

키넥트를 사용한 체감형 학습 그림자 놀이 콘텐츠 개발

손종덕*, 이병국**

*동서대학교 영상콘텐츠학과

**동서대학교 멀티미디어학과

e-mail:mmaaxx@nate.com

Shadow Playing Contents Development by Using Kinect for Interactive Learning

Jong-Deok Son*, Byung-Gook Lee**

*Dept of Visual Contents, Dongseo University

**Dept of Multimedia Engineering, Dongseo University

요 약

본 논문에서는 영상 처리 및 컴퓨터 비전 분야의 기술을 마이크로 소프트 사의 키넥트에 적용하여 효과적인 체감형 학습을 위한 콘텐츠를 구성하고자 한다. 일반적으로 사람들이 많이 알고 있는 손 그림자 놀이를 응용하여 참여자들의 움직임을 통해 인터랙션을 발생시킬 수 있도록 하였고, 깊이정보맵으로부터 카메라에서 가장 가까운 거리에 있는 영역을 검출하기 위해 Meanshift segmentation(평균이동분할) 알고리즘을 적용 시켰다. 본 시스템의 체감형 콘텐츠는 문화 콘텐츠의 한 종류로서 이의 확장된 버전이 여러 분야에서 많은 활용이 될 것을 기대한다.

1. 서론

체감형 시스템은 사용자의 움직임이나 감각을 통해 디지털 콘텐츠를 조작하는 사람-컴퓨터간의 인터랙션을 활용한 시스템 방식이다. 창의적 사고, 다양한 경험 기반의 학습 활동, 그리고 인성 및 감성 중심의 학습에 대한 욕구를 충족 시켜줄 수 있으며, 최근에는 교육기관 및 문화시설에서도 유아 및 청소년들을 대상으로 한 체험 학습 및 엔터테인먼트적 성향이 높은 체감형 시스템을 요구하고 있으며 보급 또한 활발해 지고 있다. 우리가 흔히 알고 있는 손 그림자 놀이는 빛에 비춰진 손의 그림자들이 나타내는 형상을 가지고 이야기를 이끌어 나가는 방식으로 진행 되는 놀이이다.

본 연구에서는 마이크로 소프트사의 키넥트에 영상 처리 기술을 사용하여 그림자 놀이를 이용한 교육적이고 엔터테인먼트적인 요소를 갖춘 체감형 콘텐츠를 실현하는 시스템을 개발하였다. 특히 사람이 만든 손모양을 따내기 위해 영상처리에서 영역검출에 많이 사용되는 Meanshift(평균이동분할)알고리즘[1]-[4]을 적용하였다.

2. 시스템 구성

본 시스템에서는 그림 1과 같이 키넥트와 빔프로젝터, 스크린 그리고 키넥트와 빔 프로젝터를 제어할 컴퓨터가 사용된다. 메인 컴퓨터와 서버 컴퓨터 간에는 소켓통신을 통해 네트워크 통신을 유지 하고 키넥트가 설치된 각 서버 컴퓨터들에서는 사용자들이 직접 화면에 나오는 레퍼런스 이미지와 비슷한 모양을 가지는 손모양을 만들게 된다.

