

3D 디스플레이 응용을 위한 시공간 시각주의시스템 적용 사례

- 3D 자유시점 변환 TV를 위한 원격조종기 -

□ 최경주, 박민철 / 충북대학교 전자정보대학, 한국과학기술연구원 포토닉스센서시스템센터

1. 서론

3D TV는 TV 역사에 있어 전환점이 되고 있다. 최근 세계 각국에서는 3차원 입체 방송 시스템에 관한 연구가 활발히 진행되고 있으며, 일부 국가에서는 현재 서비스가 진행 중에 있다. 한국에서는 2002년에 한국전자통신연구원 주도로 한·일 월드컵 입체 방송이 시범 중계되었으며, 중국에서는 2007년 5월 1일부터 3차원 입체오락 TV 채널이 상해에서 정식 운영된다고 보고된 바 있다. 일본의 위성방송회사인 BS 채널 11에서는 2007년 12월 1일부터 일본 내에서 세계 최초로 3D 시험방송을 시작하고, 2008년 4월부터 본 방송을 시작하였다. 각국에서 방송 서비스를 위해 개발된 3차원 입체 방송시스템은 대부분 방송 대역폭을 고려한 양안식과 입체 영상의 선명도를 고려한 안경식 방식을 채택하고 있으며 향후에 다수의 카메라를 이용하여 얻어진 영상을 제공하는

3D 자유시점 변환 TV로 전환될 것으로 예상되고 있다.

자유시점 변환 TV는 다시점 영상을 제공하여 시청자가 임의의 시점에서 방송을 볼 수 있도록 한다. 다시점 영상은 여러 대의 카메라 등의 획득 장치로부터 얻은 복수개의 영상을 이용함으로써 양안시점의 입체 영상과 달리 자유시점 즉 임의의 시점의 영상의 재현을 가능하게 한다. 현재까지 3차원 입체 방송을 위한 많은 연구가 수행 중에 있는데 다시점 영상의 경우, 시청자가 방송을 보면서 자유롭게 시점을 바꾸어 볼 수 있는 장점이 있기 때문에 대다수의 연구가 다시점 영상획득 및 시청자의 시점 변환에 의한 이미지 고속 생성 및 압축에 관한 처리에 초점을 맞추고 있다[1-3].

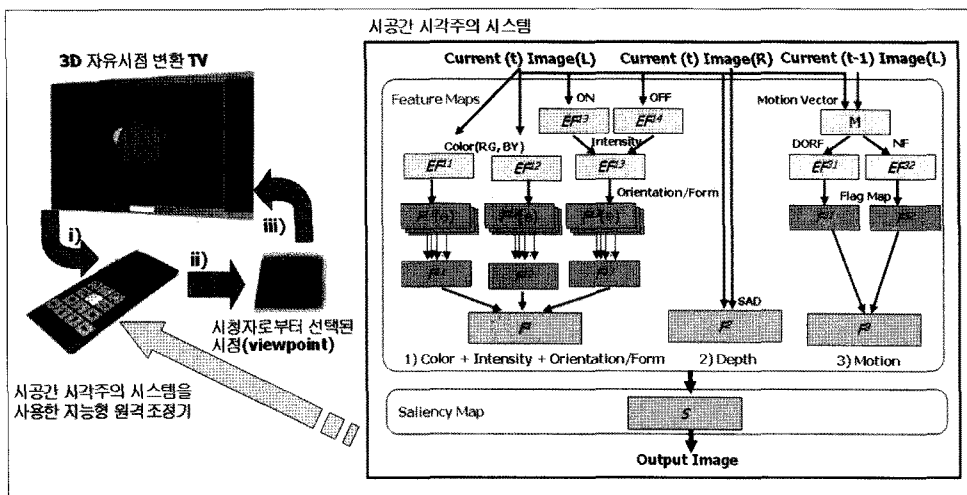
자유시점 변환 TV가 기존 TV와 크게 다른 점은 시청자가 자유시점을 갖게 함으로써 상호작용성이 높은 새로운 방송 서비스를 제공받을 수 있다는 것

