

3차원 지시 동작 인식을 위한 휴먼인터페이스

조 강현

울산대학교 전기 전자 및 자동화공학부

<요 약>

본 논문에서는, 사용자가 자신이 하는 공간지시를 시각에 기초하여 인식하여 3차원 컴퓨터 그래픽스 세계에 있는 대상체를 조작하는 휴먼인터페이스 시스템을 소개한다. 시스템은 공간기준의 공간기준계와 사용자의 몸에 고정된 사용자중심계에 대하여 손의 동작을 인식한다. 후자의 경우에는, 자신의 몸을 위치를 바꾸더라도 사용자가 앞으로 가리키면 대상체를 앞으로 움직일 수 있도록 한다. 더욱이, 액티브 카메라 시스템으로 사용자를 추적하기 때문에 사용자가 카메라의 화각을 의식하지 않아도 된다. 그리고, 아무런 카메라의 보정을 하지 않아도 된다. 이러한 특징을 갖춘 시스템을 다중 영상의 어파인 불변량 이론에 기초하여 구현을 하였다. 이를 통해 영상알고리즘과 시스템을 실험해 보았다. 실제 사용 실험을 통하여, 본 시스템의 효용성을 밝혔다.

Human Interface for Human Pointing Gesture Recognition

Kang-Hyun Jo

Dept. Control and Instrumentation Eng.

Email: jkh@uou.ulsan.ac.kr

<Abstract>

In this paper, a vision-based human interface system which enables a user to move a target object in a 3-dimensional computer graphics world by moving his hand is presented. The system can interpret hand motions both in a frame fixed to the world and in a frame attached to the user. If the latter is chosen, the user can move the object forward by moving his hand forward even if he has changed his body position.

In addition, the user does not have to keep in mind that his hand is in the camera field of view. The active camera system tracks the user to keep him in its field of view. Moreover, the system does not need any camera calibration. The key for the realization of the system with such features is vision algorithms based on the multiple view affine invariance theory. An experimental system as well as the vision algorithms are demonstrated. Human operation experiments show the usefulness of the system.

Keywords: Camera Calibration(카메라 보정), World-Fixed Frame(공간기준계), Human-Centered Frame(사용자중심계), Affine Invariance(아핀 불변량)

1. 서론

사람의 행동을 이해하여 그 의미를 이해하는 인식시스템의 중요성은 새로운 연구과제로 떠오르고 있다. 특히, 사람에게 행동에 대한 제한을 두지 않고, 일반적인 행동을 이해하는 것은 연구과제로서 중요하다. 기존의 가상현실에서의 사람의 행동을 통해 그 행동이 가지는 의미를 시스템이 이해할 수 있도록 분석하고자 하는 노력은 많이 이루어지고 있으나, 그 인식을 위한 입력장치로써 인체에 직접 부착하여 쓴 시스템이 일반적이었다.

특히, 이는 초창기의 연구로써 그 기법에 제한을 받는 까닭으로 어쩔 수 없이 이런 방법을 써왔다. 그러한 장치로부터 사용자의 사용감의 증대를 위한 목적으로, 영상신호에 의한 방법으로 이러한 행동, 제스처를 인식하는 방법을 택했던 까닭으로 사용자에게는 이 이상의 신체적 제한을 두지 않고 자유로운 행동을 인식하는 시스템에 대한 연구가 이루어지고 있다.

그러나, 사람의 일반행동을 그 대상으로 하기에는 많은 제약이 따른다. 그래서, 사람의 행동 가운데서도, 기계와의 상호작용을 목적으로 하는 경우로 제한을 해서 고려를 하는 것이 일반적이다. 특히, 가상공간에서의 에이전트에 어떤 메시지를 전달하는 것을 주목적으로 생각할 수 있다. 따라서, 여기서는 그 메시지를 에이전트에 주는 가상공간상의 위치나 방향을 가리키는 것을 목적으로 생각한다.

