

사용자 머리 추적 기술을 이용한 파노라마 비디오 직관적 제어

*,**홍유진 **황재인, **김익재

*과학기술연합대학원대학교, **한국과학기술연구원

*hyj@imrc.kist.re.kr

Intuitive Controlling Panoramic Video with Head Tracking Technique

*Hong, Yu-Jin **Hwang, Jae-In, **Kim, Ig-Jae

*University of Science and Technology, **Korea Institute of Science and Technology

요약

본 연구는 머리 추적 기술을 이용하여 파노라마 뷰어를 제어하는 새로운 방법을 제시한다. 360도 전방위로 녹화된 파노라마 비디오를 사용자의 머리를 추적하여 영상의 좌·우 회전, 위·아래 회전, 그리고 줌인/아웃이 가능하게 하며, 이는 사람의 관심 여부에 따라 머리가 움직이는 인간의 자연스러운 모습을 적용하여 쉽고 직관적인 제어가 가능하도록 한다. 실제 2009년 한국시리즈 야구 경기를 파노라마 카메라로 녹화하여 실험에 적용하여 실사용성을 확인하였다.

1. 서론

인간과 컴퓨터의 인터페이스는 전통적으로 마우스와 키보드를 사용하여 이루어져왔다. 최근 전통적인 입력 방식으로부터 나아가 더 쉽고 직관적인 인터랙션 방식에 대한 연구가 활발히 진행되고 있으며, 그 대표적 예는 인간의 행동을 인식하여 입력 신호로 이용하는 방법인데, 이는 사용자의 의도를 직관적으로 그리고 실시간으로 잘 반영할 수 있는 좋은 수단이 되기 때문이다. 최근 2-3년, Wii[1]를 시작으로 마이크로소프트사의 Kinect[2], 소니의 Move[3], 등 비전 기반 동작 인식 게임이 상용화 되어, 쉽고 직관적이며 빠른 응답성, 그로부터 오는 몰입감과 생생한 현장감을 제공할 수 있게 되었다. 인간의 제스처를 추적하기 위한 연구는 손 제스처의 추적, 눈의 시선 추적, 얼굴 추적의 예들이 대표적이다. J. Segen and S. Kumar[4]은 배경화면의 localization을 통해 실시간으로 손 제스처의 추적 연구를 하였으며, Francois[5]는 머리 움직임을 추적하여 모니터 등을 제어하는 가능성을 보였으며, L.P. Morency and T. Darrell [6]은 사용자 연구를 통해 머리 추적을 이용한 인터페이스의 실사용성을 보여주었다.

파노라마 영상은 360도 전방위 촬영된 이미지를 말하는 것으로 제한된 시야를 가진 일반 카메라의 영상보다 훨씬 많은 부의 정보를 포함하고 있으므로 흔히 투어나 몰입감, 현장감을 필요로 하는 환경에 쓰이고 있다. 대표적으로 최근 많은 이용자를 가지고 있는 Google의 Street View[7]나 Daum의 RoadView[8]가 파노라마 이미지를 사용하는 예이다. 또한 Microsoft사의 Photosynth[9]는 같은 장소에서 촬영한 이미지들을 결합하여 하나의 파노라마 이미지를 얻고, 마우스나 키보드를 조작하여 좌/우로 둘러보거나 줌인/줌 아웃이 가능하도록 하는 뷰어로서 웹에서 구현되어 있다.

본 연구는 머리 추적을 통하여 파노라마 비디오를 직관적으로 제어하는 방법을 적용하여 실제 야구 경기를 녹화한 파노라마 비디오에 적용하여 현장감과 몰입감을 높일 수 있도록 하였다. 실제 녹화된 파노라마

비디오는 360도의 시야각을 가지고 있어서, 현재의 디스플레이 디바이스의 해상도에서는 사용자가 한 눈에 시청하기엔 적합하지 않다. 따라서, 사용자의 의도에 따라 직관적으로 제어를 함으로써, 원하는 시점에서의 비디오 영상을 시청할 수 있도록 함에 그 목적이 있다.