이다. 즉 시청자가 임의의 원하는 시점에서 영상을 재현할 수 있도록 함으로써 상호작용성이 높아 흥미로운 새로운 방송 서비스를 제공할 수 있다. 하지만 TV 영상을 보면서 시청자가 매번 시점을 지정하고 그에 따라 시점을 변환하여 TV를 시청하는 일은 매우 번거로운 뿐만 아니라 불편한 작업이 될 것이다. 입체영상을 시청하면서 임의의 시점으로 변환하는 작업은 시청하는 장면에 관한 3차원 공간을 즉흥적으로 지각하고 응용할 수 있어야 하기 때문이다.

본 고에서는 3D 자유시점 변환 TV가 갖는 이와 같은 문제점을 해결하기 위해 시청자가 TV를 제어하기 위해 사용하는 원격조종기에 지능형 서비스를 제공할 수 있도록 하는 방법을 소개하고자 한다. 즉 원격 조종기가 사람이 갖는 시각기능을 모방한 인공 시각 기능을 보유하여 입력영상을 분석할 수 있도록 한 후 시청자가 보고싶어 할 수 있는 몇 개의 후보 관심영역(Region of Interest)들을 추출하여 그 결과를 원격조종기 상에 표시함으로써 시청자가 즉흥적으로 후보시점 중 원하는 시점으로 시점변환을 쉽게

할 수 있도록 도와준다. 원격조종기가 인공시각 기능을 갖도록 하는 이유는 관심영역으로부터 추출된 시점이 사람들이 실제 추출한 시점과 어느 정도 비슷하게 하기 위함이다. 이렇게 함으로써 시청자는 자기가 보고 싶어하는 시점을 원격조종기의 화면에서 보여지는 여러 후보시점 중 원하는 시점을 즉흥적으로 선택하여 시점변환을 할 수 있도록 한다. 이렇게 시청자가 선택한 시점은 자유시점 변환 TV에 디스플레이된다.

<그림 1>은 시각주의시스템을 이용하여 개발된 3D 자유시점 변환 TV를 위한 원격조종기의 구성도를 보여준다. 먼저 자유시점 TV 상에 보여지는 영상을 시각주의시스템을 적용, 분석하여 9개의 관심영역을 자동으로 추출하고, 추출된 영역의 그림이 원격조종기에 표시되도록 한다. 시청자는 원격 조종기 상에 나타난 영역(시점)을 보면서 선택을 하게 된다. 시점은 현재 추출된 영역의 중심을 시점으로 계산하고 주어진 영상에서 시점이 이동하게 된다. 시청자의 선택 정보를 자유시점 TV에 전송하게 되면



<그림 1> 시각주의시스템을 이용하여 개발된 3D 자유시점 변환 TV를 위한 원격조종기의 구성도

선택된 시점을 중심으로 시점 변환된 영상이 화면상에 출력된다. 지능형 원격조종기는 터치 스크린을 기반으로 하기 때문에 해당시점을 터치함으로써 자연스럽게 시점변환을 할 수 있다. 이러한 지능형 원격조종기의 궁극적인 목적은 시청자가 자유시점 변환 TV를 쉽게 사용하도록 하는 상호작용상의 편의를 제공하기 위함이다.