이 때, 3차원 공간을 인식하기 위해서는 지시자(혹은 사용자)가 자신이 있는 공간에서의 위치와 방향으로 이해를 할 필요가 있고, 또 어떤 경우는 지시자가 자신을 중심으로 하는 방향을 가리키는 경우를 생각할 수 있다. 즉, 어떠한 기준계에 대한 인식이 요구되는 것이다. 다만, 이 양자의 경우, 어떤 경우라도 3차원적인 위치와 방향을 알기 위해서는, 지시자의 행동에서 그 사람이 가리키는 손의 위치와 방향으로 이해하여야 한다.

본 시스템은 이러한 작업지시자의 손의 위치와 방향을 인식하는 시스템으로서, 가상공간에서의 위치, 방향을 나타내는 공간적인 의미를 이해한다. 2절에서는 이론적인 배경을 설명하고, 3절에서는 이 시스템에서 취급한 방법을 설명하였다. 그리고, 4절과 5절에서는 각기 시스템의 구성과, 실험 결과를 밝히고, 마지막 6절에서 본 시스템의 평가와 향후의 과제를 설명하였다.

2. 공간지시의 이해

공간지시를 이해하기 위해, 카메라 시스템을 약투시(weak perspective) 카메라로 가정을 하였다. 이로써, 두 대의 카메라에 비친 관심의 대상인 물체가 충분히 떨어져 있을 때는, 관심대상 물체간의 거리에 영향을 받지 않게 인식을 할 수 있게 된다.

3차원의 정보를 이해하기 위해서는 위치와 방향을 계산하여야 한다. 일반적으로 3차원의 강체(rigid body)를 표현하기 위해서는 6 자유도를 계산할 필요가 있으나, 손가락의 위치의 3 자유도와, 손가락이 가리키는 방향의 2 자유도로써도 원하는 위치와 방향을 알 수 있으며, 이를 계산하여 3차원 공간에서의 지시동작을 이해하는 것은 가능하다.

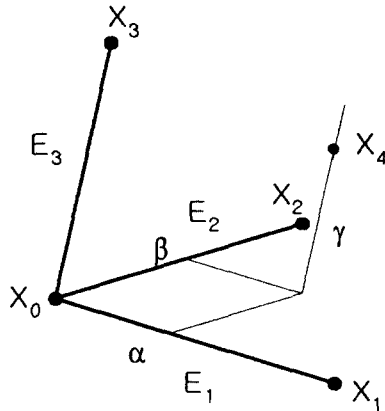


그림 1 어파인 기준계

2.1. 3차원 위치의 계산

종래의 스테레오 비전의 연구에서 3차원 위치는 두 점간의 대응을 통해서, 그 물체간의 거리를 아는 삼각측량법이었다.

그러나, 이러한 스테레오 비전에서는 정확한 3차원적인 정보를 얻기위해, 기준이 되는 카메라 내부의 패러미터, 카메라간의 패러미터를 알 필요가 있고, 이들을 미세 조정을 해야하는 카메라 보정문제 등이 있다. 이러한 방법 대신 카메라 영상내부에 기준점을 둬으로써 이를 해결할 수 있게 되었다.

즉, 동일 평면을 이루지 않는 4점으로 구성되는 프레임을 통해 두 영상간에 이들 점의 대응을 얻었을 때, 이들의 기준점에 대해서, 임의의 한점은 이들 기준점을 잇는 기준벡터를 써서, 좌표값을 계산할 수 있었다.

그림 1에서 보이는 대로, 4점의 기준점을 각기 $X_i, i=\{0, \dots, 3\}$ 라고 했을 때, 임의의 X_4 의 점은 식(1)과 같이 나타낼 수 있다.

$$X_4 = X_0 + \alpha E_1 + \beta E_2 + \gamma E_3 \quad (1)$$

여기서, $E_i, i = \{1, 2, 3\}$ 은 각각 X_0 점과 나머지 $X_i, i = \{1, 2, 3\}$ 을 잇는 기준벡터

(basis vector)이며, α, β, γ 는 임의의 점 X_4 의 이들 벡터의 좌표축에 대한 값에 해당한다.