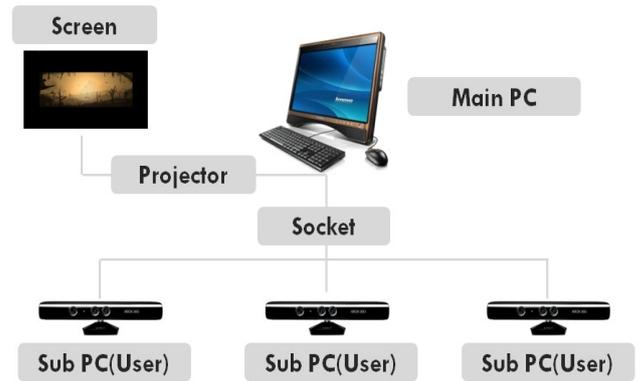


그림 1 시스템 설치 구성도

본 시스템에서는 3가지 종류의 레퍼런스 이미지(새, 토끼, 주먹)를 사용하게 되고 손 모양이 레퍼런스 이미지와 매칭이 됐을 경우 화면에 있던 새가 날라가는 모습, 토끼가 뛰어 가는 모습, 주먹이 떨어지는 모습등의 영상이 재생되고 메인 컴퓨터로 각 모델에 해당 되는 키보드 이벤트를 날려 준다. 그리고 플래쉬가 재생되고 있는 메인 컴퓨터에서는 소켓 통신을 통해 획득 되는 키보드 이벤트를 발생시켜 빔이 투사 되고 있는 스크린에 각각의 모델에 해당되는 동물이 나타나게 하고, 화면을 움직여가면서 서서히 그 동물들로 화면이 채워지는 플래쉬 영상을 플레이 시킨다. 본 시스템을 위한 기본적인 구성요소 기술들은 다음과 같다.

2.1 손 모양 검출

키넥트는 한 개의 RGB 카메라와 2개의 뎀스 카메라를 통해서 영상을 획득 하며 획득 되는 영상의 깊이 정보값을 획득하는데 용이하게 사용할 수 있다. 본 시스템에서는 깊이정보값을 얻을 수 있다는 사실 및 시스템을 사용하는 사용자는 대개 항상 키넥트를 향해 손을 뻗기 때문에 키넥트에 가장 가까이에 있다는 사실에 착안하여 키넥트로부터 일정거리 안에 들어 있는 객체에 대해서만 물체를 인식하는 방식으로 손 모양을 인식하게끔 시스템을 구성하였다. 그림 2는 이 방법으로 손모양을 추출한 예를 보여주고 있다. 그림 2(a)는 키넥트로 얻은 깊이 정보맵을 보여주고 있으며 그림 2(b)는 깊이정보맵에 mean shift segmentation(평균이동분할)과 모폴로지 연산을 수행하여 영역별로 분할된 결과를 보여주고 있다. 여기서 가장 밝기 값이 밝은 영역에 해당하는 부분이 키넥트에서 가까이 떨어진 영역이 되기 때문에 이 영역을 추출하면 그림 2(c)처럼 손 모양 영역이 얻어지게 된다.

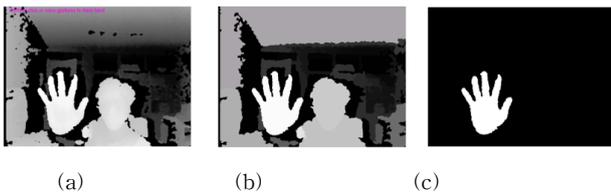


그림 2 손 모양 추출의 예

(a) 키넥트로 얻은 깊이정보맵 (b) meanshift segmentation을 수행한 결과 (c) 손모양 영역추출

Mean Shift기반의 segmentation(평균이동분할)은 객체의 경계선을 살리면서 분할을 수행할 수 있다는 점에서 획기적인 분할기술이다 [1]-[4]. Mean Shift(평균이동분할) 방법은 일정한 크기의 커널 윈도우 안에서 커널의 중심점과 커널 윈도우 내부의 모든 데이터 벡터간의 가중치 평균과의 차이 벡터를 계산하여 그 벡터만큼 커널 윈도우를 움직여가는 방식으로 데이터 벡터의 분포 중 가장 밀도가 높은 곳을 찾아가는 방법이다. 이때 데이터 벡터의 집합을 영상의 X,Y좌표와 RGB컴포넌트로 이루어진 5차원 벡터로 놓은 후, 모든 데이터 벡터들에 대하여 그 데이터 벡터를 중심으로 하는 커널을 씌우고, 각 커널에 대하여 Mean Shift(평균이동)을 반복적으로 수행하면 커널 윈도우들은 각각 가장 밀도가 높은 영역으로 수렴하게 된다. 이때 수렴된 점의 컬러를 데이터 벡터들의 대표 컬러로 놓고, 같은 점으로 수렴한 모든 데이터 벡터들을 한 영역으로 묶어줌으로써 분할이 이루어지게 되는 것이다.