II. 후보 시점 생성

생물학적인 시스템은 주어진 시각장면을 모두 병렬적으로 처리하기 보다는, 보다 복잡한 고차원적인 처리를 위해 주의가 집중되는 일정한 영역들을 순차적으로 포착하여 그 부분만을 순차적으로 처리해가는 전략을 사용한다. 이러한 생물학적 시스템에 있어서의 시각 주의는 실시간으로 입력되는 시각정보 처리에 있어서의 복잡도 문제 및 처리용량의 한계를 극복한다는 측면에서 매우 중요한 동물의 지적 능력이다. 시각주의(visual attention)란 이처럼 시각체계의 모든 시각정보 중 일부를 선택하거나 여과함으로써, 주의를 특정 영역이나 물체에 집중시켜 시각정보의 처리능력을 극대화하는 작용으로[4] 지금까지 심리학, 신경생리학, 컴퓨터과학 등 다양한 분야에서 연구되어져 왔으며, 공학적인 측면에서 영상 압축, 컴퓨터 시각(Computer Vision)에서의 목표물 자동탐지, 영상데이터베이스 등과 같은 몇몇 애플리케이션에 적용되어 왔다[5].

이러한 시각주의시스템은 자유시점 변환 TV에서 시청자가 임의의 시점을 생성할 수 있도록 하기 위해 적용될 수 있다[6]. 시각주의시스템에서 보다 복잡한 고차원적인 처리를 위해 주의가 집중되는 일정한 영역들은 바로 시청자의 주의가 집중되는 영역이

며 이러한 영역을 중심으로 시점을 생성하고 생성된 시점에 따라 영상을 생성시키면 시청자는 시점변환 영상을 감상할 수 있게 된다. 이렇게 함으로써 시청자는 보다 편하게 자신의 시점을 결정할 수 있다.

<그림 1>의 오른쪽 부분은 시각주의시스템의 전반적인 동작과정을 나타낸다. 시각주의 시스템은 크게 ‘분석’과 ‘통합’의 두 부분으로 구성되어 있다. ‘분석’ 단계에서 인간의 지각처리방식에 근거하여 색상, 명도, 방위, 형태, 깊이, 움직임에 반응하는 서로 독립적인 특징이 추출되어 특징지도로 구성되고, ‘통합’ 단계에서 특징지도가 재구성 및 통합되어 하나의 현저도지도가 만들어진다. 이 현저도지도가 시각주의 시스템에서의 출력이 된다. 시각주의시스템의 결과를 통해 ROI(Region of Interest) 또는 FOA(Focus of Attention)를 알아낼 수 있다.

1. 기준영상의 분석

1) 색상대비특징, 명도대비특징

색상대비특징은 인간의 색상지각능력에 대한 연구결과[7]를 바탕으로 한다. 색상대비(contrast)에 강하게 반응하는 동질(homogeneous) 형태의 수용야를 가진 색상대립세포 R+G-와 B+Y-를 모델링하였는데, R+G-기체의 경우, R에서는 활성화가 되고 G에서는 활성이 억제되어 R+G-세포가 흥분하면 적색을 지각하게 된다. 명도대비특징으로는 어두운 곳에서는 밝은 곳을 찾기 위해 입력되는 명도 값 그대로를 쓰고, 밝은 곳에서는 어두운 곳을 찾기 위해 원래 값의 반전값을 사용한다.

2) 형태-방위특징

인간의 형태지각은 머리에서 일어나는 인식 이전

의 과정으로서 물체를 형성하는 것이 아니라 덩어리나 경계선 등의 구분 정도를 나타낸다. Enroth-Cugell[8]는 고양이 눈에 대하여 실험한 결과 인간의 눈의 형태에 대한 반응은 좁은 폭의 가우시안과 넓은 폭의 가우시안의 역제반응으로 이루어져 있음을 알아내었다. 이를 이용하여 형태 정보를 추출해내는 Difference of Gaussian(DoG) 모형이 Marr[9]에 의해 연구되었다. DoG 모형은 형태 정보만을 알아내기 때문에 DoG 모형에 방위성을 부여한 Difference of oriented Gaussian[10]을 사용하여 색상/밝기대비 특징 지도에 대하여 8개의 방위($\theta = \{0/8\pi, 1/8\pi, \dots, 7/8\pi\}$)에 조율된 형태-방위 특징 지도를 각각 생성한다. 각각의 색상/밝기 특징 지도에 대하여 생성된 8개 방위에 조율된 형태-방위 특징 지도 $F^{\text{OK}}(\theta)$ 중 반응이 가장 큰 지도를 선택한다. 이를 통하여 이미 추출되어진 색상대비 특징 및 명도대비 특징들로부터 방위를 탐지하고, 덩어리나 경계선 등의 구분 정도를 확실히 드러내게 할 수 있다.