앞서, 카메라를 충분히 기준점과 측정대상이 되는 점과 떨어져있을 경우, 이는 식(1)이 영상면에서도 같은 꼴의 식(2)로 나타날 수 있다.

$$\mathbf{x}_4 = \mathbf{x}_0 + \alpha \mathbf{e}_1 + \beta \mathbf{e}_2 + \gamma \mathbf{e}_3 \quad (2)$$

식(1)의 대문자가, 영상면에서는 2차원 벡터로서, 소문자로 각각 나타나게 된다. 즉, 식(3)으로 표현할 수 있게 된다.

$$\begin{bmatrix} e_{1u} & e_{2u} & e_{3u} \\ e_{1v} & e_{2v} & e_{3v} \\ e'_{1u} & e'_{2u} & e'_{3u} \\ e'_{1v} & e'_{2v} & e'_{3v} \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} x_{4u} - x_{0u} \\ x_{4v} - x_{0v} \\ x'_{4u} - x'_{0u} \\ x'_{4v} - x'_{0v} \end{bmatrix} \quad (3)$$

여기서, u, v는 각각 화면상의 x, y축의 값, 프라임(')의 여부는 대응하는 영상의 순서를 나타낸다.

따라서, X_4 의 좌표값은 식(3)으로 계산할 수 있게 된다. 즉, 2차원의 식 둘을 써서, 좌표 값 α, β, γ 을 기저벡터를 좌표축으로하는 불변량으로서 계산할 수 있다.

2.2. 3차원 방향의 계산

3차원 방향은 대응된 2점의 특징점의 3차원 특징점의 위치로부터 계산할 수 있다. 그래서, 식(2)의 변형을 쓴 식(4)로 그 방향을 계산하게 된다.

$$\mathbf{x}_5 - \mathbf{x}_4 = \alpha \mathbf{e}_1 + \beta \mathbf{e}_2 + \gamma \mathbf{e}_3 \quad (4)$$

3. 어파인 불변량을 쓴 휴면인터페이스

앞서 보인 특징점의 3차원 위치계산과 방향계산의 방법을 써서, 그림 3에서 보이는 방법과 같이, 지시자의 손에 특징점을 붙인 두 점을 추출하므로써, 그 지시내용을 3차원적인 메트릭을 계산할 수 있게 된다.

즉, 그림 2의 상태에서, 그림 3에서 손의 위치를 가리키기 위한, 손가락 끝점과, 손의 동작으로 손가락이 가리키는 방향을 보이기 위한 손가락의 중심을 써서, 손가락 가리키는 방향을 계산할 수 있다. 이는 2.1과 2.2절에서 밝힌 식으로 어파인 불변량을 계산하므로써 알 수 있다.

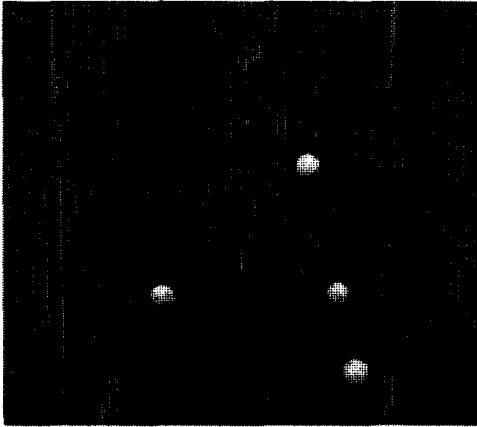


그림 2 사용자 중심 기준계



그림 3 손의 특징점: 위치 및 방향

4. 실험

전체 시스템구성을 그림 4에서 보였다. 복수 영상에서 얻은 4점의 기준점과 2점의 특징점의 어파인 불변량을 계산하기 위해, 다음과 같은 영상처리를 통해서, 간단히 실현할 수 있었다.

