

가상 양시점화 방법을 이용한 비전기반 3차원 손 인터페이스

배동희*, 김진모**

동국대학교 멀티미디어학과*, 동국대학교 영상문화콘텐츠연구원**
{vision0529, mythsjin}@dongguk.edu

Vision based 3D Hand Interface Using Virtual Two-View Method

Dong-Hee Bae*, Jin-Mo Kim**

Dept. of Multimedia, Dongguk Univ.*,
Research Institute for Image & Cultural Content, Dongguk Univ.**

요약

최근 3차원 분야의 지속적인 발전으로 인해 보다 사실적이고 실감나는 영상을 경험할 수 있게 되었고 게임과 같은 다양한 응용분야에서 활용가능하게 되었다. 특히 가상환경에서의 객체들과 상호작용하며 이를 제어하는 증강현실 분야에 많은 발전을 가져왔다. 본 연구는 한 대의 카메라를 활용한 가상 양시점화 방법을 통해 3차원 공간의 객체를 제어하는 3차원 사용자 인터페이스를 제안한다. 이를 위해 임의의 두 카메라 위치 사이의 변환 정보를 담고 있는 호모그래피(homography) 행렬을 계산하고, 한 대의 카메라에서 얻은 2차원 손 좌표, 호모그래피 행렬 그리고 카메라의 투영행렬을 이용하여 3차원 좌표의 복원을 수행한다. 이러한 방법을 통해 보다 정확하고 빠른 3차원 정보를 얻을 수 있게 된다. 이는 두 대의 카메라를 동시에 구동할 때보다 연산량이 감소하여 실시간 처리에 효과적일 수 있으며 경제적인 부담도 줄일 수 있는 장점을 가지고 있다.

ABSTRACT

With the consistent development of the 3D application technique, visuals are available at more realistic quality and are utilized in many applications like game. In particular, interacting with 3D objects in virtual environments, 3D graphics have led to a substantial development in the augmented reality. This study proposes a 3D user interface to control objects in 3D space through virtual two-view method using only one camera. To do so, homography matrix including transformation information between arbitrary two positions of camera is calculated and 3D coordinates are reconstructed by employing the 2D hand coordinates derived from the single camera, homography matrix and projection matrix of camera. This method will result in more accurate and quick 3D information. This approach may be advantageous with respect to the reduced amount of calculation needed for using one camera rather than two and may be effective at the same time for real-time processes while it is economically efficient.

Keywords : 3D Hand Interface(3차원 손 인터페이스), Virtual Two-View Method(가상 양시점화 방법), Vision Based User Interface(비전기반 사용자 인터페이스)

Received: Sep. 10, 2013 Accepted: Oct. 04, 2013

Corresponding Author: Jin-Mo Kim(Dongguk University)

E-mail: mythsjin@dongguk.edu

© The Korea Game Society. All rights reserved. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ISSN: 1598-4540 / eISSN: 2287-8211

1. 서 론

최근 게임, 애니메이션을 비롯한 3차원 분야의 지속적인 발전으로 인해 보다 사실적이고 실감나는 영상을 경험 할 수 있게 되었다. 특히 가상 환경에서 객체들과 상호 작용하며 제어하는 가상현실 분야, 현실세계와 가상공간을 혼합하여 구현한 증강 현실 그리고 이를 게임에 활용하여 보다 사실적인 3차원 콘텐츠를 제작하는 등의 다양한 분야에 활용되며 많은 발전을 가져왔다[1,2]. 인간과 컴퓨터 사이의 상호 작용을 제어하는 기존의 인터페이스들 대부분은 2차원 영상을 제어하거나 분석하는 간단한 구조로 되어 있거나 3차원 제어의 경우에도 두 대 이상의 카메라를 이용하여 영상을 입력 받아 매 프레임 많은 연산을 처리하는 비효율적인 방법을 사용하는 경우가 대부분 이다[3,4]. 또한 적외선 센서와 같은 고가의 장비를 사용하거나 데이터글러브, 스페이스 뷰 등의 물리적 장치를 이용하여 3차원 객체를 제어하는 연산을 수행하지만 이는 효율성이 떨어지고 경제적인 면에서 큰 손실이다[1,5].

본 논문은 이러한 단점을 개선하여 쉽고 편하며 경제적 부담 또한 줄임으로써 게임과 같은 콘텐츠에 효율적으로 활용 가능한 3차원 손 인터페이스를 구현한다. 이는 특별한 물리적 장치를 필요로 하지 않는다. 뿐만 아니라 기존 스테레오기반의 두 대 카메라를 활용하는 방법과 달리 한 대의 PC 카메라만을 이용하여 정확한 3차원 손 좌표를 유도할 수 있는 가상 양시점화 방법을 제안하며, 누구나 손쉽게 사용가능한 3차원 객체 제어 인터페이스를 구현한다. 이를 위해 가상 카메라의 역할을 담당하는 호모그래피(homography) 행렬을 계산하고 이를 통해 3차원 좌표를 정확하고 빠르게 계산할 수 있도록 한다.

본 연구에서 제안하는 방법은 두 대의 카메라를 이용할 때와 유사한 결과를 갖는 정확한 3차원 좌표 계산이 가능하고 카메라 설치 위치의 제약이 없으며 경제적 이점이 있다는 장점이 있다.