2. 스테레오 영상의 분석

인간의 시각체계는 서로 다른 위치에서 획득된 2개 영상을 적절히 정합함으로써 거리정보를 얻는 것으로 알려져 왔다. 스테레오 정합은 인간 시각 체계의 거리 추출 능력을 자동화하기 위한 컴퓨터시각 분야 중 하나이다. 인간의 두 눈의 기하학적 특성을 고려하여 3차원 공간상에 설치된 카메라로부터 얻어진 좌, 우 영상으로부터 3차원 정보를 획득하여 영상을 처리하게 된다. 일반적으로 스테레오 정합 결과에서 얻어진 정합점의 상대 거리를 변위(disparity)값으로 나타낼 수 있고, 이 값으로부터 물체까지의 거리(distance)나 깊이(depth)와 같은 3차원 정보를 구하게 된다. 최근 스테레오 매칭 알고

리즘들은 크게 지역적인 방법(local methods)과 전역적인 방법(global methods)으로 분류할 수 있다. 일반적으로 전역적인 방법은 보다 높은 정확도의 변위 정보(disparity information)를 추정하는 데 반해 지역적인 방법은 비교적 낮은 정확도를 가지지만 전역적인 방법에 비해 보다 단순하고 빠른 방법으로 알려져 있다. 특히 지역적 방법의 성능은 종류에 상관없이 매칭 비용 또는 유사성을 계산하는 방법에 따라 좌우된다. 일반적으로 기존의 지역적 방법은 매칭 비용 계산을 위하여 픽셀 밝기 값에 대한 유사도(pixel dissimilarity) 방법을 사용하고 있으나 최근 영상의 영역 분할을 이용한 방법[11]이나 적응적 가중치(Adaptive Support-Weight)를 이용한 방법[12] 등 전역적인 방법에 비해 보다 정확한 변위 정보를 추정할 수 있는 지역적 방법들이 제시되고 있다. 특히 [12]의 방법은 기존의 적응적 윈도우 방법과는 달리, 영역(support 또는 window) 내의 중심 픽셀과 이웃 픽셀들 간의 컬러의 차이와 기하학적 차이를 이용한 가우시안 가중치를 사용함으로써 보다 좋은 결과를 얻을 수 있다. 그러나 이 방법은 좋은 결과에 비해서 많은 시간이 걸리거나 많은 메모리를 필요로 한다. 따라서 실시간 스테레오 매칭이 불가능하다.

따라서 본 시각주의 시스템에서는 계산량이 적은 지역적인 방법을 사용한다. 영상의 깊이는 동일 축상에 정렬된 좌/우 영상에서, 각 영상의 픽셀에 대응하는 다른 영상의 픽셀간의 위치의 차이(Disparity)로 정의되므로, 두 영상의 화소들의 상관관계를 파악하여 카메라 위치로부터 물체들의 상대적 거리(깊이)를 측정한다. 깊이값은 왼쪽 영상에 대하여 오른쪽 영상과의 에러값이 가장 작은 값을 찾아 계산한다. 에러값 연산에 있어서는 그 두 지점간의 거리간의 차의 절대값 합인 SAD(sum of absolute

difference)를 이용하였다. 정확한 결과를 얻기 위해서는 좌/우 입력 영상이 정확히 수평으로 정렬되어 있어야 하고, 이미지의 각 픽셀은 주변의 픽셀들과 비슷한 깊이를 가져야 한다.

$$SAD(x, y, d) = \sum_{i=-\frac{1}{2}w}^{\frac{1}{2}w} \sum_{j=-\frac{1}{2}w}^{\frac{1}{2}w} |L(x+i, y+j) - R(x+i+d, y+j)| \quad (1)$$