이때의 mean shift(평균이동)벡터를 수식으로 표현하면 다음과 같다.

$$m(x) = \frac{\sum_{i=1}^n x_i g\left(\left\| \frac{x-x_i}{h} \right\|^2\right)}{\sum_{i=1}^n g\left(\left\| \frac{x-x_i}{h} \right\|^2\right)} - x$$

여기서 x 는 현재 커널윈도우의 중심 벡터를 나타내며, x_i 는 커널 윈도우내의 한 데이터 벡터를, n 은 커널 윈도우내에 존재하는 데이터 벡터의 갯수를, $g(\cdot)$ 는 커널함수를, 그리고 h 는 커널의 bandwidth, 즉, 크기를 나타낸다. $m(x)$ 는 meanshift 벡터가 된다. 여기서 x_i 벡터들은 X,Y 좌표와 RGB 컴포넌트들을 벡터 원소로 갖는 벡터들이다.

2.2 모델 매칭

Meanshift segmentation(평균이동분할)에 의해 손모양 영역이 검출되고 나면 이 검출된 영역을 레퍼런스 이미지랑 매칭하게 된다. 이때 손모양 영역이 레퍼런스 이미지랑 매칭이 되었다고 판단이 되면 소켓을 통해 PC에 신호를 보냄으로써 콘텐츠를 디스플레이하기 시작하게 된다. 이때 매칭을 하는 방법은 다음과 같다.

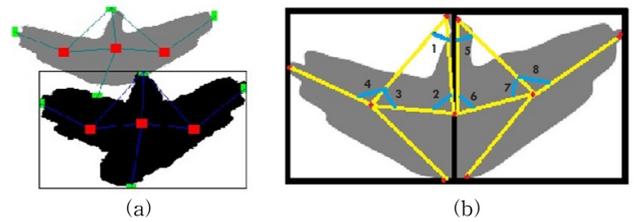


그림 3 각도 비교를 통한 모델 매칭

예를 들어 그림 3(a)는 회색의 레퍼런스 이미지와 검은색의 사용자가 직접 만들어 획득된 손모양 영역을 보여주고 있다. 이미지와 손모양영역으로부터 각각 상,하,좌,우의 최외곽포인트의 좌표를 획득한 후 그 좌표를 기준으로 가상의 사각형을 그리게 된다(그림 3(b)). 그리고 레퍼런스 이미지와 사용자가 만든 손모양의 이미지의 중심점을 찾고 중심점을 기준으로 처음의 사각형을 반으로 나누고 다시 각 사각형 마다의 중심점을 구한다. 각각의 획득된 점들 사이에 직선을 긋고 그려진 직선 사이에 나오는 1번부터 8번까지의 각들을 각각 서로 비교함으로써 레퍼런스 이미지와 사용자가 직접 만든 손모양의 매칭을 하게 된다.

이때 각도 비교는 모든 각도들에 대하여 독립적으로 수행하며 모든 각도가 레퍼런스 이미지의 각도들과의 차가 설정해놓은 임계치값보다 작을 때만 두 이미지가 매칭되었다고 판단하게 된다. 즉,

$$\max(s_0 - r_0, s_1 - r_1, \dots, s_n - r_n) < th$$

일 때 매칭이 되었다고 가정한다. 여기서 s_0, s_1, \dots, s_n 은 사용자가 만든 모양영역의 각도들이고, r_0, r_1, \dots, r_n 은 레퍼런스 이미지들의 각도들이며, th 은 미리 정해놓은 임계치값이다.

상기에서 설명한 모든 과정에 대한 순서도를 그림 4에 나타내었다.