2. 관련 연구

2.1 하드웨어기반 3차원 제어 인터페이스

기존에 사용된 3차원 제어 인터페이스는 센서가 부착되어 손에 차고 움직이면 이를 감지하여 위치 정보를 입력받는 데이터 글러브, 적외선 송수신기를 이용하여 위치정보를 획득하는 적외선 센서 3차원 인터페이스[5], 3D depth 카메라를 활용한 제스처 인식 인터페이스[6] 등이 있다. 이러한 방법들은 케이블에 의한 동작 제한을 받거나 적외선을 방출하는 물체에 의한 오작동의 우려가 있다.

이러한 단점을 극복하기 위해 적외선 센서와 카메라를 복합적으로 사용하여 인식률과 오동작 확률을 줄일 수 있는 Kinect 센서를 사용하는 방법들도 연구되고 있다[7]. 이러한 컨트롤러 입력 장치는 적외선 센서와 마커를 식별하는 카메라를 동시에 적용시켜 3차원 제어를 하는 방식이다. 하지만 하드웨어를 이용하는 대부분의 방법들은 수신 장비의 가격이 고가이기 때문에 일반인들이 사용하기에 경제적 부담일 뿐 아니라 유비쿼터스 환경에서 사용하기에는 어려움이 따른다.

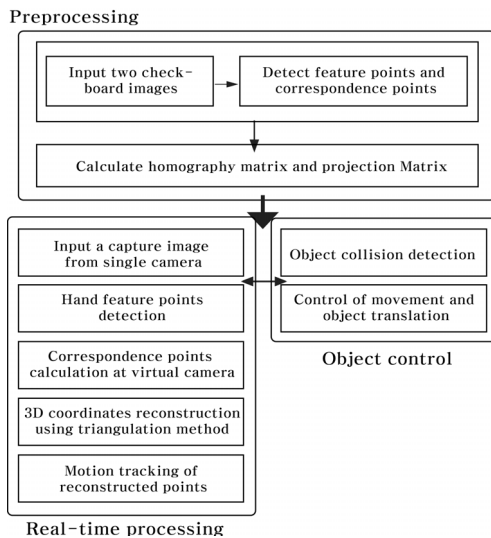
2.2 스테레오 카메라기반 인터페이스

고가의 하드웨어를 사용하는 방법들과 달리 카메라를 이용하여 3차원 공간의 객체를 제어하는 방법으로 스테레오 카메라기반 3차원 제어 인터페이스와 관련한 연구들도 진행되고 있다[8]. 또한 체감형 게임 인터페이스를 위해 전자적인 센서 방식이 아닌 동작 입력 인터페이스를 구현하는 등의 연구들도 진행되었다[9]. 카메라를 통해 입력받는 영상으로부터 3차원 좌표를 얻어내기 위해서는 일반적으로 두 대 이상의 카메라가 필요하다. 이러한 방법을 이용할 경우 행동환경의 제약이 거의 없이 3차원 객체의 제어가 가능해지며 비싼 입력장치 구입으로 인한 경제적 부담도 줄일 수 있다. 하지만 완제품으로 구성된 스테레오 카메라를 사용하지 않고 두 대의 카메라를 임의로 배치하여 사용할

경우 보정 및 조정과 같은 전처리 작업이 필요하며, 위치가 변경되면 재보정이 수행되는 번거로움이 존재한다. 이는 연산과정에서 오차를 발생하는 중요한 원인이 되기도 한다.

3. 비전기반 3차원 손 인터페이스

본 연구에서 제안하는 비전 기반 3차원 손 인터페이스는 [Fig. 1]과 같이 전처리를 통해 생성한 정보를 기반으로 실시간처리를 통해 3차원 가상 객체를 제어하는 구조이다. 전처리과정의 경우 스테레오 카메라기반 3차원 인터페이스에서의 일반적인 전처리 과정과 유사하지만, 본 연구는 한 대의 카메라만으로 스테레오 카메라의 역할을 수행하기 위하여 별도의 처리과정이 추가로 구현된다. 제안하는 3차원 손 인터페이스에서의 실시간 처리과정에서 두 대의 카메라사이의 수학적 관계를 유도하는데 발생할 수 있는 연산을 최소화함으로써 실시간 시스템에 더욱 효율적인 구조를 가질 수 있도록 하는 것이 본 연구의 새로운 제안방법이다.



[Fig. 1] Overview of our 3D hand interface system

3.1 가상 양시점화 방법을 위한 전처리

제안하는 3차원 인터페이스 구조는 스테레오 카메라를 기반으로 3차원 정보를 계산하는 방법을 활용한다. 하지만 본 연구는 한 대의 카메라만으로 두 대의 카메라를 사용하는 것과 같은 성능을 얻기 위해 한 대의 가상 카메라를 생성하는 가상 양시점화 방법을 제안한다. 우선 한 대의 카메라를 임의의 다른 위치에서 두 번 촬영하여 [Fig. 2](a)의 영상을 얻는다. 그리고 실시간 처리과정에서 가상 카메라로부터 촬영될 영상을 얻는데 필요한 정보를 전처리 과정에서 미리 계산한다. 따라서 촬영한 임의의 두 위치 중 하나는 실제 카메라의 위치, 나머지는 가상 카메라의 위치로 하여 두 카메라사이의 2차원 변환 정보인 호모그래피 행렬과 각 카메라의 투영행렬을 전처리 과정에서 계산한다.