식 (1)에서 w 값은 탐색할 영역의 범위를 결정하는 마스크 윈도우의 크기를 나타내며, L 은 왼쪽 영상, R 은 오른쪽 영상, d 는 거리를 나타낸다. 깊이를 구하기 위해 좌영상의 x, y 좌표에 해당하는 윈도우의 값과 우영상의 x 축으로 d 만큼 떨어진 윈도우의 값을 구하여 SAD볼륨 d 에서의 SAD값을 결정하고, 이 중에서 최소 SAD값을 찾는다. 최소 SAD값을 가지는 d 는 좌,우 영상간 값의 차이가 해당 d 에서 가장 작음을 의미한다. 이렇게 구한 d 값이 작을수록 해당 x, y 좌표의 픽셀은 카메라로부터의 깊이 값이 크고, d 값이 클수록 깊이 값은 작다.

3. 움직임 영상의 분석

현재 프레임(t)과 이전 프레임($t-1$)으로 이루어져 있는 연속적인 명도 영상 시퀀스를 입력받아 시간특징인 움직임 정보를 추출한다[13]. 이 영상 시퀀스에서 블록매칭기법을 사용하여 모션벡터(motion vector)를 추출한 후 모션벡터맵 M 을 구성한다. 본 시각주의시스템에서는 정확도를 위해 전역탐색기법을 사용하며, 매칭함수로는 SAD(Sum of Absolute Difference)를 이용하였다. 모션벡터를 계산하는 것은 시간 현저도지도도를 구성하기 위한 기본적인 과정이며, 이때 유의미한 모션들을 추출하는 것이 매우

중요한데 시간 현저도지도의 신뢰도가 높아질수록, 시공간현저도맵의 정확도가 향상되기 때문이다. 본 시각주의시스템에서는 심리학적 연구결과[14]를 이용하여 움직임 정보로부터 노이즈를 제거한다.

4. 분석된 결과의 통합

지금까지 기준영상인 왼쪽영상으로부터 색상/명도 특징지도 별 방위-형태 특징지도와 좌우 영상으로부터 깊이 특징지도가 구성되었으며, 움직임영상으로부터 움직임 특징지도가 구성되었다. 특징지도를 구성하고 있는 특징값들은 모두 입력된 영상의 각 지점마다 각각 특별한 의미를 가지고 있지만 각각의 특징들은 모두 서로 다른 특징추출방법으로 추출되어진 서로 독립적인 것들이기 때문에 이를 기반으로 주의 영역을 알아내기는 어려운 일이다. 따라서 영상의 각 지점에 있어서의 서로 다른 특징값들이 모두 하나로 통합되어져서 영상의 각 지점마다 하나의 전역적인 중요도 측정값인 Saliency가 계산되어야 한다. 본 시각주의시스템에서는 추출된 특징들이 모두 중요하다고 가정하였기 때문에 두드러진 특정 특징지도만을 선택하지 않고, 특징의 두드러진 정도에 따라 적절한 가중치를 부여하여 결합한다. 두드러짐 강도에 따라 높은 가중치를 부여하고, 두드러짐의 정도가 차이가 나지 않을 경우에는 차이가 나지 않는 특징맵은 서로 같은 가중치를 부여한 후 선형결합한다.