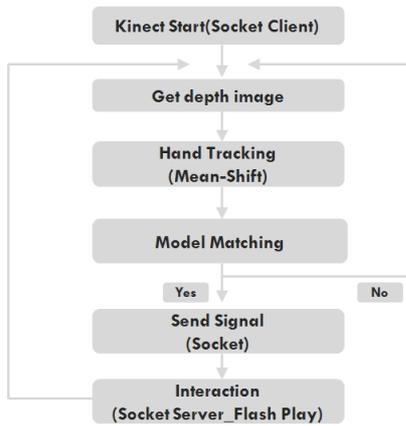


그림 4 전체 시스템 아키텍처

3. 시연 결과

상기에서 설명한 시스템으로 시연한 결과를 그림 5에 나타내었다. 왼쪽 열의 시스템은 PC에서 계산한 결과를 테이블안에 빔프로젝터가 있어서 그 빔프로젝터가 테이블에 결과영상을 빛으로 쏘아주는 시스템을 보여주고 있다. 밑줄로 갈수록 사람의 손모양이 레퍼런스 이미지에 맞추어지는 경과를 보여주고 있다. 오른쪽 그림에서는 사람의 손모양이 레퍼런스 이미지에 매칭되었다고 판단된 순간부터 화면상에 새가 나타나 날아가는 콘텐츠가 디스플레이 되는 결과를 보여주고 있다. 본 시연은 중국 상해공정대학교의 전시회때 이루어졌으며 체험형 콘텐츠로서 많은 즐거움을 선사한 엔터테인먼트적인 콘텐츠가 되었다.

4. 결론 및 향후 계획

일반적인 엔터테인먼트적인 콘텐츠나 장비들은 컨트롤러를 통해 조작하게 되지만 본 논문에서 제안한 시스템은 컨트롤러 없이 키넥트에서 획득한 영상을 처리함으로써 자유도를 얻을 수 있었고, 몸을 통한 조작과 이에 대한 인터랙션의 발생으로 인해 사용자에게 편리성뿐만 아니라 더욱더 큰 몰입감을 줄 수 있었다. 향후 시간적인 요소와 인터랙션에 의한 스토리 변화로 학습적 내용을 표현하는 콘텐츠를 추가 하며, 체험형 콘텐츠 개발 연구 분야에서

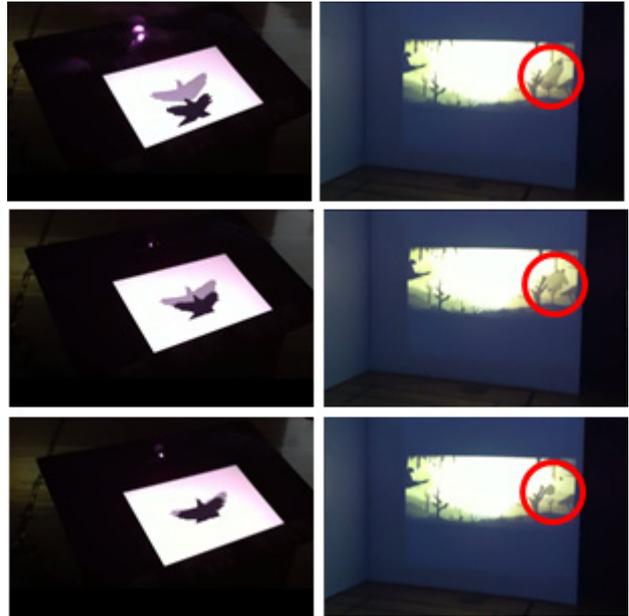


그림 5 시연결과

유용하게 사용될 것으로 예상된다. 세계 문화 콘텐츠 시장은 기존의 영화, 영상 콘텐츠 중심에서 체험 문화 및 체험 콘텐츠 중심으로 개편될 것이다. 문화, 예술 작품을 기반으로 하는 놀이와 감성적 성격이 강한 문화 콘텐츠의 중요성이 부각되고 단방향의 콘텐츠 보다는 인터랙션 형식면에서 직감적인 사용과 몰입감을 주는 콘텐츠 시장이 그 영역을 확장할 것이고, 체험형 콘텐츠 개발 연구는 이러한 세계 문화 콘텐츠 시장의 변화에 한 발 앞서 대비할 수 있는 계기가 될 것이다.

참고문헌

[1] Cheng, Y, "Mean shift, mode seeking, and clustering", IEEE Trans. PAMI, vol. 17, pp. 790-799, 1995.

[2] Dorin Comaniciu, "Mean Shift: A Robust Approach toward Feature Space Analysis", IEEE Trans. PAMI, vol. 24, no. 5, pp.603-619, May, 2002.

[3] G.R. Brandski, "Computer Vision Face Tracking as a Component of a Perceptual User Interface," Proc. IEEE Workshop Applications of Computer Vision, pp. 214-219, Oct. 1998.

[4] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," Proc. 2000 IEEE Conf. Computer Vision and Pattern Recognition, vol 2, pp. 142-149, June, 2000.