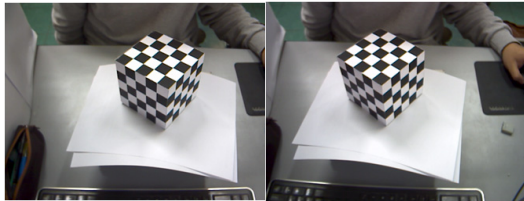
먼저 획득한 두 장의 영상들 사이의 2차원 변환 정보를 담고 있는 호모그래피 행렬을 계산한다. 두 영상으로부터 각각의 특징점(feature points)과 특징점들 사이의 대응 (correspondence)관계를 구하고 ([Fig. 2](b)), 대응점으로부터 호모그래피 행렬을 구한다[10]. 여기서 특징점은 해리스 코너 검출 방법을 이용하여 찾으며, 특징점 사이의 대응관계는 루카스 카나데 알고리즘을 사용하여 구한다. 첫 번째 영상의 특징점을 x_i , 이와 대응되는 두 번째 영상의 특징점을 x'_i 라고 할 때, 두 좌표 사이에는 x_i 에서 x'_i 으로 변환하는 2차원 변환 즉, 호모그래피 행렬 H 에 의한 $x'_i = Hx_i$ 의 관계가 성립한다. 이를 $x'_i \times Hx_i = 0$ 으로 정리하면,

$$x_i = (u_i, v_i, w_i)^T, x'_i = (u'_i, v'_i, w'_i)^T, H = \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix} \text{ 일}$$

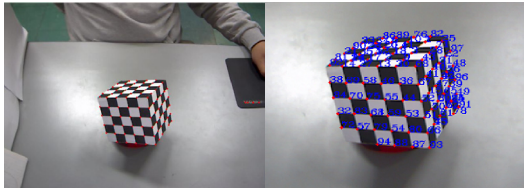
때, (eq. 1)의 $A_i h = 0$ 으로 정리된다.

$$\begin{aligned}
 x'_i \times Hx_i &= \begin{bmatrix} v'_i h_3^T x_i - w'_i h_2^T x_i \\ w'_i h_1^T x_i - u'_i h_3^T x_i \\ u'_i h_2^T x_i - v'_i h_1^T x_i \end{bmatrix} \quad (\text{eq. 1}) \\
 &= \begin{bmatrix} 0^T & -w'_i x_i^T & v'_i x_i^T \\ w'_i x_i^T & 0^T & u'_i x_i^T \\ -v'_i x_i^T & u'_i x_i^T & 0^T \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \end{bmatrix} = 0
 \end{aligned}$$

n 개의 대응점에 부합하는 A_i 행렬을 통해 호모그래피 행렬 H 를 계산한다. 이때 호모그래피 행렬의 정확성을 높이기 위해 RANSAC 알고리즘을 활용한다[10].



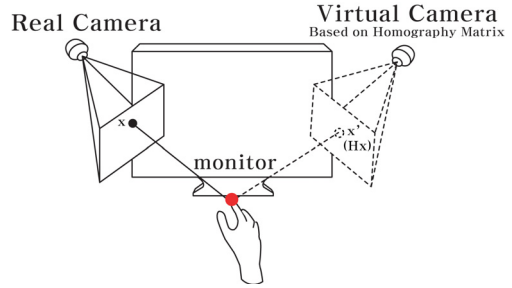
(a) two snapshot images



(b) feature points and correspondence points

[Fig. 2] Calculation of camera transform information using two images that were captured at arbitrary positions

이렇게 계산된 호모그래피 행렬은 실시간 3차원 좌표 계산에 사용되는 두 대의 카메라 중 가상 카메라 역할에 활용하게 된다. [Fig. 3]은 호모그래피 행렬기반의 가상 양시점화 방법을 통해 가상의 카메라를 설정하여 손의 특징점을 추정하는 상황을 나타낸 것이다.



[Fig. 3] Virtual two-view method based on homography matrix

다음은 두 카메라의 투영 행렬을 계산하는 것으로 [Fig. 2](a)의 체크무늬 정육면체를 활용하였다. 두 장의 영상으로 부터 [Fig. 2](b)와 같이 특징점을 검출하고 이로부터 6개의 대응점을 사용자가 직접 선택한다. 선택한 6개의 2차원 점(x) 각각에 규칙적인 간격으로 3차원 좌표(X)를 사용자가 지정하면, 두 좌표 사이에는 $x = PX$ 라는 일반적인 투영관계가 성립한다. 이를 자세히 기술하면 (eq. 2)와 같다[10].

$$x : \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} \sim \begin{bmatrix} m_{00} & m_{01} & m_{02} & m_{03} \\ m_{10} & m_{11} & m_{12} & m_{13} \\ m_{20} & m_{21} & m_{22} & m_{23} \end{bmatrix} \begin{bmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{bmatrix} : X \quad (\text{eq. 2})$$

또한 (eq. 2)는 $Am = 0$ 의 선형 방정식으로 표현 가능하며 이때 A 는

$$\begin{bmatrix} X_i & Y_i & Z_i & 1 & 0 & 0 & 0 & -u_i X_i & -u_i Y_i & -u_i Z_i \\ 0 & 0 & 0 & 0 & X_i & Y_i & Z_i & 1 & -v_i X_i & -v_i Y_i & -v_i Z_i \end{bmatrix}$$

으로 정의된다. 하나의 좌표마다 두 개의 식이 유도됨으로 3x4의 투영행렬을 유도하기 위해 최소 6개의 점이 필요한 것이다.