5. 시각주의 시스템에 대한 실험 및 결과

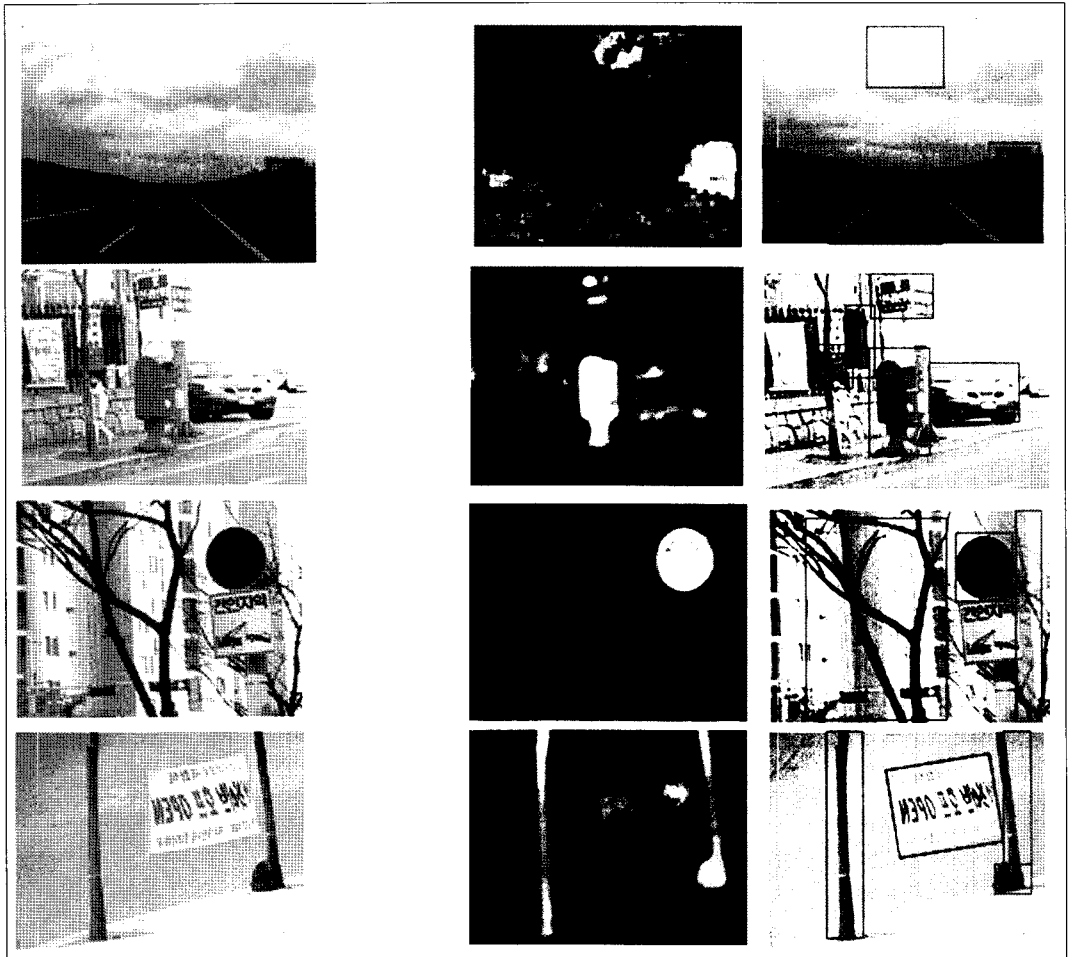
외부에서 촬영된 복잡한 실영상을 대상으로 시스템에 대한 성능을 분석하는 실험을 수행하였다. 이때 영상은 정지영상, 동영상, 스테레오 동영상 등이

될 수 있다. 본 시각주의시스템은 영상에 따라 생성되는 특징지도가 동적으로 변할 수 있다. 정지영상의 경우에는 공간특징을 사용하여 관심영역을 탐지하며, 동영상의 경우에는 시간특징도 같이 사용하여 관심영역을 탐지한다. 본 시각주의시스템은 실제 카메라를 이용하여 획득된 영상에 대한 임상실험에서도 실제 사람이 선택한 영역과 비교하였을 때 높은 확률로 탐지되었다. <그림 2>는 본 시각주