2. 시스템 설계

가. 시스템 개요

[그림 1]에서 본 연구의 전체적인 과정을 설명하였다. 먼저 PC에 연결된 전방위 특수 카메라[10]로 각 FOV 에 해당하는 이미지를 획득한다. 이런 이미지들은 스티치 작업을 통해 하나의 이미지로 재생산 되는 과정을 거치면 360°의 영상을 얻어낼 수 있다.

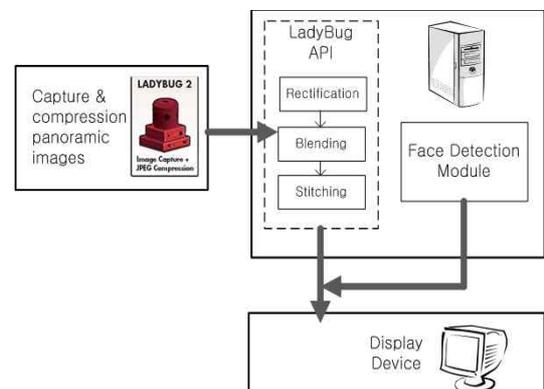


그림 1 시스템 구조
Figure 1 System Architecture

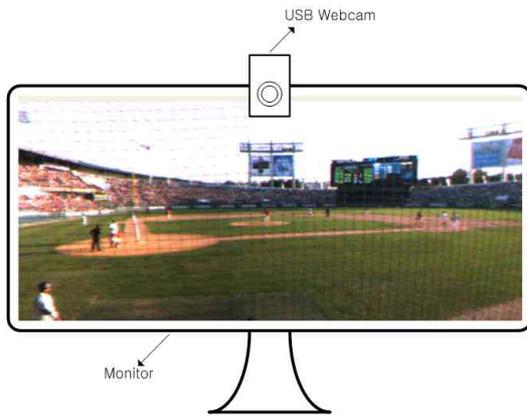


그림 2 시스템의 구성
Figure 2 System Setup

그 다음 이렇게 얻어진 파노라마 영상을 재생할 수 있는 뷰어를 구현하고, 뷰어의 시점을 직관적으로 제어하기 위하여 오직 모니터에 부착된 usb 웹캠을 통한 영상을 바탕으로 사용자의 머리의 3-D 위치, 오리엔테이션을 실시간으로 추적, 계산하여 영상을 좌우로 회전하고 줌인/아웃 할 수 있도록 한다. [그림 2]에 간단한 장치 예를 도식화 하였다.

나. 파노라마 비디오

본 연구에서는 LadyBug2[10]란 특수 파노라마 카메라를 이용하여 360°의 구 형태의 MJPEG 비디오를 획득하였다. LadyBug2는 1024 * 768의 해상도로 6개의 렌즈를 사용해 6개의 다른 뷰를 가진 이미지들을 얻어내며 IEEE-1394를 통해 12MP의 영상을 15 FPS로 Ladybug 2와 연결된 PC에 스트리밍 한다. 연결된 PC에서는 Ladybug API를 이용하여 6장을 스티치한 구 형태의 하나의 합쳐진 이미지를 재생산 하면 뷰어는 이 파노라마 이미지 프레임들을 재생시킨다. [그림 3]



그림 3 구 형태의 스티치된 이미지
Fig 3. Stitched spherical Image

다. 머리 추적

제스처는 감정과 의도를 상대방에게 표현할 뿐만 아니라 대부분이 무의식적으로 행하게 되는 즉각적이고도 직관적인 행동이다. 그러나 제스처의 종류와 그 표현 방법은 사람마다 다양하기 때문에 본 연구에서는 범위를 제한할 필요가 있다.