3.2 실시간 3차원 손 인터페이스 설계

3차원 손 인터페이스의 실시간 처리 과정은 영상처리를 이용하여 손 인터페이스에 필요한 특징점

을 구하는 과정과 전처리 과정에서 계산한 호모그래피 행렬, 투영 행렬을 활용하여 3차원 좌표를 계산하는 두 과정으로 구성된다.

우선 영상처리를 이용한 손 특징점 검출은 다음의 순서로 진행된다.

- ① 입력 손 영상을 복합 컬러 모델 변환 기법을 이용하여 검출
- ② 영상 필터를 이용한 노이즈 제거
- ③ GPU기반 세션화를 통한 손의 끝점 검출
- ④ 거리변환 행렬을 통한 손바닥 중심점 검출
- ⑤ 손 중심점과 끝점 사이의 비율을 통한 손가락 마디 점 검출

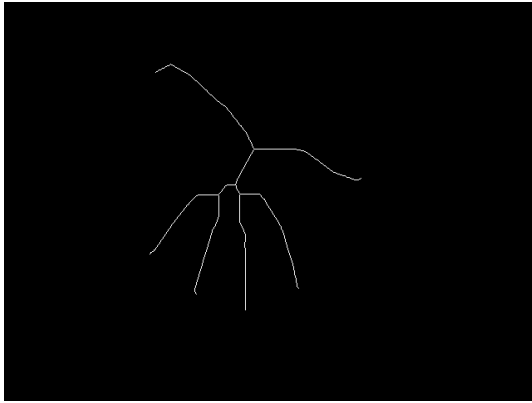
본 연구에서는 YCbCr 컬러 공간과 HSV 컬러 공간을 복합적으로 사용하여 손의 피부색 후보 영역을 검출한다. 인간 시각 특성은 색상보다 밝기에 더 민감하기 때문에 색차 신호를 이용하여 색 영역을 분리 처리할 때 효과적인 특징이 있다. 이러한 특성을 반영하여 YCbCr 공간을 활용하고, HSV 컬러 모델을 조명 밝기에 따라 손 특징점 검출 성능 차이를 줄이기 위하여 활용한다[11]. 따라서 HSV 모델을 이용하여 손 영역 후보를 검출하는 단계가 끝나면 YCbCr 컬러 모델 중 Cr 영역과 HSV 컬러 모델로 구해진 손 영역 후보를 논리곱 연산을 하여 최종적인 손 영역 후보를 분리한다 ([Fig. 4]).

입력되는 영상에서 Cr성분을 추출하게 되면 저주파로부터 고주파에 이르는 성분으로 구성되게 된다. 이때 고주파는 밝기 변화가 급격한 부분을 말하며, 저주파는 밝기 변화가 없거나 느린 부분을 말한다. 이때 공간 필터를 이용하여 잡영을 제거할 수 있다. 본 연구는 중간값 필터를 이용하여 잡영을 제거하였으며 [Fig. 4]는 이를 나타낸 결과이다.



[Fig. 4] Result of the noise removal process using median filter

다음은 손 인터페이스의 제어과정에 필수적인 손 끝점 검출이다. 본 연구는 손 끝점을 추출하기 위해 검출된 손 영역 영상에 세션화 과정을 수행한다. 이때 여러 종류의 세션화 알고리즘 중 Stentiford Algorithm을 최적화하여 사용하였다 [12]. 또한 실시간 처리에 있어 많은 연산시간이 필요한 세션화 과정의 성능 향상을 위해 nVidia의 CUDA를 활용하고, 수행속도가 최고 5.2배까지 향상됨을 확인하였다. 세션화 과정은 영상의 가로, 세로 방향으로 한 영상 안의 모든 화소들을 탐색하며, 삭제 후보 화소를 마크하고 삭제하는 과정으로 진행되며, 모든 픽셀에 반복적으로 동일한 연산이 적용되기 때문에 본 연구에서는 CUDA를 활용하여 이를 병렬처리 하였다. 따라서 GPU의 전역 메모리를 사용하여 세션화를 수행하고, 결과 값을 CPU로 반환하여 다음 연산을 수행하는 과정을 반복한다[13].



[Fig. 5] Result of a thinning hand image

마지막으로 거리변환 행렬을 사용하여 손바닥 중심에 대한 정보를 계산하고 손 끝점과 중심점 사이의 비율을 통해 손가락 각 마디 점들을 유도한다. 거리변환 행렬은 입력 영상에서 주어진 각 픽셀에서 주변 0값을 갖는 픽셀까지의 최단거리를 결과영상의 픽셀 값으로 설정하는 변환이다. [Fig. 6]은 거리변환 행렬을 통해 손바닥 중심을 찾은 결과이다. [Fig. 6]과 같이 손의 중심점을 찾으면 세선화([Fig. 5])를 통해 검출된 손 끝점을 이용해 손가락의 총 길이를 분할하는 방법을 통해 각 마디의 점들을 찾는다.