영상은 동기화된 두 화상계열을 얻는다. 각 영상에서 얻은 특징점(4점의 기준점과 2점의 측정 대상점)을 서로 대응을 찾아, 측정점의 좌표값의 계산을 한다. 나중에 얻은 계산점의 위치와 방향값을 원도우상에 있는 벡터로 나타내 보인다. 어파인 불변량의 이론을 써서, 구해지는 3차원 위치와 방향을 컴퓨터그래픽으로 그 좌표축에 대응하는 오브젝트의 위치와 방향을 가리키도록 그림을 구성하여 휴먼인터페이스 시스템으로 구성하였다. 이로써, 공간에 놓이는 공간중심의 좌표계를 써서, 계산된 3차원적인 위치와 방향을 가리키는 방법과, 그러한 기준점을 꼭 같이 사람의 몸에 붙이므로써, 사람을 중심으로한 사용자 중심기준계를 이용한 3차원적인 위치와 방향을 가리키는 시스템으로 두가지 시스템을 구성할 수 있었다. 어떤 경우라도, 이 둘의 꼭 같이 기준이 되는 기준점 4점은 두 영상의 화각에 포함이 되어야 한다. 이렇게 하므로해서, 두 영상의 상호간의 보정을 필요로 하지 않는 시스템으로 구성할 수 있었다.

4.1 기준좌표계를 이용한 경우의 실험

그림 1에서 처럼, 프레임을 만들어 보이는 경우로, 네 점을 각기 상호 직교되도록 구성한 좌표계를 마커를 달아서, 구성하였다. 이를 화면의 영상가운데 나타나도록 바닥에 놓고, 이를 기준으로 한 기준계를 공간기준계로 이용하였다. 이렇게 한 공간기준계를 써서, 사용자가 손으로 위치와 방향을 가리키는 경우, 그 내용을 보였다. 계산된 정보를 모니터 화면에 보였다.

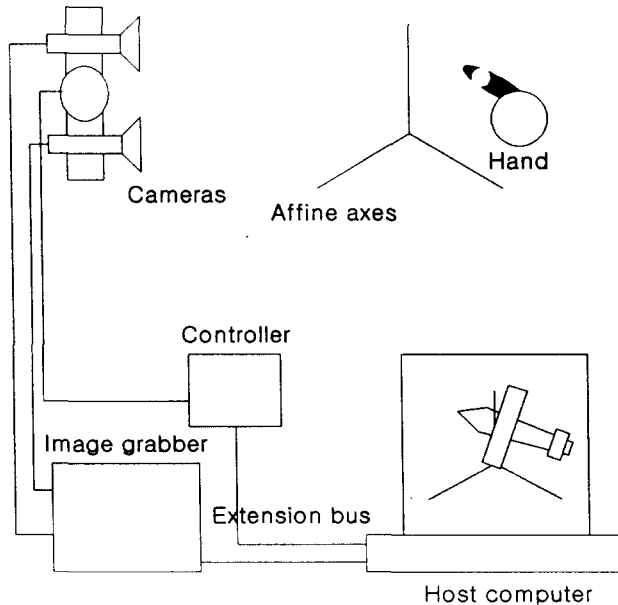


그림 4 시스템 구성도

4.2 사람에게 기준점을 붙인 경우의 실험

그림 2에서 보이는 대로, 이번에는 형광되는 점(마커)을 허리 두 군데와, 어깨 한 점, 그리고 앉은 무릎의 한 점 등 모두 네 점을 붙였다. 이렇게 놓인 네 점을 기준으로 하여 똑 같이 손의 특징점 두 점에도 색도로 구분이 되는 장갑을 끼고, 방향지시를 하였다.

5. 시스템의 평가

이렇게 한 시스템의 사용감에 의한 평가를 하였다. 이 실험에서, 두 가지를 생각해 볼 수 있다. 시스템의 측정정밀도에 대한 평가와, 시스템을 인터페이스로 사용하였을 때의 사용감에 대한 평가이다.

5.1 측정정밀도에 대한 시스템의 평가

여기서는 그림 5에서 보이는 대로, 실제 사용자가 사용환경에 있는 상황을 표현하였다. 두 카메라의 위치는 기준 프레임을 전방 L(250cm)만큼 떨어져서, 직교좌표 기준계의 각각의 길이가 20cm이며, 시점이 각기, (1.0, 0.8, 1.0)과 (0.1, 0.8, 1.0)의 위치에서 바라보게 하였다.