본 연구는 머리의 움직임을 추적하는 이유는 대부분의 사람들이, 관심 있는 대상을 보게 될 때는 자연스럽게 그 쪽으로 머리가 돌려지거나 더 가까이 보기 위해 머리를 앞으로 내미는 자연스러운 현상을

적용함으로써 단방향으로 단순히 재생되는 영상이 아닌 시청자의 질의와 응답을 실시간으로 반영 할 수 있도록 하였다.

그러나 머리 추적등의 비전 기반 알고리즘은 조명이나 안경, 수염 등의 장애물로 인해 강건한 추적이 어렵다는 단점이 있다. 우리는 이런 문제를 극복한 FaceAPI[11]을 적용하였다. FaceAPI는 머리의 방향 각(Orientation), 위치(location) 값을 1/frame 단위로 계산한다.

3. 구현

파노라마 영상은 모든 방향의 화면을 담고 있지만 실제 모니터에서 초기 재생되는 뷰는 특정한 단방향의 뷰이기 때문에, 사용자는 원하는 방향의 영상을 보기 위해서 뷰를 제어할 수 있는 방법이 필요하다. 이를 위해 먼저 제어에 필요한 머리 동작을 정의한 뒤, 정의한 동작에 가장 큰 영향을 미치는 머리의 방향각도와 위치 등의 파라미터에 대하여 생각해 볼 수 있다. [표 1]에서 정의한 동작과 파라미터를 나타내고 있다.

표 1 머리 움직임 동작과 제어 파라미터
Table 1 Head movements and Controlling parameters

머리 움직임	제어 파라미터
머리를 오른쪽/왼쪽으로 회전	Orientation의 Yaw
머리를 위/아래로 회전	Orientation의 Pitch
머리를 앞으로 움직임, 뒤로 움직임	Location의 Depth(Z)

우선 필요한 동작은 머리를 오른쪽/왼쪽으로 회전하는 동작이다. 이는 영상을 원하는 방향으로 조금씩 오른쪽/왼쪽으로 돌려보기 위함이다. 이 때 머리 축은 고정되어 있으면서 좌/우로 움직이기 때문에 머리의 수직 축을 Y축으로 본다면, Y축에 따라 회전하는 Yaw값(°)으로 계산될 수 있다. 같은 방법으로 X축의 Pitch방향각(°)을 이용하면, 머리의 위/아래 회전을 나타낼 수 있으며, 또한 머리의 3차원 위치를 계산한다면 화면에서 멀리 떨어져 있거나 가까이 다가가는 움직임을 통해 줌인/줌 아웃의 동작을 나타낼 수 있다.

[그림 4]는 사용자와 모니터 사이의 depth를 이용하여 줌인/줌아웃 되는 그림의 예시를 보여주며, [그림 5]은 카메라와 이미지, 그리고 world 좌표계의 관계를 나타낸다. [그림 6]은 Yaw 값을 이용하여 파노라마 영상을 좌/우로 회전하는 예이다.

영상을 회전하거나 줌인/아웃 하기 위한 각도나 위치 값들은 실험 환경에 따라 경험적으로 얻어진다. 본 연구에서는 사용자와 모니터 사이의 환경을 가지고 충분한 실험을 거쳐 경험적으로 얻은 값을 이용한다. Yaw값은 10°, Pitch값은 -20°, 그리고 depth값은 0.9m를 이용한다. 각 동작의 순서도는 [그림 7~9]으로 표시하였다.

한가지 더 생각해야 할 점은 회전 도중 움직이지 않는 상태를 고려해야 하는데 이는 탐색, 또는 원하는 방향을 둘러보다가 계속 집중하거나 흥미있는 부분이 생겨 움직임을 멈추고 가만히 응시하는 상태를 의미한다. 우리는 역시 실험값으로 문턱값을 정해 각 동작들마다 적용하였다.