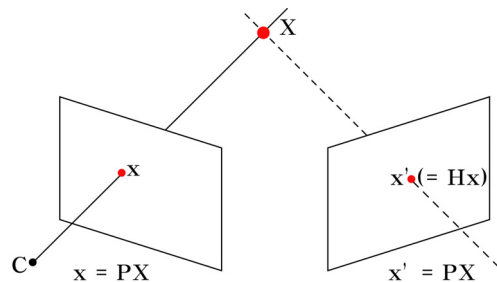
한 장 영상으로부터 세선화와 거리변환행렬을 적용하여 총 21개의 손끝과 마디 점, 그리고 중심점이 검출되면 전처리 과정에서 구한 호모그래피 행렬을 통해 가상의 두 번째 카메라에서의 2차원 영상 좌표를 계산한다. 두 대의 카메라로 촬영된 영상을 각각 영상처리하고 이들 사이의 대응관계를 구하여 3차원으로 복원하는 방법과 비교하면 제안하는 방법은 한 대의 카메라만을 활용하여 유사한 계산 결과를 얻으면서 영상처리에 필요한 연산이 절반으로 줄어드는 장점을 갖는다.

4.5	3.5	3	3	3	3.5	4.5
3.5	3	2	2	2	3	3.5
3	2	1.5	1	1.5	2	3
3	2	1	0	1	2	3
3	2	1.5	1	1.5	2	3
3.5	3	2	2	2	3	3.5
4	3.5	3	3	3	3.5	4



[Fig. 6] Result of a center point detection of hand using the distance transform matrix

최종적으로 호모그래피 행렬과 한 대의 카메라로부터 검출된 손 좌표를 적용하여 가상의 두 번째 카메라의 손 좌표가 모두 계산되면 전처리 과정에서 유도된 투영행렬과 삼각화 방법(triangulation method)을 이용하여 3차원 좌표를 계산한다[10]. [Fig. 7]은 삼각화 방법을 통해 3차원 손 좌표를 계산하는 과정을 나타낸 것이다.



[Fig. 7] A 3D coordinate reconstruction using the triangulation method

한 대의 카메라로부터 추출된 손 좌표와 호모그래피 행렬을 통해 계산된 가상 손 좌표를 각각

$x, x' (= Hx)$ 로 하고, 전처리 과정을 통해 계산된 투영 행렬을 P, P' 그리고 복원된 3차원 좌표를 X 라고 하면 $x = PX, x' = P'X$ 의 관계가 성립된다. 그리고 이를 $x \times PX = 0, x' \times P'X = 0$ 으로 유도하여 $AX = 0$ 의 선형 방정식으로 정리하면 A 는 (eq. 3)과 같이 정리된다.

$$\begin{bmatrix} up^{3T} - p^{1T} \\ vp^{3T} - p^{2T} \\ u'p'^{3T} - p'^{1T} \\ v'p'^{3T} - p'^{2T} \end{bmatrix} = \begin{bmatrix} up^{3T} - p^{1T} \\ vp^{3T} - p^{2T} \\ \frac{h^1 \cdot x}{h^3 \cdot x} p'^{3T} - p'^{1T} \\ \frac{h^2 \cdot x}{h^3 \cdot x} p'^{3T} - p'^{2T} \end{bmatrix} \quad (\text{eq. 3})$$

여기서, $x_i = (u_i, v_i, w_i)^T, x'_i = (u'_i, v'_i, w'_i)^T$ 이고 $H = \begin{bmatrix} h_1^T \\ h_2^T \\ h_3^T \end{bmatrix}$ 정의한다. 이처럼 한 번의 전처리 과정

을 통해 투영행렬과 호모그래피 행렬을 계산하면, 매 프레임 새로운 좌표가 입력되어도 정확한 3차원 좌표를 매우 효율적인 방법으로 구할 수 있다. 본 연구에서 제안하는 방법은 카메라를 두 대 사용하여 매 프레임 스테레오 영상에서 손의 특징 좌표를 검출하고, 이를 통해 3차원 좌표를 얻는 것과 비교하여 많은 연산 과정을 감소시켜 주며, 두 장 영상에서 서로 다른 특징점을 검출하여 발생할 수 있는 오류를 줄이는 역할을 하게 된다. 이는 스테레오 카메라와 비교하여 3D 깊이 정보를 정확히 고려하지는 못하나 3차원 depth 카메라와 같은 별도의 장비 없이 하나의 카메라만으로 3D 정보를 보다 근사하게 유도함으로써 활용성을 극대화하게 된다.

4. 실험 및 검토

본 연구는 검출된 3차원 손 좌표를 효과적으로 활용할 수 있음을 증명하기 위해 3단계로 나누어 실험을 진행하였다. 우선 한 대의 카메라를 통해

입력받아 검출되는 손의 특징 좌표를 정확히 추정하여 검출해주는 호모그래피 행렬 계산의 정확성을 확인하고, 추정된 좌표를 통해 3차원 좌표를 복원할 경우 복원 좌표의 정확성에 대한 실험을 진행하였다. 마지막으로 복원된 3차원 좌표를 가상공간에 적용하여 활용하는 3차원 손 인터페이스를 구현하였다.