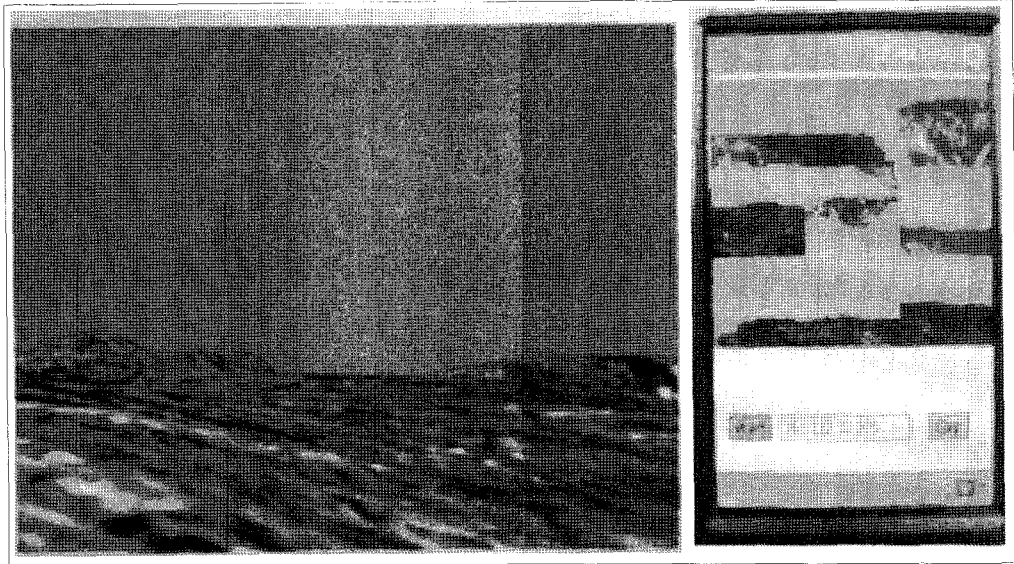
의시스템의 결과 예를 보여준다. 입력되는 영상에 대해 다수의 객체를 강도별로 탐지할 수 있음을 알 수 있다.

III. 구현된 지능형 원격조종기

지금까지 기술된 시각주의시스템을 이용하여 실



<그림 2> 시각주의시스템 결과 : (좌) 입력영상 (우) 현재도지도, 탐지된 후보주의영역



〈그림 3〉 구현된 3D 자유시점 변환 TV(좌)와 지능형 원격조종기(우)

제 구현된 지능형 원격조종기가 <그림 3>에 보여진다. 지능형 원격 조종기는 PDA와 무선인터넷을 사용하여 구현되었으며, PDA가 지능형 원격조종기로 사용된다. 3D 자유시점 변환TV는 PDA와 네트워크로 연결되어 있고, 무선인터넷으로는 지그비(Zigbee)와 UWB(Ultra-Wideband)가 사용될 수 있다. 원격조종기의 화면을 통해 시각주의시스템의 결과인 후보시점들을 볼 수 있으며, 시청자는 이를 기반으로 원하는 시점을 선택할 수 있다. 원격조종기에는 그 외에도 수평, 수직 이동 및 회전 등의 시점변환을 쉽게 할 수 있도록 기능이 설계되어 있다.

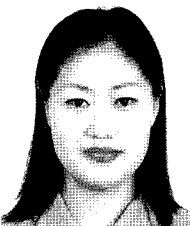
IV. 결론

본 고에서는 3D 자유시점 변환 TV가 갖는 시점변환에 관한 문제점을 일부 해결하면서 시청자에게 조작상의 편의를 제공하기 위한 지능형 원격조종기를 소개하였다. 본 고에서 이루어진 실험은 그래픽 영상에 대해서 이루어진 것으로 향후에 실제 다수의 카메라로부터 얻어진 영상에 대해 보다 구체적인 실험과 평가가 이루어져야 할 것이다. 3D 자유시점 변환 TV 기술은 기술구현상 다소 어려운 부분이 있음에도 불구하고 시청자에게 더욱 실감있는 방송을 제공하기 위한 최선의 솔루션으로 기대되며 점차 이러한 기대를 최대한 충족시켜 가는 방향으로 발전해 나갈 것이다.