그리고, 그 기준계의 길이가 그림6(a)에서와 같이, 10cm, 15cm, 20cm로 변경하여, 얻은 값의 차이를 나타내 보이고, 그림 6(b)에서는 한 축의 길이를 일정하게 둔 채로, 각도를 각각

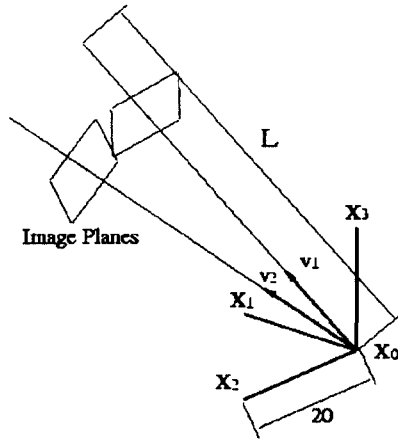


그림 5 측정 상태 구성

90, 75, 60도로 변화시켜 그 측정오차를 계산하였다.

이 때의 측정오차를 표1과 표2로 나타내 보였다. 이 표에서 알 수 있듯이, 실제 이론적인 좌표값과 그 길이의 차이는, 오차로 10%이내임을 알 수 있다. 이는 길이 단위로 길이 3cm 이내임을 뜻한다.

표 2 좌표축 길이의 변화에 따른 오차

축 길이의 변화	이론 좌표값	계산 좌표값
20cm	(0.5, 0.5, 0.0)	(0.47, 0.51, -0.02)
15cm	(0.67, 0.5, 0.0)	(0.66, 0.53, -0.01)
10cm	(1.0, 0.5, 0.0)	(0.96, 0.52, -0.04)

표 3 좌표축 각도의 변화에 따른 좌표오차

축 각도의 변화	이론 좌표값	계산 좌표값
90도	(0.5, 0.5, 0.0)	(0.47, 0.51, -0.02)
75도	(0.5, 0.46, 0.0)	(0.48, 0.40, -0.03)
60도	(0.5, 0.33, 0.0)	(0.55, 0.23, -0.02)

5.2 사용감에 대한 평가

사용감에 대한 평가는, 실제 사용자의 감도에 대한 평가로, 평가기준이 모호하다. 여기서는 다만, 사용자가 실제의 시스템에 얼마만큼 적응하느냐하는 것이 관심이 될 수 있으므로, 이를 배제한 내용으로 실제, 사용자가 자신을 기준으로 한 사용자 중심 기준계와, 환경을 기준으로 한 고정 기준계에서 느끼는 감에 대한 평가를 하였다.

1. 사용자 중심 기준계의 경우: 사용자의 시야를 방해하지 못하도록, 간단한 HMD(head mount display)를 붙여 이를 보면서, 자신의 몸을 회전하는 가운데 조작을 하는 경우, 좌

우 45도 정도까지는 부자연함을 느끼지 않고 지시동작 결과를 얻을 수 있었다.

2. 공간 중심 기준계의 경우: 한번의 기준계를 잡은 다음 자신의 지시동작을 좌우 5도 정도까지는 부자연스러운 느낌이 없이 동작할 수 있었다.

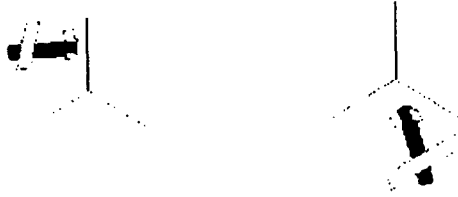
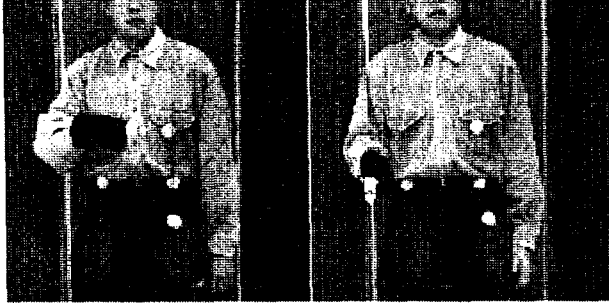


그림 6 실험 결과 : 사용자의 지시동작에 따른 결과표의 표현

그림 6에서는 전체시스템의 이용상태를 보인 것이다. 상단의 두 화면은 사용자가 자신을 중심으로 보인 왼쪽, 아래쪽을 가리키는 지시동작에 각각 컴퓨터그래픽스 화면에서 좌표축으로부터 왼쪽과 아래쪽을 가리키는 결과를 얻을 수 있다.