실험환경은 일반적인 실내의 형광등 조명에서 이루어졌고, 실험 PC 사양은 Intel core2 Duo CPU, DDR2 3GByte Memory, Geforce 9600GT VGA, aInfo PC camera (10M Pixel)을 사용하였다. 또한 Visual C++ 2005, MFC, OpenCV, OpenGL 라이브러리를 이용하여 구현하였다. 본 연구에서 사용되는 aInfo PC camera는 사용자가 쉽게 구할 수 있는 저가 사양의 USB 웹캠이다.

본 연구는 손 끝점과 각 마디 좌표 추출은 물론 보다 정확한 손의 3차원 정보를 계산하여 사실적인 손 인터페이스를 구축함을 목표로 한다. 따라서 제안하는 방법의 성능을 비교하기 위해 우선 입력 영상으로부터 손의 인식률과 인식 속도를 측정하였다. 실험은 초당 24프레임 환경에서 1분 동안 입력된 1440장의 영상에 대해 진행하였다. 본 논문에서 사용한 손 좌표 검출 방법은 기존의 영상처리 방법을 활용하였기 때문에 기본적으로 인식률의 큰 차이는 없다. 그러나 두 대의 카메라를 배치하여 각각 영상을 입력받고, 제안한 방법을 통해 한 대의 카메라만으로 손을 인식하는 과정에서 가려진 손이 발생하는 경우 다른 한 대의 카메라를 작동시켜 보완함으로써 인식률을 증가하게 된다. 이 경우 계산된 호모그래피 행렬의 역행렬(H^{-1})을 사용하여 반대 영상의 대응 좌표를 계산하도록 한다. 그리고 두 대의 카메라로 영상을 입력받아 손 좌표를 찾고 대응관계를 분석한 후 3차원 좌표를 계산하는 일반적인 두 대의 카메라를 활용한 인터페이스는 실험결과 한 프레임에서 대응 좌표 계산이 평균 0.63초, 3차원 좌표 복원은 평균 0.72초 소요되었다. 결과적으로 인식 속도가 현저히 감소하여 실시간 시스템에 사용하기에 부적합함을 알 수 있

다. [Table 1]은 이를 나타낸 것이다. 본 연구에서 제안하는 방법과 두 대의 카메라를 활용한 인터페이스의 인식 속도 결과는 3차원 좌표 계산까지의 속도로 2차원 좌표만을 고려할 경우 성능이 더욱 좋아질 것이다.

[Table 1] Comparison of the hand recognition result with the existing method[3]

	recognition speed	recognition rate
our method	24~26 fps	92.2%
existing method[3]	38~43 fps	91.4%
two-view method	9~15 fps	96.2%

다음은 호모그래피 추정 좌표의 확인 실험이다. 두 대의 카메라를 이용하여 사전에 제작된 체크무늬 패턴의 정육면체를 촬영하고, 촬영된 이미지로부터 카메라의 투영정보와 호모그래피 행렬을 계산하였다. 이를 통해 한 대 카메라로부터 촬영된 입력 손 영상으로부터 호모그래피 행렬을 이용해 계산된 가상 추정 좌표와 해당 위치에 실제 카메라를 배치하여 촬영된 영상의 좌표를 비교하여 오차를 확인하였다.

두 대의 카메라를 임의의 위치에 배치하여 손 영상을 입력받아 제안한 영상처리 방법을 통해 손 특징 좌표를 계산한다. 그리고 호모그래피 행렬을 통한 가상 좌표 또한 동시에 실시간으로 계산한다. 그런 다음 다양한 동작으로 움직이는 손 영상 가운데 임의의 한 프레임 영상을 선택하여 각각 손 특징 좌표를 계산하여 정확성을 실험하였다. [Table 2]는 15개의 손 특징 좌표를 비교하여 유클리디안 거리 공식을 통해 실제 좌표와 오차 좌표의 거리를 실험한 결과이다. 실험결과 전체 오차의 평균은 약 1.4 픽셀 정도로 크지 않음을 확인할 수 있다. 단, 전처리 과정에서의 정육면체 위치와 카메라로 입력받는 손 영상의 위치가 멀어지면 호모그래피 행렬의 정확성이 떨어질 수 있다. [Table

2]에서 실험에 사용된 임의의 영상 내에 엄지손가락의 위치가 호모그래피 행렬 계산에 사용된 중심 영역 밖에 존재하여 유독 큰 오차가 발생하였다. 이는 손 인식의 관심 영역을 설정함으로써 해결할 수 있을 것이다.