참고 문헌

- [1] T. Fujii and T. Tanimoto, "Free-Viewpoint TV System Based on Ray-Space Representation", SPIE ITCOM, Vol. 4864, pp.175-189, 2002.
- [2] 조석환, 이도현, 김윤섭, 유창동, "자유 시점변환 비디오 시스템의 개발", 영상처리 및 이해에 관한 워크샵, pp.225, 2008.
- [3] Dongbo Min, Donghyun Kim and Kwanghoon Sohn, "Virtual View Rendering System for 3DTV", 영상처리 및 이해에 관한 워크샵, pp.226, 2008.
- [4] W. Osberger and A. J. Maeder, "Automatic identification of Perceptually important regions in an image", Proceedings of Fourteenth Intl. Conf. On Pattern Recognition 1, pp.701-704, 1998.
- [5] Min-Chul Park, Sang Ju Park, and Takayuki Hamamoto, "A Selective Visual Attention Module Based on Motion Stimuli", The Institute of Image Information and Television Engineers, Vol.59, No.12, pp.1862-1868, 2005.
- [6] Min-Chul Park, Sung Kyu Kim, and Jung-Young Son, "A perceptive remote-control for 3D TV interface", SPIE Newsroom, No.52, 2007.
- [7] L. Spillmann and John S. Werner, 'Visual Perception : the neurophysiological foundations', Academic Press Inc., 1990.
- [8] Enroth-Cugell C, Robson J G, The contrast sensitivity of retinal ganglion cells of the cat, 1996.
- [9] D.Marr, and E.Hildreth, Theory of Edge Detection, proc. Roy. Soc. London, Vol.207, pp.187-217, 1980.
- [10] R.Milanese, H.Wechsler, S.Gil, J.Bost, and T.Pun, Integration of Bottom-up and Top-down Cues for Visual Attention Using Non-Linear Relaxation. Proc. IEEE conf. Computer Vision and Pattern Recognition, pp.781-785, June, 1994.
- [11] F. Tombari, S. Mattocchia and L. Di Stefano, "Segmentation-based adaptive support for accurate stereo correspondence", In Proc. IEEE Pacific-Rim Symposium on Image and Video Technology, 2007.
- [12] K. J. Yoon and I. S. Kweon, "Adaptive Support-Weight Approach for Correspondence Search", IEEE Trans. PAMI, Vol.28, No.4, pp.650-656, 2006.
- [13] S. Li and M. Lee, "An Efficient Spatiotemporal Attention Model and Its Application to Shot Matching", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 17, No. 10, pp.1383-1387, 2007.
- [14] A. Hiroshi, "Visual Psychophysics (8) : Visual Motion Perception and Motion Pictures", Journal of Image Information and Television Engineers, Vol. 58, No. 8, pp.1151-1156, 2004.

필자 소개



최경주

- 1996년 : 충북대학교 컴퓨터과학과 학사
- 1999년 : 연세대학교 컴퓨터과학과 석사
- 2002년 : 연세대학교 컴퓨터과학 · 산업시스템공학과 박사
- 2002년 ~ 2005년 : LG CNS 연구개발센터 과장
- 2005년 ~ 현재 : 충북대학교 전자정보대학 컴퓨터공학부 조교수
- 주관분야 : 컴퓨터비전, 영상처리, 바이오컴퓨팅, 유비쿼터스컴퓨팅

필자소개



박민철

- 1993년 : 홍익대학교 전자공학과 학사
- 1997년 : 일본 동경대학 전자정보공학과 석사
- 2000년 : 일본 동경대학 전자정보공학과 박사
- 2001년 ~ 현재 : 한국과학기술연구원 포토닉스센서시스템센터 선임연구원
- 주관심분야 : 내용기반 영상검색, 멀티미디어, 3차원영상 디스플레이, 컴퓨터비전