6. 결론

본 시스템의 평가에 대한 것은 먼저, 앞서 든 시스템의 정밀도에 의한 평가에서 보이는 바대로, 측정시스템으로서의 의미를 든 것과, 다음으로는 휴먼인터페이스 시스템으로서의 영향에 대한 평가를 하였다. 이러한, 사용자 중심 기준계와, 공간 중심 기준계 두 경우에서 기준계와 측정대상인 사람의 손의 특징점의 불변량 계산으로 휴먼인터페이스시스템을 구현해 보였다. 이렇게 함으로써, 카메라의 보정문제와 같은 복잡한 과정을 거치지 않고서도, 쉽게 인터페이스를 이용할 수 있음으로써, 휴먼인터페이스 본래의 목적인 사람에게는 장치로부터 자유로운 채로, 지시동작을 할 수 있는 시스템을 갖추게 할 수 있었다. 또한, 사용자 중심 기준계를 씌우으로써, 사용중 자신의 자세에 대한 염려를 할 필요가 없이 작업을 수행함으로써, 보다 편하고 사람 중심이 되는(user-friendly) 시스템으로 갖출 수 있게 하였다.

그리고, 좌표계를 같은 길이의 직교좌표로부터 조금의 차이를 나타내는 경우의 시스템의 영향은 커다란 지장이 없음을 보였다. 이는 실제적으로 시스템을 이용하는 사람의 수용능력이 시스템이 가지는 능력을 초월하여 이를 이용하는 사람의 인식 능력 범위 내에서 수용을 하며, 주로 사용자의 시각적인 피드백이 작용하기 때문이다. 특히, 시스템의 가정에서 보인 것과 같이 기준계를 길이를 변형시켜보거나, 그 각도를 변형시켜보아도 이들이 이론적으로는 동일 평면을 이루지 않는 네 점의 기준으로는 어떤 경우에도 성립하기 때문인 것으로 알 수 있다. 따라서, 어파인 불변량을 쓴 방법으로써, 휴먼인터페이스를 구축하는 경우에는 잘 적용된다는 것을 알 수 있다.

현재의 시스템을 쓴 실험의 가용범위에 대한 평가는 이것을 어떤 에이전트에 적용하는가 하는 것이 과제로 남는다. 자율 이동로봇을 대상으로하는 원격지 접근 제어나, 접근 후 동작제어에 대한 것이 대상이 될 수 있으며, 가상공간에서의 캐릭터에 대한 상호작용(interaction)도 그 대상이 될 수 있다. 그리고, 당면한 앞으로의 과제로는, 사용자가 보다 자유롭기 위해서, 마커를 이용하는 것을 지양하는 영상처리기법의 도입과, 사람의 자유로운 행동의사를 폭넓게 다루는 연구가 병행되어야 한다.

참고문헌

- [1] J.J. Koenderink and A.J. van Doorn, "Affine structure from motion", Opt. Soc. Am. A. vol.8, No.2, pp. 377-385, 1990.
- [2] J.L. Mundy and A. Zisserman, ed. "Geometric Invariance in Computer Vision", pp.1-39, pp.463-519, MIT Press. 1992.
- [3] M. Fukumoto, K. Mase, and Y. Suenaga, "Realtime detection of pointing actions for a glove free interface", Proc. IAPR Workshop on Machine Vision Application, pp.473-476, 1992.
- [4] R. Cipola, Y. Okamoto, and Y. Kuno, "Robust structure from motion using motion parallax", Proc. IEEE 4th International conference on computer vision, pp.374-382, 1993.
- [5] K.-H. Jo, "Human Gesture Recognition using Human-Centered Frame and Task Knowledge for Smooth Human-Computer Interaction", Ph.D Thesis, 1997.