[Table 2] Comparison between real coordinate and calculated coordinate based on the virtual two-view method

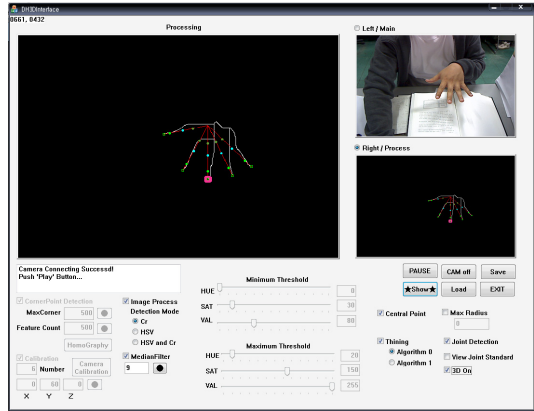
		real coordinate	calculated coordinate using homography
thumb	end point	427, 212	430, 216
	first knuckle	394, 202	410, 205
little finger	end point	175, 302	179, 299
	first knuckle	210, 256	210, 256
	second knuckle	190, 284	190, 284
index finger	end point	354, 343	356, 345
	first knuckle	333, 271	333, 271
	second knuckle	344, 305	344, 305
ring finger	end point	231, 350	233, 346
	first knuckle	249, 277	247, 276
	second knuckle	239, 310	239, 310
middle finger	end point	290, 369	290, 369
	first knuckle	290, 282	287, 281
	second knuckle	290, 317	289, 316
center point	center	284, 166	284, 166
error sum		21.5 pixel	
average error		1.4 pixel	

그리고 투영행렬을 통한 3차원 좌표 복원에 대한 확인 실험이다. 두 대의 카메라를 사용하였을 경우 카메라를 통해 투영된 두 영상의 대응점 정보를 계산하여 카메라 투영행렬을 통해 3차원 좌표를 복원한 결과와 호모그래피 행렬을 통해 추정된 대응점과 한 개의 카메라를 통해 입력된 대응점을 이용하여 3차원 좌표를 복원하는 결과를 비교하였다. 위의 실험환경과 동일하게 임의로 프레임임 30장을 선택하여 10개 복원 좌표의 오차를 실험하였다. [Table 3]과 같이 평균 오차가 0.5정도로 크지 않음을 확인할 수 있다. 3차원 좌표를 OpenGL을 활용하여 계산된 값으로 OpenGL 공간에서 1 값을 기준으로 0.5정도의 오차를 갖는 것을 의미한다. 여기서 최대오차 역시 전처리 과정에서 호모그래피 행렬 계산에 사용된 중심영역을 많이 벗어나는 위치의 손 좌표로 호모그래피 행렬 연산 과정에서 오차가 발생하여 자연스럽게 3차원 좌표 복원 또한 오차가 커지게 된다.

[Table 3] Comparison of reconstructed 3D coordinates error

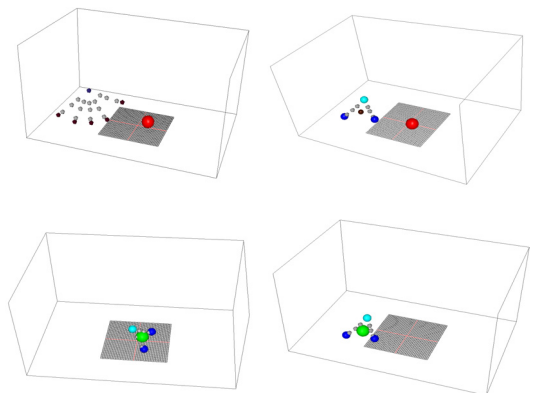
number of reconstructed points	error sum	average error	maximum error
10	4.77563	0.47756	0.88456

마지막으로 가상공간에서의 객체 제어 3차원 손 인터페이스를 구현한 결과이다. [Fig. 8]은 한 대의 카메라로부터 입력된 손 영상을 제안한 영상처리과정을 통해 손 좌표를 계산하는 시스템이다.



[Fig. 8] Proposed hand interface system

손 좌표 추출에 필요한 모든 영상처리 과정이 끝나면 팝업 창을 이용하여 3차원 가상공간 안에 실시간 복원된 손 좌표가 자유롭게 이동하며 다른 객체와 상호작용할 수 있도록 구현하였다. [Fig. 9]는 이를 나타낸 것으로 손의 모든 좌표를 구로 표현하였다. 그리고 가상의 객체를 손을 이용해 잡은 후 이동시키는 처리를 표현하였다. 손의 모양에 따라 객체를 잡아 이동시키고, 다시 손을 잡고 있던 객체를 놓게 된다. 제안하는 손 인터페이스는 특정 제스처에 대한 패턴을 고려하지 않고 비교적 자유자재로 가상공간을 움직이게 하는 것을 목표로 하였다.



[Fig. 9] Result of object control in virtual space using proposed 3D hand interface

5. 결 론

본 연구는 한 대의 카메라를 이용하여 입력받은 손의 위치를 3차원 좌표로 복원하고 이를 통해 3차원 공간의 객체를 제어하는 인터페이스 설계에 새로운 방법으로 접근하여 효율적인 처리를 할 수 있도록 제안하는 것이다. 실험에 사용된 카메라와 같이 고가의 카메라가 아닌 일반인들도 누구나 쉽게 구입하여 사용할 수 있는 PC 카메라로서 이를 이용하여 누구나 쉽게 사용 가능한 3차원 손 인터페이스를 구현하였다. 영상처리를 이용하여 입력받은 손의 영상에서 특징점 좌표를 검출하는 과정과 비전 기반으로 3차원 좌표를 복원해 나가는 과정을 통해, 한 대의 카메라만으로 3차원 가상공간을 제어하는 인터페이스를 구현 가능하게 하였다. 현재 시스템 환경에서는 불가피하게 전처리 과정을 통해 3차원 복원의 기본이 되는 호모그래피 행렬과 투영 행렬을 구하는 작업을 수행하여야 하지만, 한번 계산된 호모그래피 행렬과 투영행렬을 저장 후 프로그램이 수행될 때 자동으로 로드하는 방식을 사용하기 때문에 재계산의 번거로움을 충분히 해결하였다.