그리고 이런 정해진 문턱값을 사용하였을 때 머리 회전에 비해 영상의 전환이 빠르거나 느리면 자연스럽지 못한 제어가 되므로 충분한

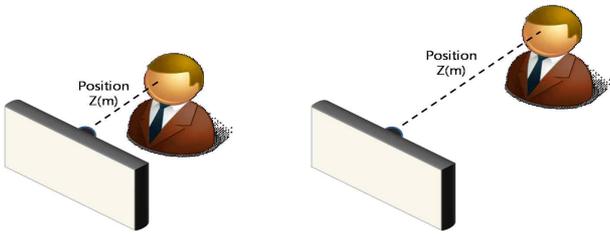


그림 4 모니터에 부착된 카메라와 사용자의 거리를 이용한 줌인/줌 아웃

Figure 4 The relation with the distance between attached camera and user and zooming in/out

실험을 통하여 얻어진 값들을 사용하여 최대한 어색함이 없이 자연스럽게 고도 즉각적인 응답이 될 수 있도록 하였다.

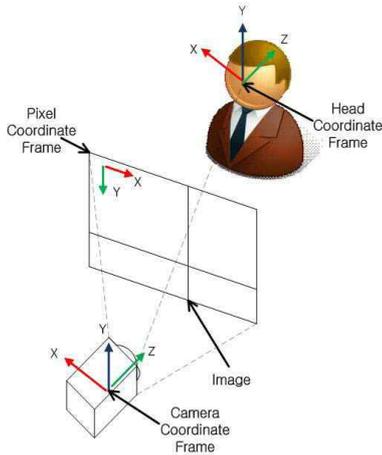


그림 5 카메라, 이미지, 그리고 world 좌표계

Figure 5 Camera, Image and World coordinates



그림 6 영상을 좌/우로 회전한 예

Figure 6 Turning left/right

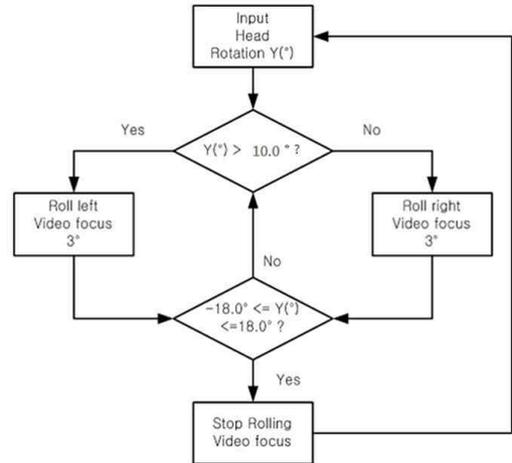


그림 7 머리 회전에 따른 영상의 좌/우 회전

Figure 7 Turning left/right with head motion

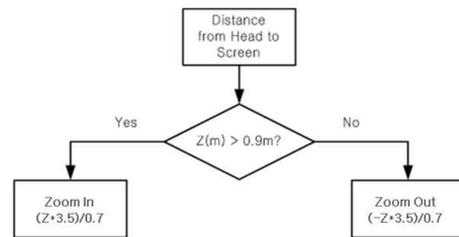


그림 8 머리와 모니터 간격에 의한 줌인/아웃

Figure 8 Zooming in/out with head position

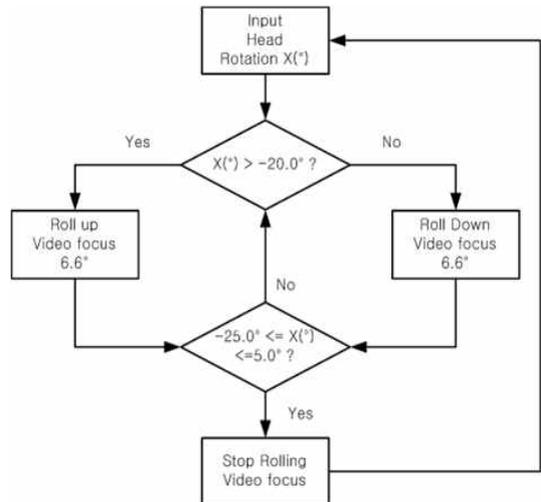


그림 9 머리 회전에 따른 영상의 위/아래 회전

Figure 9 Turning upwards/downwards with head motion

4. 결론 및 토의

[그림 10~12]은 정의한 동작들의 결과를 나타낸다. 실험 영상은 실제 2009년 야구 한국시리즈에서 Ladybug2로 녹화한 파노라마 비디오를 가지고 수행되었다. 얼굴이 먼저 인식되면 실시간 머리의 움직임 추적이 가능하며 이는 [그림10~12]에서 왼쪽 상단에 현 사용자의 상태와 추적 상태가 보여지게 된다. 사용자가 좌/우로 머리를 회전할 때 파노라마 영상 또한 좌/우로 회전하며, 좀 더 가까이 보고 위해, 모니터로 가까이 가는 경우에는 영상이 확대(zoom in)되며 멀어지면 축소(zoom out)된다. 그리고 머리를 위로 들거나 아래로 내렸을 때 영상 또한 위/아래로 움직이는 것을 볼 수 있다. 일반적으로 머리의 회전각은 정해져 있어 다양한 실험 환경에 동일하게 쓰일 수 있지만 모니터 혹은 스크린과 사용자의 위치에 따라 depth값이 변할 수 있으므로 depth값은 조정해 사용할 필요가 있다.

