영상을 선택한다. 이렇게 생성된 동영상은 VCR 테이프에 저장하면 TV 등에 연결하여 입체 영상을 보여줄 수 있다. DSP 하드웨어 보드를 이용하여 의료분야 및 Direct TV에 직접 연결되어 2차원 동영상의 입체 영상 시청 가능하다. 운동이 크지 않을 때에는 입체감이 좋으나, 운동량이 커지면 고스트(ghost) 현상이 나타난다.

## 3. XenoTech

먼저 물체 추적(object tracking)을 위하여 운동 기반 물체 분할(object segmentation)을 이용하여 운동 물체를 추출한 후 각 물체마다 적절하게 입체 영상을 생성한다. 오프라인 처리에서는 기존의 영화 필름 및 동영상의 고화질 입체 영상 제작이 가능하다. 실시간 처리에서는 하드웨어가 입체 변환기에 저장되어 케이블 TV 및 위성 방송 시청을 가능

하게 한다. PC에 플러그인 보드로 장착이 가능하다.

## 4. VISIDEP

VISIDEP(VISual Image Depth Enhancement Process) 방법은 미국 University of South Carolina에서 연구된 기술이다[8]. 항공 사진의 remote sensing 영상 데이터를 입체 영상으로 시청하도록 되어있는 방법이다. Ogly의 실험 결과를 응용하였는데, 양 눈의 교차(alternating) 영상의 시간차가 100ms 이하면 입체 안경을 사용하지 않더라도 입체감을 느낄 수 있다고 한다.

## 5. MPEG 변환기술

상기 변환 기술들은 일반적인 NTSC 신호에 적용이 가능한 것이다. 이에 반해 MPEG(Moving Picture Experts Group)을 이용하는 Video CD, DVD, 디지털 방송에서 응용될 수 있는 입체 변환기술이 연구되어 왔다[9,10]. NTSC 신호 변환에는 반드시 운동 예측(motion estimation)에서 영상 운동 정보를 추출해야한다. 따라서 알고리즘 및 변환 방법의 복잡도가 증가할 뿐만 아니라, 운동 예측의 연산량 때문에 하드웨어가 수반되어야 한다. 이에 반해, MPEG 데이터에는 DCT (Discrete Cosine Transform) 데이터, 운동 정보 등의 압축 데이터를 포함하고 있기 때문에[11], NTSC 신호 처리보다는 보다 간단하게 제작될 수 있다.

MPEG 데이터의 입체 변환은 다음과 같은 전처리 과정을 필요로 한다. 보다 자세한 내용은 [12]를 참조하기 바란다.



- 1) GOP(Group of Picture)를 구성하는 I(intra-coded), P(predictive-coded), B-picture (bidirectionally-predicted coded)의 추출
- 2) 각 픽처의 매크로블록(macroblock)의 운동 벡터(motion vector)의 추출
- 3) 운동 벡터의 스케일링: 픽처간의 간격인 파라미터 M을 이용하여 각 픽처의 운동벡터를 스케일링을 하는 것이 필요하다.
- 4) I-picture의 운동 벡터 생성: I-picture는 JPEG과 유사하게 intra-coded되기 때문에, 이전 P-picture 또는 B-picture를 이용하여 운동벡터의 생성이 필요하다.

#### IV. 결론 및 향후 연구

본 논문에서는 2차원 동영상의 3차원 입체 변환 기술 및 응용에 대해 서술하였다. 이 분야

는 잠재성이 큰 분야이나, 입체감의 향상을 위해서는 보다 우수한 영상 처리 기술을 필요로 하기 때문에, 아직 연구 분야가 많은 미완성 분야이기도 하다.

기술적 완성을 위해서 연구되어야 하는 기술은 1) 영상 운동 해석, 2) 장면 변환 검색(scene change detection), 3) 고속 운동 영상 처리 등이 있다고 할 수 있다. 1)에서는 우수한 영상 해석 방법이 필요하고, 2)에서는 실제 이전 영상을 사용하여 입체 변환을 하는 경우에는 장면 변환이 발생한 현재 영상과 무관한 이전 영상이 사용되므로 심한 flicker 현상이 발생한다. 따라서 장면 변환 검색과 이에 따른 입체 변환 기술이 연구되어야 한다. 3)에서는 물체의 운동 크기 때문에 운동 시차의 양안 시차 변화가 어려운 경우이다.

#### 참고 문헌

- [1] B. J. Garcia, "Approaches to stereoscopic video based on spatio-temporal interpolation," SPIE Photonic West, vol. 2635, pp. 85-95, San Jose, 1990.
- [2] T. Okino, M. Murata, K. Taima, T. Inimura and K. Oketani, "New television with 2D/3D image conversion technologies," SPIE Photonic West, vol. 2653, pp. 96-103, 1995.
- [3] M. B. Kim, M. S. Song, D. K. Kim and K. C. Choi, "Stereo conversion of monoscopic video by the transformation of vertical-to-horizontal disparity," SPIE Photonic West, vol. 3295, pp. 65-74, Jan. 1998.
- [4] Y. Matsumoto, H. Terasaki, K. Sugimoto and T. Arakawa, "Conversion system of monocular image sequence to stereo using motion parallax," SPIE Photonic West, vol. 3012, pp. 108-115, 1997.
- [5] D. C. Burr and J. Ross, "How does binocular delay give information about depth?," Vision Research, vol. 19, pp. 523-532, 1979.
- [6] J. Ross and J. H. Hogben, "Short-term memory in stereopsis," Vision Research, vol. 14, pp. 1195-1201, 1974.
- [7] D. F. McAllister (editor), Stereo computer graphics and other true 3D technologies, Princeton, NJ: Princeton University Press, 1993.
- [8] A. P. McLaurin, et al., "Visual image depth enhancement process: an approach to three-dimensional imaging," DISPLAYS, July 1986.
- [9] M. B. Kim, K. C. Choi, D. K. Kim and M. S. Song, "Conversion of 2D video to stereoscopic video," Proc. 11th Korean Signal Processing Conference, pp. 923-926, Oct. 1998.
- [10] M. B. Kim, J. Y. Park and H. S. Lee, "On the implementation of Ross phenomenon for stereoscopic conversion of 2-D video," Proc. 11th Workshop on Image Processing and Image Understanding, pp. 187-192, Feb. 1999.
- [11] K. R. Rao and J. J. Hwang, Techniques and standards for image, video and audio coding, Prentice Hall, 1996.
- [12] M. B. Kim and S. H. Lee, "A new method for the conversion of MPEG encoded data into stereoscopic video," J. of the Society for 3D Broadcasting and Imaging, No. 1, Vol. 1, pp. 48-59, June 2000.

필자소개



김만배

- 1983 한양대 전자공학과 학사
- 1986 University of Washington 전기공학과 공학석사
- 1992 University of Washington 전기공학과 공학박사
- 1992 - 1998 삼성종합기술원 수석연구원
- 1993 Georgetown University 의과대학 객원연구원
- 1996 University of Rochester 전기공학과 객원연구원
- 1998 - 현재 강원대학교 전기전자정보통신공학부 조교수
- 주관심분야 : 비디오신호처리, 영상통신, 입체영상처리, 의료영상처리