3차원 복원 기술과 비전기술이 나날이 발전하고, 하루가 다르게 3차원 분야에 새로운 기술들이 출현하는 오늘날, 3차원 공간의 객체를 보다 효율적이고 체험적으로 제어할 수 있는 인터페이스 개발에 대한 연구가 굉장히 시급한 실정이다. 특히 게임 산업과 가상현실 분야에서 이러한 연구가 보다 깊이 있고 넓게 이루어져야 할 것이다. 본 연구에서 제안한 방법은 저렴한 한 대의 PC 카메라만을 이용하여 3차원 객체를 제어할 수 있다는 장점이 있다. 때문에 여러 대의 카메라를 장착할 수 없는 휴대용 컴퓨터에서도 쉽게 3차원 객체 제어가 가능하다. 제안하는 방법은 한 대의 카메라만으로 두 대의 카메라 역할을 수행하기 위한 새로운 방법을 제안한 것이지만 정확한 3차원 정보를 담고 있지 못하는 한계가 있다. 향후 에피폴라 기하 정보를 고려하여 정확한 3차원 정보를 유도할 수 있도록

보완할 것이며, 손동작 제스처를 몇 가지 선정하여 인터페이스를 구축함으로써 게임과 같은 응용 시스템에 효율적으로 활용 가능하도록 할 것이다.

REFERENCES

- [1] Dong-Ha Lee and Mun-Hee Lee, "Current status of human - computer interaction (HCI) technology policy", The Magazine of the IEEK, Vol. 34, No.6, pp. 628-637, 2007.
- [2] Woon-Tack Woo, Namgyu Kim, Karen Wong and Makoto Tadenuma, "Sketch on dynamic gesture tracking and analysis exploiting vision-based 3D interface", Proceeding of SPIE, Vol. 4310, pp. 656-666, 2001.
- [3] Boo-Nyon Kim, Jongho Kim and Tae-Young Kim, "A method of hand recognition for virtual hand control of virtual reality game environment", Journal of Korea Game Society, Vol. 10, No. 2, pp. 49-56, 2010.
- [4] T. Maung, "Real-time hand tracking and gesture recognition system using neural networks", Proc. of World Academy of Science: Engineering & Technology, Vol. 50, pp. 466-470, 2009.
- [5] C. J. Lim, Hun Lee, and Yun-Guen Jeong, "Validation test of vision-based eye controlled user interface", Journal of Korean Society for Computer Game, Vol. 24, No. 2, pp. 141-151, 2011.
- [6] Hyun-Cheol Yoon and Jong-Seung Park, "Gesture-based game character animation using motion data captured from the kinect sensor", Journal of Korean Society for Computer Game, Vol. 26, No. 2, pp. 147-153, 2013.
- [7] Guan-Feng He, Jin-Woong Park, Sun-Kyung Kang and Sung-Tae Jung, "Development of gesture recognition-based 3D serious games", Journal of Korea Game Society, Vol. 11, No. 6, pp. 103-114, 2011.
- [8] D. Goldin, S. Smolka, P. Wegner and D. Eds, "Iterated dynamic programming and quadtree subregioning for fast stereo matching", Image

and Vision Computing, Vol. 26, Issue 10, pp. 1371-1383, 2008.

- [9] J. Lee, Hyung-Gi Kim, and Won-Hyung Lee, "A multimedia implementation using tangible game interface", Journal of Korean Society for Computer Game, Vol. 7, pp. 16-22, 2005.
- [10] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision", Cambridge University Press, 2004.
- [11] Ae-Kyung Kang, "Image processing- based hand interface to control 3D graphic objects", Master thesis, Dongguk University, 2008.
- [12] J. R. Parker, "Algorithms for image processing and computer vision", John Wiley and Sons, 1996.
- [13] David B. Kirk, Wen-mei W. Hwu, "Programming massively parallel processors", Elsevier Inc., 2010.



배 동 희 (Bae, Dong Hee)

2009년 2월 동아방송예술대학교 방송기술학과 졸업 (전문학사)

2011년 8월 동국대학교 멀티미디어학과 졸업(공학석사)

2013년-현재 TJB 대전방송 기술국 기술운영팀

관심분야 : 컴퓨터그래픽스, 컴퓨터비전, 방송기술



김 진 모 (Kim, Jin Mo)

2006년 2월 동국대학교 멀티미디어공학과 졸업(공학사)

2008년 2월 동국대학교 멀티미디어학과 졸업
(예술공학석사)

2012년 8월 동국대학교 멀티미디어학과 졸업(공학박사)

2012년 9월-현재 동국대학교 영상문화콘텐츠연구원 전
임연구원

관심분야 : 게임 프로그래밍, 컴퓨터그래픽스, 게임공학,
실시간 렌더링, 영상처리