향후 계획으로는 본 논문에서 제안된 인터페이스를 통해서, 다수의 사용자들로부터 사용자 평가성을 받고, 이러한 피드백을 통해 시스템을 향상시키는 연구가 진행되어야 할 것이다. 앞으로 대화면, 고해상도 디지털 영상물이 많아짐에 따라, 이에 적합한 직관적 인터페이스가 많이 필요할 것이다. 특히 본 연구에서 제안하는 방법처럼, 사용자에게는 어떠한 추가적인 장비를 활용하지 않는 직관적인 인터페이스가 주류를 이룰 것으로 예상됨에 따라, 이러한 필요성에 적합한 인터페이스 연구가 활발히 이뤄져 할 것으로 생각된다.



그림 10 좌/우 회전

Figure 10 Turning Left/Right



그림 11 위/아래 회전

Figure 11 Turning Upwards/Downwards



그림 12 줌인/줌 아웃

Figure 12 Zooming In/Out

5. 참고 문헌

- [1] "Wii", <http://wii.com/>, Nintendo

- [2] "Kinect", <http://www.xbox.com/ko-KR/Kinect/Home>, Microsoft
 [3] "Move", <http://us.playstation.com/ps3/playstation-move/>, SONY
 [4] J. Segen, S. Kumar, "Shadow gestures: 3D hand pose estimation using a single camera", Proc. of the Computer Vision and Pattern Recognition Conference, CVPR99, v. 1: 485, 1999.
 [5] F. B&ard, "The perceptual window: Head motion as a new input stream", Proc. of the IFIP Conference on Human-Computer Interaction, INTERACT99, pp. 238-244, 1999.
 [6] L.-P. Morency, T. Darrell, "Head Gesture Recognition in IntelligentInterface: The Role of Context in Improving Recognition", Proc. of the 11th International Conference on Intelligent User Interfaces, Sydney, Australia, IUI06,
 [7] "Street View", <http://maps.google.com/help/maps/streetview/>, Google
 [8] "Road View", <http://local.daum.net/map/index.jsp>, Daum
 [9] "Photosynth", <http://photosynth.net/>, Microsoft
 [10] "Ladybug2", <http://ptgrey.com/>, PointGrey
 [11] "FaceAPI", <http://www.seeingmachines.com/product/faceapi/>, Seeing Machine